Statistical Risk Estimation for Communication System Design

by
Alessandra Babuscia
B.S., Politecnico di Milano (2005)
S.M., Politecnico di Milano (2007)

Submitted to the Department of Aeronautics and Astronautics
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
May 2012

Author .....
Department of Aeronautics and Astronautics
May 24, 2012

Certified by .....
David W. Miller
Professor of Aeronautics and Astronautics, Thesis Committee Chair

Certified by .....
Moe Z. Win
Associate Professor of Aeronautics and Astronautics

Certified by .....
Kar-Ming Cheung
Technical Group Supervisor at NASA Jet Propulsion Laboratory

Certified by .....
Alvar Saenz-Otero
Research Scientist at MIT Space System Laboratory

Accepted by .....
Eytan H. Modiano
Professor of Aeronautics and Astronautics,
Chair, Graduate Program Committee

Statistical Risk Estimation for Communication System Design

by

Alessandra Babuscia

Submitted to the Department of Aeronautics and Astronautics
on May 24, 2012, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

Spacecraft are complex systems that involve many subsystems and multiple relationships among them. The design of a spacecraft is an evolutionary process that starts from requirements and evolves over time. During this process, changes can affect mass and power at component, subsystem, and system level. Each spacecraft has to respect overall constraints in terms of mass and power. The current practice in system design deals with this problem by allocating margins to individual components and to individual subsystems. However, a statistical characterization of the fluctuations in mass and power of the overall system (i.e. the spacecraft) is missing. This lack can result in a risky spacecraft design that might not fit the mission constraints and requirements, or in a conservative design that might not fully utilize the available resources. This problem is especially challenging at the initial stage of the design, when high levels of uncertainty due to lack of knowledge are unavoidable.

This research proposes a statistical approach to quantify the likelihood that the design of a spacecraft would meet the mission constraints in mass and power consumption, focusing on the initial stage of the design. Due to the complexity of the problem and the different expertise required to develop a complete risk model for a spacecraft design, the scope of this research is focused on risk estimation for a specific spacecraft subsystem: the communication subsystem. The current research aims to be a "proof of concept" of a risk-based design approach, which can then be further expanded to the design of other subsystems as well as to the whole spacecraft.

The approach presented in this thesis includes a baseline communication system design tool, and a statistical characterization of the design risks through a combination of historical mission data and expert opinion. Different statistical techniques are explored to ensure that the amount of information extracted from data and expert opinion is maximized. Specifically, for statistics based on data, Kernel Density Estimator is selected as the preferred technique to extract probability densities from a database of previous space missions' components. Expert elicitation is generated

through a four-part model which quantifies experts' sensitivity to biases, and uses this measurement to compose properly the assessments from different experts. Finally, an optimization framework is developed to compare multiple possible design architectures, and to select the one that minimizes design objectives, like mass and power consumption, while minimizing the risk associated with the same metrics.

Examples of missions are applied to validate the model. Results show that the statistical approach recognizes whether the initial estimate of the system is an overestimation or an underestimation, providing a valuable tool to measure the risk of a communication system at the initial state of the design. Specifically, statistics based on historical data and on expert elicitation allow the designer to size contingency properly, providing a reliable estimation of mass and power in the initial stage of the design.

Thanks to this method, the communication system designers will be able to evaluate and compare different communication architectures in a risk trade-off prospective across the evolution of the design. Extensions to different subsystems and to additional metrics (like cost) make this model applicable to a wider range of problems.

Thesis Supervisor: David W. Miller
Title: Professor of Aeronautics and Astronautics, Thesis Committee Chair

Thesis Supervisor: Moe Z. Win
Title: Associate Professor of Aeronautics and Astronautics

Thesis Supervisor: Kar-Ming Cheung
Title: Technical Group Supervisor at NASA Jet Propulsion Laboratory

Thesis Supervisor: Alvar Saenz-Otero
Title: Research Scientist at MIT Space System Laboratory

Acknowledgments

This work has been supported by the MIT Department of Aeronautics and Astronautics, the NASA Jet Propulsion Laboratory Graduate Fellowship Program, DARPA, and the Zonta International Amelia Earhart Fellowship. I am grateful to all of these institutions for their financial support throughout my graduate work.

I would like to thank my advisor, Prof. Dave Miller, for your guidance. I appreciate all of the advice, challenges, support and most of all the fact that you gave me a chance, betting on me when nobody else was willing to. I hope it paid off.
I would like to thank Dr. Alvar Saenz Otero for being such a great mentor through all my graduate studies at MIT. Thank you for your suggestions and for always asking the right question at the right moment...it really helps.
I would like to thank Prof. Moe Win for providing guidance and advice as a part of my thesis committee. Your insights and suggestions have helped to shape this work.
Thank you to Dr. Kar-Ming Cheung for challenging me with this fascinating topic, for having hosted me at NASA JPL, and for your guidance during the development of this research. I truly enjoy solving complex problems with you.
I would also like to thank the META team (Prof. John Deyst, Prof. Karen Willcox, Dr. Doug Allaire, Chelsea He, and Emily Clements) for your support and insights in the last part of this work.

Thank you to Prof. Sara Seager for having been such an incredible inspiration for me. I truly appreciate all your guidance during my years at MIT.
Thank you to Jennifer Craig and Jane Connor for all your support and your friendship. I enjoyed every moment we worked together.
Thank you to Prof. Eytan Modiano for initially welcoming me to MIT and for your technical suggestions over the years.
Thank you to Marilyn Good, Sharon Brown, and Paul Bauer for your assistance and for making my life so much easier.
Thank you to the Aeronautics and Astronautics Department Staff over the years, including Barbara Lechner, Beth Marois, and Marie Stuppard for your support during my graduate studies.

I would also like to thank the entire Space System Lab over the years, including Farah Alibay, Bruno Alvisio, Zach Bailey, Ingrid Beerer, Thomas Chiasson, Lucy Cohan, Corey Crowell, Phillip Cunio, Sydney Do, Greg Eslinger, Jack Field, Gwendolyn Gettliffe, Erica Gralla, Koki Ho, Jake Katz, Matthew Knutson, Whitney Lohemeyer, Christopher Mandy, Annie Marinan, Matthew McCormack, Ryan McLinko, Swati Mohan, Sreeja Nag, Jameson Nash, Sarah Nothnagel, Michael O' Connor, Sung Wook Paek, Eric Peters, Chris Pong, Jaime Ramirez, John Richmond, Joe Robinson, Hal Schmidt, Matt Smith, George Sondecker, Kevin Stout, Brent Tweedle, and Evan Wise.

A special thanks to my office mate Dr. Rebecca (Becky) Masterson for patiently reading my thesis and for being such a nice person to share the office with: I truly enjoy working, laughing, and chatting with you.

Thanks to my friends Mary Knapp, Rebecca Jensen Clem, to all the friends from MIT Gospel Choir, MIT Grad Student Choir, and MIT Isshinryu Karate Dojo for making my life at MIT more balanced and enjoyable.

Finally, thank you to my family and to all of my friends (especially Lorenza Deluchi and Valentina Vitale... I miss you girls).

Thank you to my boyfriend Jeremy Stone for your love, encouragement, and for always being there for me.

Special thanks to my parents, Patrizia e Renato, for your love, support and for inspiring me with your example all my life. I could not have done this without you...We made it.

Contents

List of Figures

List of Tables

List of Acronyms

AFOSR - Air Force Office of Sponsored Research

ANOVA - ANalysis Of VAriance

CASTOR - Cathode Anode Satellite Thruster for Orbital Repositioning

CDIO - Conceive Design Implement Operate

CDF - Cumulative Distribution Function

CDR - Critical Design Review

COUHES - Committee on the Use of Humans as Experimental Subjects

COTS - Commercial-Off-The-Shelf

DCFT - Diverging Cusped-Field Thruster

DNP - Develop New Projects

DSN - Deep Space Network

EELV -Evolved Expendable Launch Vehicles

EIRP - Equivalent Isotropic Radiated Power

ESA CDF - European Space Agency Concurrent Design Facility

ESPA - EELV Secondary Payload Adapter

FEA - Finite Element Analysis

GEO - Geostationary Earth Orbit

GRB - Gamma Ray Burst

GUI - Graphical User Interface

HETE - High Energy Transient Explorer

HGA - High Gain Antenna

KDE - Kernel Density Estimator

LEO - Low Earth Orbit

LGA - Low Gain Antenna

MCR - Mission Concept Review

MGA - Medium Gain Antenna

MSDO - Multidisciplinary System Design Optimization

MSE - Mean Square Error

PDF - Probability Density Function or Probability Distribution Function

PDR - Preliminary Design Review

PRA - Probabilistic Risk Assessment (Analysis)

SNR - Signal to Noise Ratio

SQP - Sequential Quadratic Programming

SRI - Stanford Research Institute

STK - Satellite Tool Kit

SSPA - Solid State Power Amplifier

TERSAT - Trapped Energetic Radiation Satellite

TRL - Technology Readiness Level

TWTA - Traveling Waves Tube Amplifier

UHF - Ultra High Frequency

UNP - University Nanosatellite Program

VLF - Very Low Frequency

VHF - Very High Frequency

Chapter 1

Introduction

1.1 Motivation and Problem Description

Spacecraft are complex systems composed of many subsystems with multiple dependencies between them. The complexity of such designs requires a team of engineers to develop different subsystems and to coordinate the design at the system level. The design starts from requirements and evolves across multiple iterations. The time intervals between the initial Mission Concept Review (MCR) [95], the Preliminary Design Review (PDR), and Critical Design Review (CDR), varies according to the mission. However, months or years can pass between one stage and the next. The complexity of the design process, the number of people involved, and the time intervals between the reviews inevitably leads to changes in the design. Particularly, different studies (some conducted at NASA Jet Propulsion Laboratory, Team X [96]) reveal that in the design of a mission, one of the major challenges is given by:

"Significant deviations from expected mass, power, cost or performance for any element of the spacecraft." [91]

At each new stage of the design, engineers perform new analyses and acquire better understandings of their respective subsystems, reducing uncertainties in key metrics such as power and mass. However, prior to initial MCR and the PDR, engi-

neers are forced to estimate quantities without a complete knowledge of the system. Some components are not finalized, or they have never been fabricated before. Some components may be fabricated in house, while some of them can be fabricated externally.

As a result, a problem arises: engineers are forced to estimate values for mass and power consumption at the component level and at the subsystem level. These values inevitably fluctuate over time.

This problem is critical in spacecraft design because each mission is subjected to constraints in total mass and power allocation. On one hand, the fluctuations can cause the system to exceed its limitations, causing a consistent process of redesign, with associated cost. On the other hand, fluctuations can generate too conservative designs, whenever excessively overestimated contingencies are applied.

Contingencies represent the traditional approach to counteract risks associated with design fluctuations. They are generally applied at the component level, at the subsystem level, and at the system level. Contingencies are deterministic additions that do not statistically model the probability that the mass or power will effectively increase. Contingencies try to overcome the risk of designing a system that does not respect constraints, but they can also cause overestimation. Overestimated systems result in inefficient designs with consequent reduction of the utility of a system.

As an example, consider a commercial satellite system to provide communication services at different bands. The number of channels at different bands carried on board has not yet been finalized. The design goal is to maximize the number of channels while respecting mass constraints. We define as x the total mass allocated to the channels. In this simple model, the communication equipment (antennas, transceivers, amplifiers) for each channel has an estimated mass of y. To avoid solutions that can exceed the total mass allocated (x), contingencies are applied: the extra mass for each channel is defined as z. The utility of a commercial satellite can be defined as the number of channels that the spacecraft can carry on board. Given

the previous assumptions, the maximum number of channels N is:

$$N \leq \frac{x}{y+z} \tag{1.1}$$

The greater the value of z is, the more conservative are the solutions, with consequent reductions in the number of channels and the total utility of the satellite. Also, it can happen that the system at the PDR is equipped with a certain number of channels ($N_{PDR} = N_1$) and then, at CDR, design contingencies (z) decrease because engineers have gained a better understanding of the components, discovering that the initial mass was overestimated. At that point, it may be possible to increase the utility of the system by carrying another channel ($N_{CDR} = N_2 > N_1$), but this operation will require redesign and corresponding cost.

In conclusion, the problem of uncontrolled fluctuations can cause expensive redesigns, whenever the system results too conservative or too risky. This concept is summarized in Figure 1-1, which illustrates the design process and the loop caused by redesign. A better statistical understanding of the fluctuations has the potential to improve the efficiency of the design process by reducing the number of design iterations.

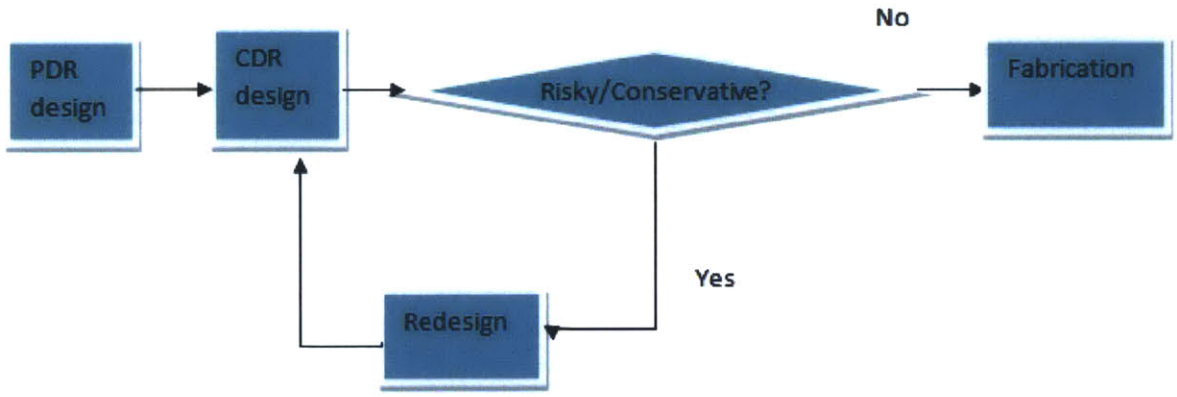

Figure 1-1: Flux diagram of design process. The objective of the thesis is to improve the statistical characterization at PDR to minimize the iterations with consequent savings in cost and human resources.

The problem is critical for all the subsystems, especially for communication. Different studies ([140], [135], [120], [124]) suggest that certain communication systems

represent a consistent fraction of total spacecraft mass and power. Hence, fluctuations can have an impact on the entire spacecraft design. This concept is valid for specific applications:

- Commercial satellites: The communication subsystem is the *payload.* Hence, statistical risk estimation for this application can consistently impact the utility of the entire system.

- Space relays: In space explorations spacecraft are designed to act as communication relays for other spacecraft that may be orbiting, or on the surface of another planet. Relay spacecraft are very similar to commercial satellites in the sense that they both provide a communication service. Moreover, given the distance at which they are located, they need to carry high energy communication equipment. This equipment drives spacecraft mass and power consumption, with consequently high impacts on the respective fluctuations.

- Small satellites (ChipSat, CubeSat, ESPA class satellites): Mass and power consumption constraints play a key role in increasing launch opportunities for small satellites. Hence, communication systems have to be highly optimized and the fluctuations in mass and power need to be controlled in order to assess the feasibility of the system.

More details regarding the research focus are presented in Section 1.4.3.
In summary, the problem of unpredicted mass and power fluctuations is critical for communication systems. A description of the fluctuations' causes and a presentation of the state of the art approaches to counteract the problem are discussed in the next section.

1.2 State of the Art

A consistent amount of work has been done in identifying the reasons for which the final value of masses and powers diverge with respect to the initial values. Specifically, the causes for fluctuations can be categorized as:

- Fluctuations due to lack of human interaction across designers ([91], [36] , [14]): A team composed of different engineers working on a project can, for lack of interaction, develop subsystems which do not fit together and which require a partial redesign with consequent mass and power fluctuations.

- Fluctuations due to lack of knowledge ([91], [14]): A general lack of definitive knowledge in mission implementation, or in fabrication of the components leads to values of mass and power affected by fluctuations.

Fluctuations due to lack of human interaction are generally overcome through the development of interactive facilities. An example of this kind of facility is Team X at NASA Jet Propulsion Laboratory (JPL). Pictured in Figure 1-2, Team X [96] is a multidisciplinary facility that utilizes concurrent engineering methodologies to complete rapid design, analysis and evaluation of mission designs. This center consists of networked workstations, a data management infrastructure, interactive graphic displays, simulation tools, historical data repositories, and a shared project model that the design team updates [96].

Figure 1-2: Engineers in the Project Design Center during one of the daily Team X sessions. The networked stations are completely interconnected. A single change in any of the stations is automatically reflected in the subsystems analyzed by the other stations. Source: [96].

Another example is ESA Concurrent Design Facility (CDF), shown in Figure 1-3. Similarly to JPL Team X, Concurrent Design Facility (CDF) is equipped with a network of computers, multimedia devices and software tools, that allow a team of experts from several disciplines to apply the concurrent engineering method to

the design. It facilitates an effective interaction of all disciplines involved, ensuring consistent results in a much shorter time [44].

Figure 1-3: Engineers during an ESA CDF session. The networked stations are completely interconnected. A single change in any of the stations is automatically reflected in the subsystems analyzed by the other stations. Source: [44].

Although much work has been done to overcome the lack of interaction among engineers, less work has been done to overcome the lack of knowledge, since in the initial stage of the design it is impossible to avoid a certain level of uncertainty. However, statistical instruments can be used to quantify fluctuations. The development of a statistical characterization for mass and power fluctuations can be a key factor in the design of a mission. Recalling the satellite example where the number of channels was conditioned by the contingency metric z, a more accurate definition of the value of z at the PDR can lead to a more precise assessment of the number of channels, avoiding redesign at CDR.

Hence, this research proposes an approach to statistically characterize mass and power fluctuations due to the lack of knowledge. Some key aspects of the proposed approach are discussed in the next section.

1.3 Proposed Approach

Current practice in system modelling deals with fluctuations by allocating margins (contingencies). As an alternative, this research proposes a methodology that uses

statistical estimation to quantify risks at the initial stage of the design (PDR). The methodology aims to obtain a probabilistic assessment of how likely the system is to exceed constraints in mass and power consumption. This methodology allows engineers to evaluate the PDR design by performing risk-performance analysis. Moreover, it should give accurate indications on how much the current PDR design is likely to vary up to CDR, allowing engineers to take appropriate countermeasures. A synthesis of how the methodology fits in the design process is given in Table 1.1.
The table shows an example of the process of sizing a communication system. Total mass and power values at PDR are m_{PDR} and p_{PDR}. The contingency method takes into account risk by adding a 30% margin ([135], [95]). The CDR values are m_{CDR} and p_{CDR}. Statistical risk estimation aims to act on the risk characterization by identifying margins (α_{risk}, β_{risk}) at PDR which are more reliable than the state of the art contingency and which are closer to the final CDR value.

Table 1.1: Difference between traditional approach and statistical risk estimation. Statistical risk estimation aims to act on the risk characterization by identifying values of margins closer to the final CDR value.

Methodology	PDR		PDR + Risk		CDR	
	Mass	Power	Mass	Power	Mass	Power
State of the art	m_{PDR}	p_{PDR}	$m_{PDR} + 30\%$	$p_{PDR} + 30\%$	m_{CDR}	p_{CDR}
Statistical risk estimation	m_{PDR}	p_{PDR}	$m_{PDR} + \alpha_{risk}$	$p_{PDR} + \beta_{risk}$	m_{CDR}	p_{CDR}

Hence, the methodology should quantify fluctuations at PDR in order to:

- Avoid too risky solutions: If risk estimation is too high, the margins allocated by statistical risk estimation can be excessively high. Hence, engineers will be aware of the level of risk of the solutions, and they can decide to accept/modify the design.

- Avoid too conservative solutions: If risk estimation is low, the margins allocated by statistical risk estimation can be excessively low, and resource allocation will

be more efficient.

To develop the statistical estimation approach, two information sources are exploited:

1. **Historical data:** Real mission data for communication system components are used to shape the probability distributions for the fluctuations.

2. **Expert opinion:** If the amount of historical data is excessively limited, experts in the field provide additional information.

The information conveyed by data and by experts is elaborated through different statistical techniques (kernel density, expert elicitation, statistical link analysis, linear composition of experts' assessments, Bayesian composition) to calculate risks. Finally, optimization is used to select the best architecture for which risk is below a certain threshold.

A final remark can be made regarding the risk metrics used. The approach aims to quantify risks in power and mass; however cost is also a very important metric in the design process. Unfortunately, it is very difficult to have access to cost information to validate any methodology. Hence, in this thesis the approach will be focused on mass and power fluctuations, while some methodological considerations on expanding the model to cost risk analysis are discussed in Chapter 8.

1.4 Problem Statement and Research Focus

Statistical risk estimation aims to improve risk estimation at PDR by quantifying the probability that the design will exceed constraints over the evolution of the design. To obtain such estimation, it is necessary to exploit different sources of information to maximize the precision of the estimations, and to evaluate the total effect of different components. The next section defines the main terms presented in the thesis. The following sections give a detailed explanation of problem statement, research focus and thesis objectives.

1.4.1 Definitions

In this subsection the principal terms used in the thesis are defined to avoid ambiguity.

- **Design value:** average value of a design metric (mass, power, cost).
- **Fluctuation:** deviation from an initial design value. As an example, component x has an initial estimation of mass of 3 Kg (design value). However, at CDR the final value of mass for component x becomes 3.5 Kg. Hence, the fluctuation is the 0.5 Kg mass increment from the initial design value.
- **Cap:** limit which cannot be exceeded. Example: mass cap indicates the maximum mass allocated to the system.
- **Design risk:** Risk that a fluctuation will exceed a cap during the evolution of the design.
- **Probabilistic Risk Assessment (PRA):** a probabilistic approach to evaluate risk.
- **Probability Distribution Function (PDF, also Probability Density Function):** a function of a continuous random variable whose integral over an interval gives the probability that its value will fall within the interval [90].
- **Cumulative Distribution Function (CDF):** the integral over a defined interval of the Probability Distribution Function (PDF). It is used to express the probability that a random variable will be less than or equal to a given value.
- **Tail function (Risk function):** the complement of the CDF. It expresses the probability that a random variable will be greater than a given value. It is generally used in PRA problems, where the focus is on quantifying whether or not a risk will be greater than a certain threshold.
- **Non-parametric density estimation:** "the construction of an estimate, based on observed data, of an unobservable underlying probability density function [115]."

- **Tails' divergence:** distance between the real and estimated value of a tail. The mathematical expression is in Chapter 4.

- **Estimation error:** integral of the divergence of the random variable. It is analogue to the *L1* [43] metric, but it is calculated using tail functions. The mathematical expression is given in Chapter 4.

- **Expert elicitation:** the construction of a probability density function based on expert information.

- **Experts composition:** the process of combining information from different experts in a unique probability function.

- **Bias:** In expert elicitation a bias (also called cognitive bias) is the "human tendency to make systematic errors in certain circumstances based on cognitive factors rather than on evidence [25]."

- **Heuristics:** simple, efficient rules, applied to make decisions, to express judgments, to solve problems, typically when facing incomplete information. These rules can lead to systematic errors or biases.

- **Calibration:** measurement of the agreement between expert opinion (subjective probability) and observed relative frequency.

- **Seed variables:** variables used to perform calibration. Generally, seed variables are similar to the ones used to perform elicitation, but for them a probability density or at least a realization is known.

- **Adverse tolerance and favorable tolerance:** upper and lower bounds of a bounded probability density function in statistical link analysis (Chapter 5).

- **Communication architecture:** set of components necessary for a spacecraft to satisfy communication requirements.

- **Baseline design:** a model that receives as inputs certain communication parameters (frequency, data rate, number of channels, level of redundancy, receiver

characteristics) and generates as outputs one or more iso-quality communication architectures. Each component in the baseline design is specified by design values of mass, power, and gain.

- **Risk model:** a model that receives as inputs one or more communication architectures calculated in the baseline design model, and evaluates the design risks in mass and power consumption.
- **Power:** power consumption if not differently specified.
- **Generalized expert opinion:** expert knowledge extrapolated from the interviews and applied to the mathematical optimization framework (Chapter 7).

The following subsection describes the problem statement.

1.4.2 Problem Statement

In the previous sections, the problem of fluctuations in mass and power for a communication system has been described. Hence, the research question becomes: how do we develop a model to quantify the risk that a communication system design at PDR will exceed the mission caps as the design matures? Specifically, the problem statement can be decomposed into the following sub-questions:

1. What information can be used to quantify fluctuations in mass and power consumption?
2. How can different sources of information (data and expert opinion) be combined to generate statistics?
3. Which statistical technique or combination of techniques can be used to exploit the information collected?
4. How can the different fluctuations be combined to quantify total risks?
5. How can risk estimation be embedded in a communication system design tool?

6. How can risk estimation be integrated in a design optimization process?

Answering these questions requires the categorization of different information sources, the identification of the optimal statistical techniques for the different sources, and the development of a risk assessment model.
This problem concerns the design of an entire spacecraft. However, the choice of the thesis is to explore the problem from the point of view of design of communication systems. The following subsection clarifies the motivations for focusing the research on this specific area.

1.4.3 Research Focus

The focus of this research is on risk estimation for communication systems. The motivations are the following:

- The communication system is a key factor in different spacecraft [140]: It is the case for relay communication satellites or small satellites where the fractions of mass and power allocated to the communication system are consistent. Figures 1-4 and 1-5 show that in certain spacecraft, communication mass fraction reaches up to 28% of the total mass, while power fraction reaches up to 32% of the total power. It is difficult to quantify the correspondent fluctuations of mass and power consumption. However, according to the traditional approach of contingencies [135] the expected fluctuations in mass and power between PDR and final prototype are in the range of 30%. Hence, in the spacecraft illustrated in Figures 1-4 and 1-5, the fluctuations due to the communication subsystem impact up to 9% (a 30% of the 32%) of the total spacecraft mass and power. An impact between 5% and 9% can significantly affect the final spacecraft design; hence a statistical assessment of risks is important. The satellites selected for the analysis in Figures 1-4 and in 1-5 are JPL missions and commercial satellites. The selection does not represent an exhaustive set of all possible spacecraft, but it includes different satellites (space exploration, commercial) of various sizes. Hence, it can be considered meaningful for the purpose of this

analysis.

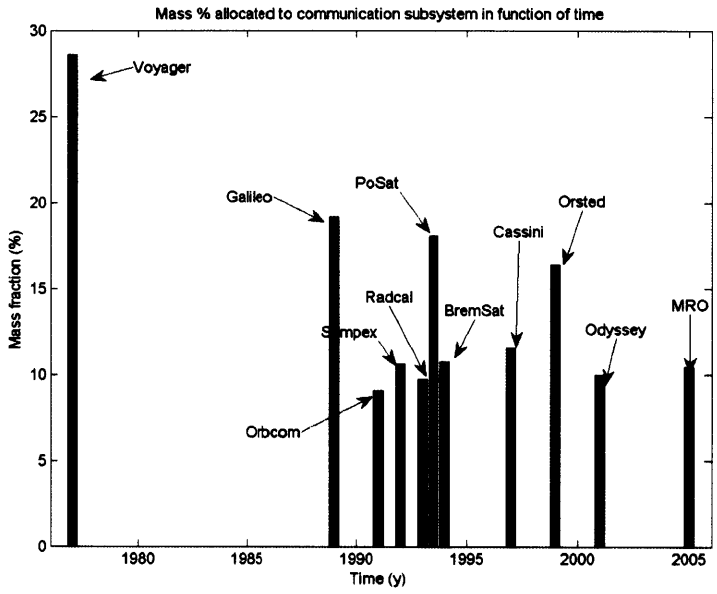

Figure 1-4: Mass of the communication system as a fraction of the total mass in selected spacecraft. The communication mass fraction reaches up to 28% of the total mass.

- Communication represents the core of commercial satellites applications. In these cases, a better statistical assessment for mass and power fluctuations will allow the designers to improve the payload design, by increasing the number of channels that the satellite can carry as described in Section 1.1.
- The current research aims to be a "proof of concept," which can then further be extended to the different subsystems, as well as to the whole spacecraft design process (Chapter 8).

The following section underlines thesis objectives.

1.5 Thesis Objectives

The objective of the research is to develop a mathematical approach to improve PDR design by assessing the likelihood that the design of a space communication

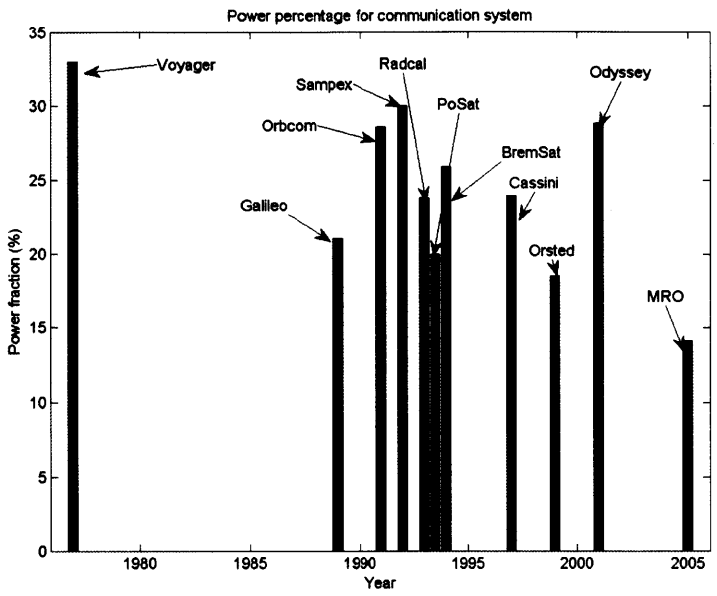

Figure 1-5: Power consumption of the communication system as a fraction of the total power consumption in selected spacecraft. The communication power fraction reaches up to 32% of the total power consumption.

system would meet the spacecraft and mission constraints as the design matures. This methodology would allow designers to quantify risk trade-offs. The thesis's sub-objectives can be categorized according to five main research areas, which are discussed in the different chapters of the thesis: historical data statistics (Chapter 4), expert elicitation statistics (Chapter 5), risk model (Chapter 6), optimization (Chapter 7), and model extensions and generalizations (Chapter 8). A break-down of the thesis sub-objectives is as follows:

1. Historical data statistics: identification of the statistical technique that most appropriately describes the sets of data.

   (a) Development of a comparison test.

   (b) Determination of the number of samples required for a maximum tolerable error.

   (c) Population of database and probability density function generation.

2. Expert elicitation statistics: identification of an opportune approach to elicit expert opinion in risk characterization.

   (a) Quantification of expert's sensitivity to biases and heuristics.

   (b) Quantification of expert's calibration.

   (c) Selection of opportune elicitation techniques.

   (d) Development of an approach to combine multiple experts' judgments.

   (e) Development of an approach to combine historical data statistics and expert statistics.

3. Risk model: development of a modelling tool to perform risk estimation at the PDR stage.

   (a) Development of a baseline design to identify sets of communication architectures.

   (b) Integration of data approach and expert approach in the risk estimation.

   (c) Computation of risks.

   (d) Development of comparisons with current approach to infer guidelines on risk quantification for communication system design.

4. Optimization: development of a mathematical framework to represent the risk estimation problem as a Multidisciplinary System Design Optimization (MSDO) problem.

   (a) Formulation of objectives and of design constraints.

   (b) Integration of risk assessment in MSDO methodology.

   (c) Optimization of the system under risk.

5. Model extensions and generalizations: identification of an approach to extend the model to different subsystems and to multiple metrics.

The following section summarizes thesis contributions.

1.6 Contributions

The primary contributions of the thesis are summarized below. They will be discussed in greater detail in Chapter 9.

1. Creation of a statistical model to improve PDR design by assessing the risk that a certain component or a group of components in the communication system will exceed the caps in mass and power as the design matures.

2. Creation of an approach based on known benchmarks to compare density estimation techniques and identification of Kernel Density Estimator as the most appropriate technique to model space mission historical data.

3. Creation of a methodology to elicit expert opinion by measuring expert sensitivity to biases and mis-calibrations.

4. Creation of an approach to aggregate opinions from multiple experts into an unique estimate.

5. Development of a methodology derived from Bayesian analysis to integrate expert opinion and historical data.

6. Application of the model with test cases and improvement with respect to traditional approaches.

7. Development of a mathematical framework to model risk estimation for a communication system as a constrained optimization problem.

8. Identification of guidelines on performing risk estimation for space communication systems.

9. Identification of guidelines to generalize and to extend the methodology to other subsystems up to the system level.

1.7 Thesis Outline

Figure 1-6 shows an overview block diagram of the organization of the thesis. Chapter 2 presents a review of literature that is relevant to the various aspects of this thesis. Chapter 3 presents the methodology used throughout the thesis. Chapters 4 and 5 illustrate the statistics used to perform risk estimation. Specifically, Chapter 4 is focused on the use of historical mission data to perform statistical risk estimation: analysis and comparison of different density estimation techniques, the methodology used to categorize historical data, and the process of shaping probability densities from data are presented. Chapter 5 describes the methodology developed to apply expert opinion to statistical risk estimation: the methodology, the content of the three-part interview, and results of the experimental tests performed are described. Chapter 6 presents how the statistics are applied to perform statistical risk estimation: the baseline design model, its validation, the risk model, and test case examples are discussed. Chapter 7 addresses the optimization problem: a framework to incorporate statistical risk estimation in an MSDO methodology, and results are discussed. Chapter 8 proposes the extension of the model to cost risks, and the generalization of the model to other spacecraft subsystems. Finally, Chapter 9 concludes the thesis with a summary, contributions, and suggestions for future work.

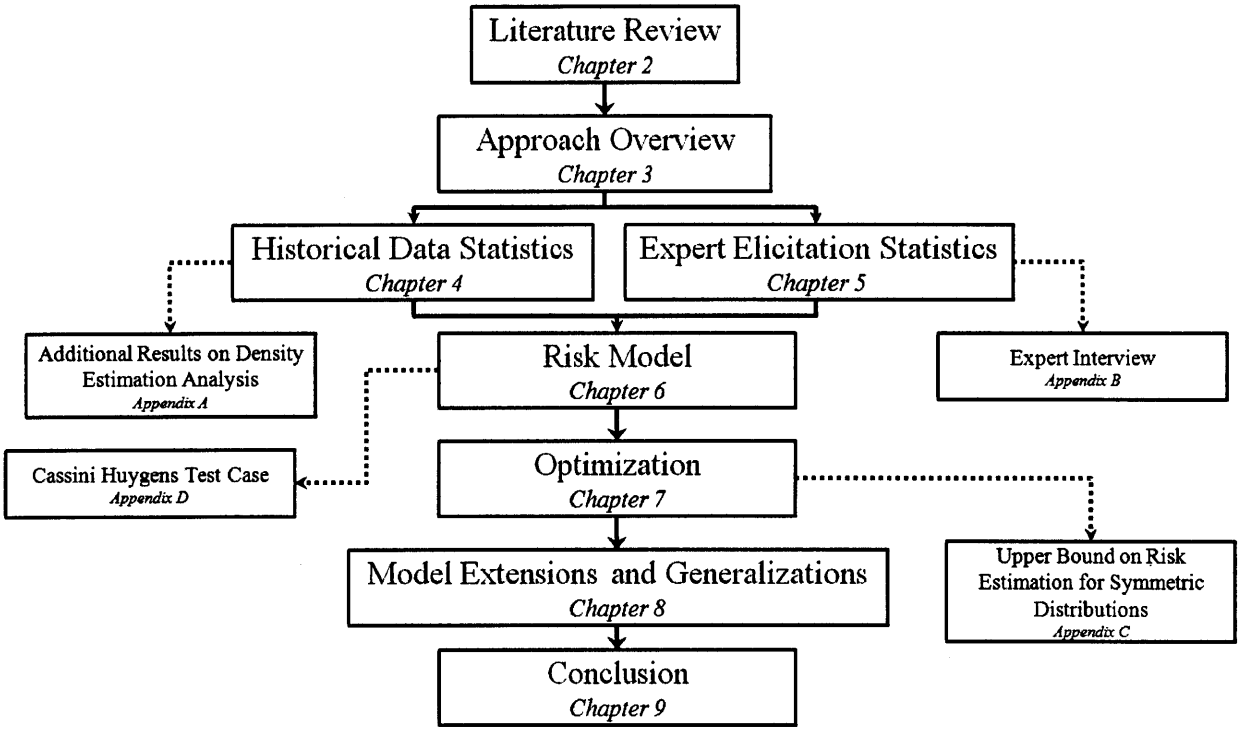

Figure 1-6: Thesis Outline.

Chapter 2

Literature Review

2.1 Overview

The literature review is organized into three main areas:

- Risk characterization and modelling: presentation of the state of the art in the fields of design risk identification, classification, analysis and quantification.
- Statistical techniques: discussion of the principal techniques used to perform non-parametric estimation, expert opinion elicitation, and composition of multiple distributions.
- Modelling and optimization for communication systems: analysis of the current techniques to model space communication system, and to perform validation and optimization.

Each of the following sections is dedicated to one of the three areas. The principal works in the field are discussed, and the gaps fulfilled by the thesis are underlined.

2.2 Risk Characterization and Modelling

Different publications analyze the problem of design risk. Literature in this area can be divided in two main categories: qualitative and quantitative approaches to design

risk. Most of the literature works belong to the first category, and the few works in the field of quantitative approaches apply only statistics based on experts' opinions and not on data.

In the category of qualitative approaches to design risk, an article on risks by Cortellessa [36] categorizes the different risks and defines design risks as the risks of exceeding mass and power constraints over the evolution of the design. Meshkat ([91], [92]) analyzes design risk on the bases of her experience in Team X, and she describes different cases in which she observed *significant deviations in mass and power* over the evolution of the design. She suggests a possible way to quantify risks using historical data, one of the techniques that is explored in this research. Barrientos [14] analyzes the causes of design risks focusing on lack of human interaction across engineers. Her study is mainly qualitative and focused on the interaction developed in cooperative environments like Team X or in university satellite projects (ref. to Stanford). Oberkampf et al. [98] study the causes of risks distinguishing between aleatory and epistemic: aleatory risks are due to aleatory phenomena, while epistemic risks are due to lack of knowledge. This research is focused on epistemic risks. Oberkampf et al. also distinguish possible available data to quantify risks. Three categories are identified: strong statistical information, sparse statistical information, and intervals. The type of data used in this thesis is given mostly by sparse statistical information from a database (Chapter 4), and by intervals (bounds) given by the experts (Chapter 5). Asnar [5] developed qualitative risk analysis techniques to deal with the problem of selecting across design alternatives. Other qualitative analyses focused on human interactions and flux decision trees are described in the works of Cornford [35], Du [139], and Grote [57]. Lough et al. [84] are focused on counteracting early design risks, but they refer mostly to failure risks and not to design risks.

Fewer publications can be found in the field of quantitative characterization of design risk. Dillon et al. [41] propose a model focused on programmatic cost-risk analysis for space missions. However, the model is focused on managerial risks: decision vs. cost as a managerial problem. The emphasis is on flexibility between architectures

and cost trade-offs, while the emphasis in this thesis is on performing the statistical risk characterization for specific architectures. A similar approach can be found in the works of Bearden [15], and of Dean [40].

A quantitative work in statistical estimation for design risk is presented by M. Fuchs and A. Neumaier. Their publications ([49], [51], [52], [50]) describe an approach based on the creation of a n-dimensional cloud of uncertainties in which the different architectural solutions lie. The dimensions of the cloud are given by the different metrics on which the uncertainty is evaluated, and the cloud becomes smaller as the desired confidence in the solution increases. The shape of the cloud is defined by expert opinion. A three- and two-dimensional example of the cloud is shown in Figure 2-1. It is possible to notice the different design points and the shape of the cloud, defined only by expert opinion.

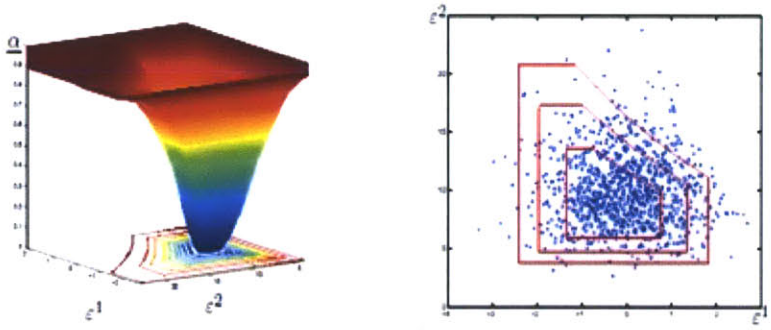

Figure 2-1: Cloud theory: uncertainty clouds in function of two metrics and of the confidence alpha. The shape of the cloud is determined by applying expert opinion. Source: [51].

This methodology represents the first attempt to statistically model the lack of knowledge in space system design. However, cloud theory has two main disadvantages. It is extremely complex and computationally expensive, and it is based only on expert opinion and not on historical data. The key concepts of the literature in the field of risk characterization can be summarized as follows:

- A consistent amount of effort has been put into categorizing risks and identifying

causes.

- Different qualitative models focused on improving human interactions across engineers have been developed.
- Quantitative models are few and generally lack statistical estimation of the fluctuations.
- The only attempt at developing a model similar to the one proposed in this thesis is complex, computationally expensive, and based on expert opinion only.

The work proposed in the thesis is different since it proposes a simpler statistical approach to quantify risks in mass and power using a combination of data and expert opinion. The next subsection is dedicated to a literature review in the field of statistical techniques.

2.3 Statistical Techniques

To implement a model able to quantify risks, different statistical techniques have been explored. Specifically, this survey is divided into three main categories:

- Density estimation: techniques to estimate the shape of a probability distribution given a set of data.
- Expert elicitation: techniques to express expert opinion in the form of a probability density function.
- Saddle-Point approximation: a technique to approximate the tail function of multiple independent random variables.

2.3.1 Density Estimation

Concerning density estimation, the literature lacks a comparison of different techniques for small sample sizes. As presented below, initial research started in the 1950's. The first and most intuitive technique is the Histogram, followed by the Naïve

Estimator documented in the work of Fix and Hodges [116], and Rosenblatt [108]. One of the most important contributions in the field is represented by the development of the Kernel Density Estimator by Parzen [102]. The Nearest Neighbor Estimator and the Variable Kernel Estimator are more recent methodologies. They are the results of the works of Loftsgaarden and Quesenberry [83] and of Breiman, Meisel, Purcel [23]. Scott [111] and Shawe [113] explore the multi-dimensional Kernel Density Estimator. Techniques for bandwidth selection are discussed by Jones [74]. Density estimation techniques are also interesting for the fact that they can identify multi-modal behaviors in data sets, as underlined by Izenman [65]. A useful description and presentation of all these methods can be found in the monograph of Silverman [115], where the principal methods are described and qualitatively compared. However, we are interested in a quantitative comparison across the different methodologies. Some work was developed by DasGupta [38] on an asymptotic scale. DasGupta analyzes the convergence of the different methods for a large number of samples. A gap in the current density estimation literature is represented by the absence of a statistical comparison of these methods for limited sample sizes. This absence is due to the fact that this analysis is not tractable in an analytical way, but it needs to be performed via numerical simulation as Silverman [115] underlines. This research tries to fill the gap by investigating different density estimation techniques, and by developing a test to compare the techniques for small sample sizes using known distributions as benchmarks.

2.3.2 Expert Elicitation

Regarding expert elicitation, different works were developed, but a quantitative methodology, for the case in which seed variables with known probability distributions are unavailable, is missing. As presented below, expert elicitation approaches can be organized [94] into two categories of methods: elicitation and aggregation. Elicitation methods are used to extract a probability density function from the inputs of an expert. Aggregation methods are used to compose the opinions of multiple experts into a unique estimation, and then into a unique probability function.

Regarding elicitation methods, an important piece of literature concerns the study of experts' ability to model probabilities. This field is also defined as "Mathematical Psychology," and it investigates the behavior of the brain when it is forced to perform quantitative probability assessments. Initial studies in this field were conducted in the 1970s by Tversky and Kahneman. They discovered ([130], [128], [127]) that when people think about probability they tend to rely on biases and heuristics, shortcuts that affect the way in which experts model probabilistic phenomena. Specific bias includes conjunction fallacy ([47], [129]), law of small numbers [127], base rate neglect [77], anchoring [138], and others [99]. An interesting review of biases and heuristics is presented by Garthwaite [53]. All these biases are explored in the interview process developed for this thesis.

Studying biases is important to assess how experts model probabilities, but to complete the measurement of experts' performance it is also necessary to assess the ability of experts to model quantities in their field of expertise. This process is called calibration. Calibration measures the agreement between subjective probabilities and real probabilities. Calibration has been studied by different authors. Lichtenstein [79] identifies a general problem of overconfidence in the estimates made by the experts. However, studies from authors like Wright [138] discover the opposite problem. Carlson's [26] research indicates that experts perform better in assessing probabilities of future events than of past events. Hagan [100] presents a review of the principal results in calibration. Calibration is normally performed by using seed variables for which the probability density function is already available. However, it is challenging to calibrate when no empirical data exist to compute the probability densities for the seed variables, as in the specific case of this research. In similar cases, authors Cooke and Hagan ([34], [100]) propose the use of previous expert elicitation information as probability densities for the seed variables. However, to the best knowledge of the author no previous expert elicitations in the field of design risk analysis for space communication system have been performed. Hence, this thesis will fill the gap of performing calibration when seed variables with correspondent probability densities are unavailable. The calibration proposed in this thesis is performed by point esti-

mation and it is discussed in Chapter 5.

In terms of techniques to generate the probability densities during the elicitation, two principal approaches are discussed in the literature: Bayesian analysis and statistical link analysis. Bayesian analysis ([18], [43], [78], [22], [48]) models expert opinion by using two probability functions: a priori and a posteriori. It is a preferable technique whenever there is already an initial knowledge of the uncertainty of the system, which means that the a priori distribution is known ([81], [16]). It is also a preferable technique in the cases in which new information is introduced into the system later. Even in the cases in which previous information is not available, Bayesian analysis can be applied by introducing non-informative priors or reference priors ([33], [17]). A method proposed to derive the a posteriori information from expert opinion is the quantile method described by Gelfand et al. [54]. In the context of the thesis, Bayesian methods are applied to update information from data statistics through expert opinion, when both sources of information (historical data and expert) are available. Similar approach which pair Bayesian analysis with density estimation are proposed by Escobar [45], Hagan [99], and Blattberg [20].

However, when very little data is available, and we have to rely on expert opinion only, it can be convenient to apply a technique which does not require modelling two different probability distributions. In this thesis, we employ a probability density elicitation approach similar to the one used in statistical link analysis. Statistical link analysis was developed by Yuen [75] in 1975 with the goal of providing a new statistical approach to model link budget uncertainties. Yuen's method gives an intuitive way to elicit expert opinion, which assumes that each expert has to guess the most adverse value and the most favorable value plus an idea of the confidence that he/she has in the distribution. This methodology was applied to link analysis, but not to design risks. However, Cheung [27] presents a refinement of statistical link analysis in a risk analysis perspective. It is interesting to notice that statistical link analysis presents some similarities with an expert opinion approach based on possibility theory discussed by Dubois [42].

Regarding aggregation methods, mathematical compositions have been intensively

studied by authors like Cooke [34] and Hagan [100]. Cooke developed the so-called Cooke Classical Model. Hagan's work ([100], [99]) describes different ways to exploit at maximum expert opinion using Bayesian analysis, elicitation processes, subjective probabilities, and experts cooperation. Lindley [82] presents a formal procedure to use expert opinion in reliability. Zhou, Zhu, and Tang [141] present an approach based on cloud model and evidence theory, which is useful in the case of experts' opinions that are changing over time. Clemen [32] reviews the principal techniques for expert aggregation: linear weighted combination, logarithmic combination, Cooke generalization, Bayesian approach, and Mendel-Sheridan composition. A comparison between linear and Bayesian compositions is proposed by Jacobs [66]. A combined statistical model is discussed by Martz and Bryson [88]. A copula approach to model dependence across experts' opinions is analyzed by Jouini et al. [93]. Jacques and Pidgeon [68] apply expert opinion to perform risk analysis in actuarial science. Goulet [56] presents an expert R software package which implements Cooke convex combination of experts opinion. A general critique on expert approaches for risk analysis is discussed in [94]. The principal approach applied in this thesis to aggregate multiple expert opinions is a linear combination similar to the one proposed by Cooke [34]. However, our method is different from Cooke's in how calibration is performed and how the aggregation weights are computed (more details in Chapter 5).

2.3.3 Saddle-Point Approximation

Saddle-Point approximation is useful in a lot of situations. Some details can be found in the work of DasGupta [38], Daniels [37], Reid [104], and Cox [12]. Cheung [28] presents an interesting application of this methodology to the problem of estimating the amount of data produced by instruments in a space probe for planetary exploration. The core of Saddle-Point approximation is the possibility of using the moment generating function to obtain an approximation of the tail of a distribution. The computational advantage is consistent, since Saddle-Point allows computation in the transformation domain by substituting convolutions with less computationally expensive products. It can be argued that in some cases convolutions are not neces-

sary, since if the number of distributions to convolve is large enough, then the Central Limit Theorem can be used as a very good approximation. However, in our model the number of distributions to convolve will be in many cases much less than 30, which is the rule of thumb number for which the application of the Central Limit Theorem is appropriate. However, when distributions are less than 30, the calculation can still be computationally consuming, and the possibility of using Saddle-Point approximation becomes extremely important. In the context of our work, Saddle-Point estimation is used in two contexts. First, it is studied as a possible density estimation technique through the construction of the characteristic function using sample moments. Second, Saddle-Point approximation is used to reduce the computational cost of the model every time the convolution of different probability densities is required.

2.3.4 Summary

The key concepts of the literature in the field of statistical techniques can be summarized as follows:

- Density estimation: different techniques have been developed, but a comparison based on small sample size is missing.

- Expert opinion approaches: a lot of work has been performed to categorize biases and heuristics, but a quantitative process to map them in the elicitation process is missing. Also, previous studies in calibration did not address the problem of performing calibration when seed variables do not have a correspondent probability density. Finally, statistical link analysis has never been applied as a possible approach to model the probability densities assessed by the experts.

- Saddle-Point approximation: some work has been done in the application of this technique. However, in this thesis the approximation will be applied for the first time as a density estimation technique.

The next section is dedicated to the literature review in the field of modelling and optimization.

2.4 Modelling and Optimization for Communication Systems

This section of the literature presents works in the field of communication system modelling and optimization to identify developed methods/models similar to the one proposed for this research. Literature in communication system modelling and design includes works of Maral and Bousquet [87], Richaria [106], Evans [46], Brown [24], Wertz [135], and Gilchriest [55]. All these authors are focused on identifying average design values for the communication system, and on allocating contingencies. However, contingencies are computed as a fixed percentage of the design values without a statistical assessment of the real probability for these values to fluctuate. The work of this thesis aims to fill this gap by developing a model to size the communication system while assessing design risk.

When the model is developed, it is necessary to validate it. Balci [11] presents an interesting overview of different possible validation techniques. Specifically, in our case the *predictive* nature of the model (which aims to characterize risks) indicates that a *predictive validation technique* is required. This technique uses past system input data to generate forecasts which are compared with the correspondent past system output data. This validation technique has been applied to the test cases discussed in Chapter 6.

The field of optimization has known great progress since the 1990s through the development of the Multi-disciplinary System Design Optimization (MSDO) [3] methodology. An interesting survey of the different aspects and developments of MSDO is the one presented by Sobieski and Haftka [117]. Additional aspects of modelling and

of MSDO can be found in the work of Palambros [101]. Another promising work in the field of MSDO and Robust Optimization is the one of Bertsimas [19]. Bertsimas identifies the optimal solution for any realization of the uncertainty in a given data set. Bertsimas' s approach is different from the one developed in this thesis, because in this thesis the uncertainty is characterized through a set of statistics.

The application of MSDO methodologies to space system design is relatively recent. The work of Jilla and Miller presents a first attempt to apply MSDO to the design of distributed satellite systems ([70], [72], [69], [71]). Additional work has been done by De Weck and Miller ([132], [133]) in the design of Nexus spacecraft as an MSDO problem, and in the development of the concept of Iso-performance methodology. However, limited work has been done in applying MSDO to the characterization of design risks. De Neufville and De Weck ([134], [97]) analyze the problem from a reconfigurability/flexibility point of view, which is different from a statistical characterization when the architecture is not flexible. Hassan, Crossley, Pullen, and Parkinson ([61], [60], [103]) characterize the uncertainty in terms of reliability, and not in terms of design risk. The work of this research aims to fill a gap by incorporating statistical risk estimation in MSDO methodology.

Finally, an example of modelling and optimization similar to the one presented in this thesis is explained in the work of Cheung ([30], [29]) for risk analysis in mission planning and sequencing. The mathematical framework is similar to the one proposed in this thesis; however the risk characterization proposed by Cheung refers to mission operation, while this research is focused on design risk.

The key concepts of the literature in the field of modelling and optimization for communication systems can be summarized as follows:

- Current models of communication systems do not include a statistical assessment of design risk. This thesis fills this gap by pairing traditional communication system design tools with a risk estimation model.

- Previous works in MSDO do not integrate optimization with statistical risk estimation. In this thesis, a mathematical formulation to include risk estimation

in the optimization framework is proposed.

A summary of the literature review is presented in the next section.

2.5 Summary

The previous three categories –risk analysis and characterization, statistical techniques, modelling and optimization for communication systems– represent the bulk of the literature relevant to this work. This research will draw upon pieces of each of these areas to develop a statistical estimation model to perform design risk estimation. While each of the fields contains a great deal of work, there are a few aspects lacking:

- **Quantitative models for design risk**: The majority of work in this field is composed of qualitative assessments. Only a few works are dedicated to quantitative assessments. A methodology based on integration between expert opinion and data is lacking.

- **Statistical techniques to perform estimation**: A consistent amount of literature discusses different density estimation techniques, and different approaches to model expert opinion. Nevertheless, a performance analysis for small sample sets of data, a mapping of experts' biases, an expert approach which redefines calibration, an elicitation based on statistical link analysis, and the use of Saddle-Point approximation as a density estimation technique are missing.

- **Inclusion of risk analysis in modelling and optimization for communication systems**: Current communication system design models do not include statistical risk estimation. Also integration between statistical risk estimation and MSDO methodology is lacking.

The work proposed here will fill in these gaps to advance the state of the art of statistical risk estimation and design for communication systems.

Chapter 3

Approach Overview

The objective of the thesis is to develop a mathematical approach to quantify the likelihood that a communication system will meet the spacecraft and mission constraints as the design matures. This chapter presents an overview of the approach developed to accomplish this objective.

3.1 Overview

The approach is summarized by the block diagram in Figure 3-1. Figure 3-1 shows the main modules that compose the approach. Baseline design generates a parametric model of a communication system. Expert and data statistics are used to compute risks for the risk module. The optimization module identifies the best architectural solutions across alternatives.

A short description of the modules is as follows:

- **Baseline design:** This module uses traditional analytical models/equations (link analysis, coverage) to generate an initial point estimation of a communication system. Baseline design receives as inputs specific communication parameters (channels, frequencies, signal quality, and receiver structure) and elaborates vectors of possible baseline communication architectures that include lists of components and correspondent design values for gain, mass, and power. The baseline design emulates what it is done in the initial design stage before

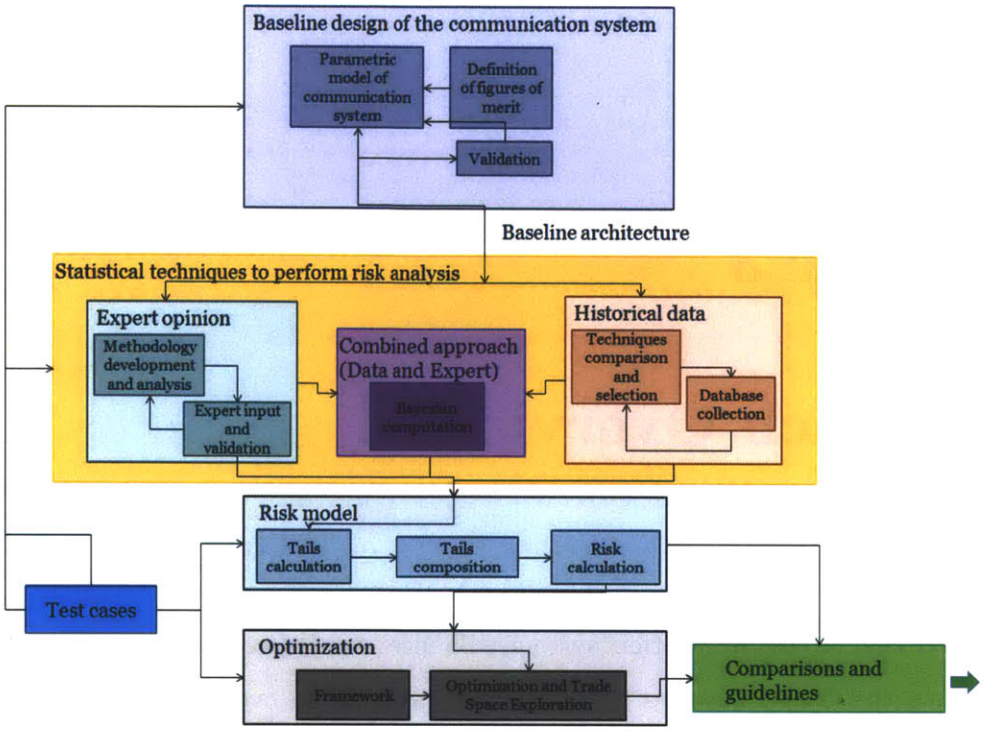

Figure 3-1: Approach overview. Baseline design generates a parametric model of a communication system. Expert statistics, data statistics, and Bayesian statistics are used to compute risks at the component level. Risks are convolved at the system level, and the model performs optimization to identify the best architectural solution across alternatives. Test cases are used in different parts of the model as a validation tool. Comparisons with traditional approaches and guidelines on communication system design risk analysis are outcomes of the approach.

performing risk analysis or before applying contingencies: it creates an initial estimation of the design. The output of baseline design is used to quantify risks in the risk model. The output of baseline design does not necessarily need to be correct: it is an initial estimation which can eventually be corrected by the risk model through the assessment of the probabilities that the predicted values of mass and power consumption will fluctuate across the design's evolution.

- **Historical data approach:** One way to quantify components' fluctuations in mass and power is to use previous samples of the same components to develop a statistical analysis. The use of historical data requires: an identification of the most appropriate statistical technique to use, a determination of the number

of samples required, the construction of a database, the shaping of probability distributions, and the validation of the approach.

- **Expert elicitation approach:** If samples are not available, an approach based on expert opinion is developed. Expert elicitation requires: a technique to measure biases and to calibrate expert opinion, a methodology to elicit probability distributions, and an approach to aggregate different experts' opinions in a unique estimate.

- **Bayesian approach:** a Bayesian combination of the previous two sources of information into an unique probability density.

- **Risk model:** The outputs of baseline design are design values for power and mass for each component. These values are used together with data and expert statistics to compute the risks of exceeding caps in mass and power. Results are presented using tail functions (probability of exceeding a certain value) for each component.

- **Optimization:** The risk model can be used to explore different architectural solutions and to optimize the design. For this part of the approach it is necessary to develop a mathematical optimization framework to capture all constraints of the system.

- **Test case analysis and guidelines:** The approach needs to be tested on test cases. Test cases are used to compare the traditional approach in risk evaluation with the risk model proposed in this thesis. The comparison is used to demonstrate how statistical risk estimation can improve the state of the art in design of communication systems.

3.2 Baseline Design

The baseline design is the initial module of risk estimation. The objective of this module is to provide an initial estimation of mass and power consumption for a

communication system. Baseline design uses standard analytical equations and fitting curves to model a communication system by calculating coverage, link analysis, mass and power consumption. Baseline design reflects the state of the art in communication system design. The outputs of baseline design are sets of possible space communication architectures that satisfy specific requirements. Baseline design is used whenever an initial PDR level design is not available, or whenever the designer wants to generate different possible architectures to compare. However, if an initial design is already available, this step can be skipped, or the baseline design can be used as a validation/comparison tool.

As previously mentioned, baseline design is derived from similar communication system design methodologies (see Maral and Bousquet [87], Richaria [106], Evans [46], Brown [24], Wertz [135], Gilchriest [55]). However, one difference is that it is paired with a coverage software tool to calculate the minimum transmission rate required. In this way, the data rate selected is the minimum to accomplish a certain mission. The baseline design is organized in three submodules: coverage, link analysis, and average mass and power calculator. More details on each of these submodules are given in Chapter 6.

The inputs for the baseline design are mentioned in the following list, but they are described in Chapter 6: number of communication channels, central frequency for each channel, level of redundancy, receiver characteristics, orbit parameters, simulation time, quality requirement, mission data requirement, and link analysis margin.

The baseline design performs computations for each communication channel to identify the feasibility of the system (similar to [9]) defined as the ability of the system to transmit the amount of data required in a specified interval of time. If the system is feasible, the model identifies a set of iso-EIRP (Equivalent Isotropic Radiated Power) solutions. Each solution corresponds to a different combination of amplifiers, transceivers and antennas, which achieves the same EIRP. Any architecture represents a collection of components which can fluctuate in mass and power. The risk of exceeding caps is computed in the risk model. The baseline design is validated using data from current missions. Specifically, the link analysis submodule is validated by

comparing the EIRP calculated by the model, and the one obtained from technical documentation of different missions. A similar validation is performed for the other submodules by comparing the values of mass and power computed by the model, and the real values obtained from technical documentation. Results are given in Chapter 6.

Baseline design generates the initial condition for the computation of design risks. However, the correctness of baseline design is not critical for the performance of the entire statistical risk estimation model. In fact, the goal of the risk model is to identify whether or not the initial design is likely to exceed expected values of mass and power consumption. Hence, if the estimation created by the baseline design is an overestimation or an underestimation, the risk model will reveal it.

3.3 Historical Data Statistics

The statistics based on data use historical mission data to estimate mass and power distributions of a component. A pre-requisite to construct a database is the existence of sufficient samples to generate representative statistics for the quantities of interest. This problem can be challenging as components for space communication systems have low market demands, and are thus produced in low quantities. Traditional probability density estimation techniques are divided into two main categories: parametric and non-parametric estimations. In the case of this research, it is difficult to estimate the shape of the probability distribution as many factors affect the process. For example, if we consider the probability distribution of the mass of a component, it is influenced by functionality of the component, materials used, fabrication processes, etc. It is difficult to model all the effects with governing equations; hence a parametric distribution cannot always be assumed. It is therefore necessary to model data through non-parametric density estimation.

Non-parametric density estimation techniques process observed data to construct an estimate of the underlying probability density function [115]. Many of these techniques can be shown to converge to the true density function when the sample size

approaches infinity [38]. However, there are few works in the literature that assess the usefulness of these techniques when the sample size is small, as is the case for components of spacecraft communication system. In this thesis, a numerical analysis to investigate the convergence rates of non-parametric density estimation functions is developed. The analysis is performed for small sample sizes between 10 and 100. The density estimation techniques compared are: Histogram, Naïve Estimator, Kernel Density Estimator (KDE), Nearest Neighbor Estimator, Variable Kernel Estimator, and Saddle-Point Estimator. To compare the goodness of the chosen density estimation techniques, benchmark distributions are applied. The comparison metrics are defined in Chapter 4, and they represent a measure of the distance between estimated tails (calculated using the different density estimation techniques) and real tails (computed using benchmarks). The results of the analysis are given in Chapter 4, and they show that the Kernel Density Estimator technique achieves the lower bound in estimation error for all cases analyzed. For this reason, the Kernel Density Estimator is selected to model data statistics.

The database for the thesis has been developed using articles on design of components for communication systems, and NASA JPL's publications. The database is parameterized such that each sample collected is converted into a coefficient that represents a relation between two parameters (example: mass fractions per unit of gain). More details on the probability density generation process and on the way in which the approach is adapted to compensate for non-linearity in the data are in Chapter 4.

The next section presents key aspects of the expert elicitation statistics described in Chapter 5.

3.4 Expert Elicitation Statistics

In the area of spacecraft communication system design, there are experts in the field, who can provide subjective assessments of mass and power of future components based on their experience and understanding of the market trends, product availability, and technology readiness. In the cases in which the required sample size to apply database

techniques is missing, the statistician can rely on expert elicitation techniques. Expert elicitation can be viewed as a form of external knowledge that is introduced in the model, with the objective of substituting or improving the knowledge given by historical data. However, as Hagan [100] points out, expertise also involves how the expert organizes and uses the knowledge. In fact, to have an effective elicitation of an expert's knowledge, the expert has to be able to express his or her uncertainty of the knowledge accurately. In order to properly model expert opinion, it is necessary to:

- Identify biases that the expert may have in expressing his or her knowledge and the corresponding uncertainty. Biases can lead to systematic errors in the way in which a person models probabilities.

- Identify the level of adjustment required to calibrate the expert opinion. Calibration is important to measure the ability of the expert to assess quantities in his/her field of expertise.

- Elicit the probability densities from experts. This last part involves the selection of appropriate statistical quantities that can accurately express the belief and the uncertainty of the experts. They can be: bounds, mean, variance, quantiles, etc.

To accomplish the three above steps, a three-part interview is developed in this thesis. Part 1 of the interview is called Probabilistic Thinking. This part contains general questions on probability that are used to model the expert's ability to think probabilistically. The result of this part is the generation of a quality index used to compare and to weight opinions from different experts. Probabilistic Thinking is focused on mapping the possible biases and heuristics that experts may have. The biases mapped are the following: hindsight bias, small sample bias, judgment by availability, judgment by representativeness, awareness of underlying conditional probabilities, anchoring and adjustment, and coherence.

Part 2 of the interview is called Calibration. It contains technical questions on mass

and power for communication system components. Experts are required to estimate mass and power for typical components, and their answers are used to identify if the expert has the tendency to underestimate or overestimate certain quantities. This part results in a calibration coefficient used to shift the probability assessments performed by the experts, and in a calibration score used to weight opinions from different experts.

The last part of the interview is the Elicitation. It aims at helping the expert to express his or her belief on the mass and power values for certain components and their corresponding uncertainties, and to translate the experts' estimates into probability density functions. Two different approaches are chosen to construct the density function based on the expert's belief: an approach derived from statistical link analysis, and the quantile method. The approach derived from statistical link analysis [75] first provides the initial design value of the component as the starting point. The expert will then assess three quantities: lower bound, upper bound, and the form of the distribution (uniform, triangular, or normal). In the quantile method, the expert assesses the 50% quantile, the 16% quantile, and the form of the distribution (uniform, triangular, or normal). A recursive tool allows the expert to visually check the shape of the distribution that results from the elicitation process, and to modify it until it properly expresses his or her belief.

The interview was performed on three different populations: MIT undergraduate students, MIT graduate students, and JPL engineers. The students are not experts in the design of spacecraft communication systems; hence they were tested only for Part 1 (Probabilistic Thinking). The JPL engineers participated in all three parts of the interview. Results indicate that the composition of multiple experts' opinions weighted with the scores computed from Probabilistic Thinking and Calibration improves the quality of the estimations with respect to the case of experts' probability densities weighted equally.

Finally, to combine data and expert approaches, a Bayesian composition of probability densities using expert and data information is discussed in Chapter 5. Test case validation shows that the Bayesian combined approach performs better than the

approaches based only on data or only on expert. The complete description of the expert methodology is discussed in Chapter 5, while the full set of questions used for the interviews is included in Appendix B.

3.5 Risk Model

The input of risk model is a given design (computed through baseline design), and the output is the probability that the mass or power would exceed the given allocations. This probability is computed using the approaches previously presented: an approach based on data, an approach based on expert opinion, and an approach based on data and expert opinion combined.

In the case of the data approach, Kernel Density Estimator is used to compute the probability distributions. Currently, the size of the database is limited to no more than 40 samples for each category of components, and samples tend to take on a wide range of values. The distributions constructed are inclined to have a larger dispersion. This problem is discussed in the test cases (Chapter 6), and it is one reason for which the use of expert opinion is strongly encouraged. In the case of expert elicitation, the accuracy of experts' judgments is difficult to measure, but the three-part interview with quality estimation (Part 1) and calibration analysis (Part 2) are used to quantify experts' reliability. A problem with expert statistics is that expert estimation can be in some cases conservative in the risk assessments. This problem is also discussed in Chapter 6.

Hence, as shown in Chapter 6, the Bayesian approach is the most promising since it inherits the positive aspects of the other two statistics. Bayesian estimates are less conservative than expert estimates: similarly to data estimate, Bayesian probability densities are centered in the real values of mass and power of the components tested. However, Bayesian estimates show a smaller dispersion than data estimates, since they take advantage from the knowledge of the experts.

Figure 3-2 is a graphical example of the comparison of two architectures on the bases of risk. For each of the two architectures the tail function is plotted. The tail function

represents the probability for the architecture to exceed a certain value in a certain metric. The green solid line represents the cap which cannot be exceeded. The corresponding risks of exceeding the cap, for each of the two architectures, are given by the intersection of the solid line and each of the tail functions.

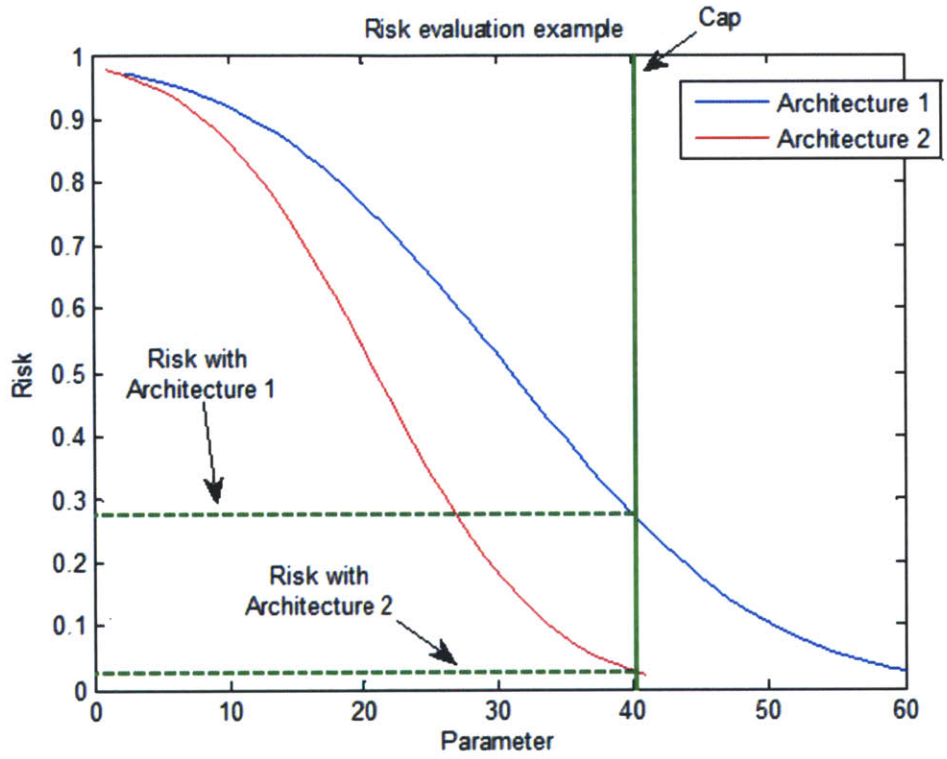

Figure 3-2: Comparison of two architectures on the basis of risk. The tail function represents the probability for the architecture of exceeding a certain value in a certain metric. The correspondent risks to exceed the cap for each of the two architectures are given by the intersection of the solid green line and each of the tail functions.

To obtain the output pictured in Figure 3-2, the model takes as inputs the outputs of the baseline design: components' design values in mass and power. Then, it computes the risk of exceeding mass and power constraints by applying the different statistical approaches. Finally, risks at the system level (defined as the sum of risks of different components) are computed by composing multiple distributions under the hypothesis of independence. Hence, the total probability density function is the

convolution of the single densities, as shown in Equation 3.1.

$$f_{tot}(x) = f_{mass_1}(x) * f_{mass_2}(x) * \ldots * f_{mass_n}(x) \tag{3.1}$$

The complete description of the model and of the test cases is presented in Chapter 6. The next section summarizes the optimization module developed for the thesis.

3.6 Optimization

As shown in Figure 3-1, the optimization module aims to explore multiple architectural solutions, to select the one that maximizes certain objectives while maintaining design risks below certain thresholds. The inputs of this module are the architectures calculated in the baseline design, and the risk assessment computed by the risk model. Some basic characteristics of the module are the following:

- Variables: communication architectures. Any architecture is composed of a list of components, described in mass, power, and gain.
- Objectives: minimize mass and power consumption.
- Constraints:
  - Quality constraint for each of the channels: The sum of gain for each of the components has to be greater than the minimum EIRP required (as determined by the baseline model).
  - Cap constraints: The sum of the design values of mass and power cannot exceed caps.
  - Risks constraints: The probability for the whole system (including fluctuations) of exceeding caps in mass and power has to be less than a certain percentage.

To perform the optimization, the non-linear nature of the problem requires the use of non-linear solvers. Finally, the set of architectures is visualized to identify the Pareto front of solutions which maximize/minimize the optimization objectives. The inclusion of risk analysis in the optimization is used to underline how risk analysis can change the Pareto front of optimal solutions as discussed in Chapter 7. The next section lists test cases used to validate the model.

3.7 Test Cases

Test cases are examples of real mission designs for communication systems, for which the initial mass (at PDR) and the final mass (at CDR) are known. The model analyzes the designs by comparing risk evaluation performed by the model with the traditional approach based on contingencies. The objective of the comparison is the validation of the model, as well as the determination of guidelines in applying statistical risk estimation to the design of communication systems.
A challenge regarding this part of the thesis concerns the availability and the accessibility of data. Therefore, test cases are mostly spacecraft developed in universities where fewer restrictions apply in terms of data accessibility. The missions used as test cases are as follows:

- **CASTOR (Cathode Anode Satellite Thruster for Orbital Reconfiguration):** CASTOR is a small size (ESPA class, 50 Kg) spacecraft designed at MIT Space System Laboratory [1]. Hence, technical documentation of the fluctuations for the different components at any point of the design is completely available. The analysis performed shows that the application of statistical risk estimation in the initial stage of design of this mission (before PDR) would have identified overestimations and/or underestimations of the mass and power for the different components (more details in Chapter 6).

- **HETE-2 (High Energy Transient Explorer):** HETE-2 is a medium size (100 Kg) spacecraft developed by MIT in cooperation with other institutions [131],

to detect cosmic gamma-ray bursts (GRBs) to determine their origin and nature. HETE-2 was launched and successfully operated during the 1990s. The design data (PDR and CDR) are available. The analysis performed shows that statistical risk estimation (especially through expert opinion) would have helped to identify overestimations and underestimations, and would have reduced design risk for the system (more details in Chapter 6).

- **CASSINI-HUYGENS:** CASSINI-HUYGENS is a big size (2,523 Kg) spacecraft developed to study the planet Saturn and its satellites. CASSINI-HUYGENS launched in 1997 and it is still operative today. This mission has been selected for this thesis, because it is a space mission designed to reach a very far point in the solar system, and as a consequence it is equipped with high energy components. Hence, this test case is very different from the previous two analyzed. Unfortunately, the unavailability of expert elicitation data for this test case has limited the analysis. For this reason, CASSINI-HUYGENS is discussed in Appendix D.

3.8 Model Extensions and Generalizations

Model extension is an effort to expand the methodology to different uncertainty metrics other than power and mass. Specifically, cost as a metric for design risk analysis is discussed in Chapter 8. Differently, model generalization is an attempt to show applications of the previously described approach to other subsystems as well as to the entire spacecraft. Generalization examples are presented in Chapter 8.

3.9 Summary

Chapter 3 presents an overview of the thesis's approach. The following chapters detail the modules here introduced. Specifically, baseline design is described in Chapter 6. Data statistics are presented in Chapter 4, while the expert elicitation approach and

the Bayesian combination are in Chapter 5. The risk model is described in Chapter 6, and the optimization is detailed in Chapter 7. Test cases are mostly discussed in Chapter 6, but they are also used to validate some aspects of data and expert statistics (Chapters 4 and 5). Model extensions and generalizations are presented in Chapter 8.

Chapter 4

Historical Data Statistics

Design risk analysis for space communication systems is the process of estimating the risk that certain architectures will exceed constraints in mass and power over the design's evolution. To perform such estimations different techniques can be implemented: data statistics, or expert statistics. If data statistics are used, no a priori assumption on the shape of the probability distributions can be made. Consequently, non-parametric density estimation is required. However, selecting the most suitable density estimation technique for the analysis is a challenge due to the limited number of samples available.

Chapter 4 presents the methodology to model probability densities for mass and power fluctuations using historical data information. A comparison across different density estimation techniques is described, and the results show that the Kernel Density Estimator achieves the lowest estimation error. Additionally, the chapter describes the organization of the database, and the process followed to generate the probability densities for the cases of interest.

4.1 Motivation

As mentioned in Chapter 3, a data approach to design risk is the use of historical mission data to quantify/predict the likelihood that a component will exceed caps in mass and power. Historical mission data are considered sources of information for

this research, due to:

- Heritage: Many missions inherit their components from previous missions. Hence, CDR values of mass and power for components used in the past can help to predict mass and power values for new components.

- Independence from human opinions: An approach which depends as much as possible on data and not on human opinions/experience can help to automate the design process. However, in some cases, expert opinion is necessary as explained in Chapter 5.

One of the key aspects of a model based on data is the number of samples required to infer reliable results. This depends on the estimation technique applied. Traditionally, estimation techniques are divided into parametric and non-parametric estimations. Regarding parametric estimation ([18], [43]) techniques, the shape of a probability distribution is assumed, and data are used to compute the parameters of the distribution. Parametric estimation is performed using different techniques which include the Maximum Likelihood Estimator [18] and the Method of Moment Estimator [43]. However, the fundamental assumption of parametric estimation is the knowledge of the distribution's shape, which is the main disadvantage for this kind of technique, and it represents the motivation for the use of non-parametric techniques in this research.

Regarding non-parametric estimation (or density estimation), the shape of the distribution is unknown, and it becomes part of the estimation problem. Data are used to infer not only the parameters, but the shape of the overall distribution as well. These techniques are more recently developed than classical parametric estimation, and they involve different challenges. They converge slower than parametric estimations [38], and their performance can vary strongly for different problems [115]. However, they do not require any assumption on the shape of the distribution; hence, they are suitable for this research. In fact, in this research, it is difficult to assume the shape of a distribution, since multiple factors affect the components' development process.

In Chapter 2, literature works in the field of density estimation are described. In the

context of this research, we are interested in a quantitative comparison across the different methodologies. Some work was done by DasGupta [38] on an asymptotic scale. However, a statistical comparison of the methods for a limited number of samples is missing. This comparison is developed in this chapter.

4.2 Comparison of Different Density Estimation Techniques

Density estimation techniques "aim to construct an estimate, based on observed data, of an unobservable underlying probability density function [115]." Different techniques in this field have been developed, but the identification of the best techniques to use is still a challenge.

Research ([115], [111], [65]) in this field obtained conflicting results. The main reason is that the performance of the techniques depends on the estimation problem considered, and on the amount of information available. In the field of design risk for space communication systems, a comparison of the performance across these techniques is lacking, especially for small sample sizes.

Hence, the following comparison aims to characterize the behavior of the different techniques as the sample sets become small. To perform this study, a numerical analysis is required (as pointed out by [115]), and it is performed using benchmark distributions.

The following subsections describe the statistical techniques compared, the numerical analysis, and the results.

4.2.1 Review of Density Estimation Techniques

The density estimation techniques compared are: Histogram, Naïve Estimator, Kernel Density Estimator, Nearest Neighbor Estimator, Variable Kernel Estimator, and Saddle-Point Estimator. In each of the equations that describe the techniques, x

indicates the support of the distributions: the set of values on which the PDF is computed. In the specific case of this thesis, x is generally a continuous set of mass values (measured in Kg), or power values (measured in W). n indicates the number of samples.

Histogram

This technique is the oldest among the different density estimators. The definition, according to Silverman [115], is the following: given an origin x_0, and a bin width h, the bins of the histogram are the intervals $[x_0 + m \cdot h, x_0 + (m+1) \cdot h]$, where m is a positive integer. The expression is as follows:

$$\hat{f}(x) = \frac{1}{n \cdot h} \cdot \tilde{x_i} \tag{4.1}$$

$\tilde{x_i}$ indicates the number of x_i in each of the bins. This method presents two problems. First, the dependence of the estimate on the selection of origin and bin width causes changes in the shape of the distribution. Second, the discontinuity of the technique causes problems whenever the estimation is used as part of a continuous model.

Naïve Estimator

According to [115], if h is the bandwidth of the estimator, x is the support of the distribution, and $P(x)$ the probability that sample X belongs to an interval of size h centered in x, then the estimator is:

$$\hat{f}(x) = \lim_{h \to 0} \frac{1}{2 \cdot h} \cdot P(x - h < X < x + h) \tag{4.2}$$

The Naïve Estimator can be considered "a histogram where every point is the center of a sampling interval, thus freeing the histogram from a particular choice of bin positions [115]." The choice of bin width still remains governed by the parameter h, which controls the amount by which the data are smoothed to produce the estimate.

Kernel Density Estimator

$K(\cdot)$ is a generic and symmetric probability function that represents the kernel. According to [115], the Kernel Density Estimator is defined as:

$$\hat{f}(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h}\cdot K\left(\frac{x - X_i}{h}\right) \tag{4.3}$$

where h is the bandwidth, the amplitude of each kernel function around the sample X_i. The Kernel Density Estimator is a sum of functions centered on each of the samples. Figure 4-1 describes the process of construction of a kernel estimator.

The bandwidth selected is the optimal in the L2 sense [38] for the normal distri-

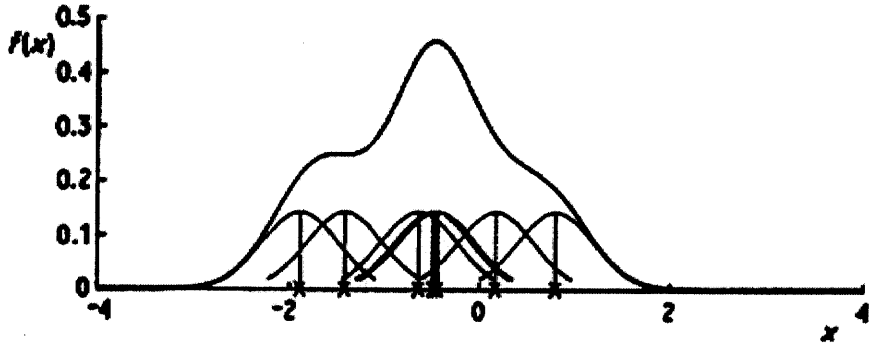

Figure 4-1: Process of construction of Kernel Density Estimator. A Kernel Density Estimator is a sum of functions centered on each of the samples. Source: [115].

butions. Kernel Density Estimator generates a well defined probability density: it is everywhere non-negative, and it integrates to 1. Also, the kernel density inherits all the continuity and differentiability properties of the kernels from which it derives. Different kernels can be used to generate the estimation. In this analysis, four kernels have been used: uniform, rectangular, normal, and Epanechnikov (a special kernel which is designed to minimize the Mean Square Error [115]). This technique is still sensitive to the fact that the bandwidth h is fixed. The following two techniques are examples of density estimations with variable bandwidth.

Nearest Neighbor Estimator

The Nearest Neighbor Estimator [115] adapts the amount of smoothing to the local density of the data. The level of smoothing is determined by an integer k (smaller than the sample size, generally $k \approx \sqrt{n}$). Defining the distance between two points as the Euclidean distance, then, the kth nearest neighbor is defined as follow:

$$\hat{f}(x) = \frac{k}{2 \cdot n \cdot d_k(x)} \tag{4.4}$$

It is possible to generalize the nearest neighbor as a Kernel Density Estimator:

$$\hat{f}(x) = \frac{1}{n \cdot d_k(x)} \sum_{i=1}^{n} K\left(\frac{x - X_i}{d_k(x)}\right) \tag{4.5}$$

It can be observed that the nearest neighbor method is the Kernel Density Estimator evaluated at x with window width $d_k(x)$. The overall smoothing is governed by the choice of the integer k, but the window width used at any particular point depends on the density of observations near that point. This approach has multiple drawbacks: the estimate is not a smooth curve, and it is a continuous technique but its derivative is not. Finally, Silverman [115] showed that the integral of the distribution tends to infinity and not to one.

Variable Kernel Estimator

This method, like the Nearest Neighbor, adapts smoothing to the local density of data. The estimate is constructed similarly to the classical Kernel Density Estimator, but the width varies from one data point to another. Let $K(\cdot)$ be a kernel function and k a positive integer. Define $d_{j,k}$ to be the distance from X_j to the kth nearest point in the set comprising the other $n - 1$ data points. The Variable Kernel Estimator [115] is defined as (h is a smoothing parameter in this case):

$$\hat{f}(x) = \frac{1}{n} \sum_{j=1}^{n} \frac{1}{h \cdot d_{j,k}} \cdot K\left(\frac{x - X_j}{h \cdot d_{j,k}}\right) \tag{4.6}$$

The window width of the kernel is proportional to $d_{j,k}$, so that data points in regions where the data are sparse will have flatter kernels associated with them. For any fixed k, the overall degree of smoothing will depend on h. The choice of k determines how responsive the window width choice will be to very local details.

Saddle-Point Estimator

This technique is discussed in Cheung [28] and also DasGupta [38]. The Saddle-Point estimator does not estimate the shape of the distribution, but the tail of the distribution only. The concept is to approximate the tail of a distribution by estimating the components in the transform domain –the characteristic function. In this case, instead of using the true characteristic function for the distribution, which is not known, an approximation of the characteristic function is constructed from the samples. Using Taylor Series expansion, the characteristic function can be regarded as a collection of moments:

$$\Psi_z(s) = 1 + \frac{m_1}{1!} \cdot s + \frac{m_2}{2!} \cdot s^2 + \frac{m_3}{3!} \cdot s^3 + \ldots + \frac{m_n}{n!} \cdot s^n + R_n(s) \tag{4.7}$$

where $m_1, m_2, \ldots, m_n$ are the first, second, n-moment of the distribution. In the case of this analysis, the exact moments will be substituted by the sample moments, namely:

$$\hat{m_k} = \frac{\sum_{i=1}^{N} \left(X_i^k\right)}{N} \tag{4.8}$$

N is the number of samples. The approximated transform $\hat{\Psi}_z(s)$ can be computed as:

$$\hat{\Psi}_z(s) = 1 + \frac{\hat{m_1}}{1!} \cdot s + \frac{\hat{m_2}}{2!} \cdot s^2 + \frac{\hat{m_3}}{3!} \cdot s^3 + \ldots + \frac{\hat{m_n}}{n!} \cdot s^n + R_n(s) \tag{4.9}$$

Using the approximated transform $\hat{\Psi}_z(s)$, the function $\hat{\Psi}(s)$ is defined as:

$$\hat{\Psi}(s) = -s \cdot \alpha + \log \hat{\Psi}_z(s) - \log s \tag{4.10}$$

The value s_0 represents the minimum of $\hat{\Psi}(s)$. Once that s_0 is computed, the tail approximated by using Saddle-Point results:

$$T_+(\alpha) \approx \frac{e^{\hat{\Psi}(s_0)}}{\sqrt{2 \cdot \pi \cdot \hat{\Psi}''(s_0)}} \tag{4.11}$$

The main problem with this technique is that it sometimes requires many sample moment terms (more than 40) for the tail probability to converge, and high-order sample moments require a large number of samples to be accurate.

Detailed descriptions of the different techniques can be found in [115], [28]. In the next subsection, the numerical analysis developed to compare the different techniques is described.

4.2.2 Numerical Analysis

As mentioned, the objective of this numerical analysis is to assess the quality of the different density estimation techniques for small sample sizes. The challenge in performing this assessment is that the distribution from which data originate is unknown. Hence, it is difficult to assess how far the estimated distributions are from the real distribution. To alleviate this problem, samples from known distributions are used. Figure 4-2 summarizes the methodology followed. The model uses as inputs

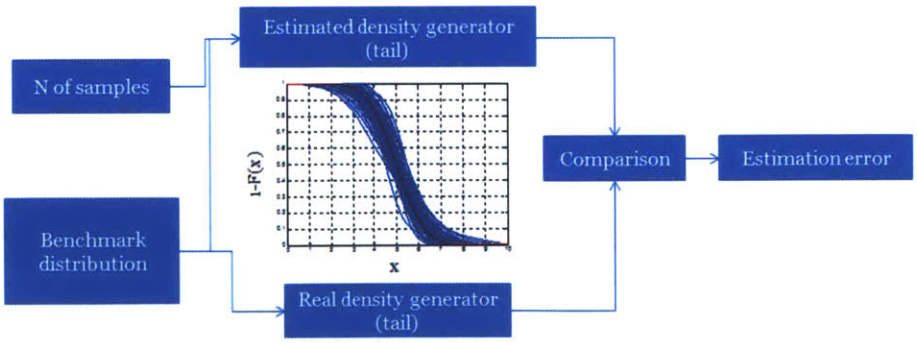

Figure 4-2: Block diagram of the test. Benchmarks are sampled, and the samples are used to feed the density estimation techniques. Each technique generates an estimated tail, which is compared with the real tail, mathematically computed using the benchmark.

the number of samples, and a benchmark distribution. The number of samples is a quantity motivated directly by the nature of the numerical analysis itself. Since the goal is to analyze the behavior of different density estimation techniques for small sample sizes, the number of samples will be small: from a minimum of 10 up to a maximum of 100 samples (with increments of 10 samples for subsequent iteration). The benchmark distribution is a known distribution used to measure the quality of the estimator. Common distributions generally used to describe random phenomena have been selected as possible benchmarks, and they are listed as follows:

- Normal distribution: It is generally used to model aleatory phenomena in different branches of science.
- Lognormal distribution: In space system design estimated mass and power consumption tends to increase rather than to decrease. Hence, the asymmetric shape of the lognormal is particular suitable for this analysis.
- Beta distribution: It can be used to model non-informative distribution, such as the uniform distribution.
- Gamma distribution: It can model very different phenomena just by a change in the parameters.
- Exponential distribution: Like the lognormal, it presents an asymmetric profile.

The number of samples n and the benchmark distribution used (with its parameters) are the inputs of the calculation. The model samples randomly the benchmark to extract n samples. This sampling operation is repeated multiple times (k times generally 100) to stabilize the numerical analysis on a consistent number of trials. The model collects a set of a total of k vectors of n samples. Each vector is elaborated by the density estimation techniques to generate the estimated tail functions of the distribution. The tail function or risk function (as explained in Chapter 1) is the reverse of the Cumulative Distribution Function, and it is defined as:

$$T(x) = \int_c^{\infty} f_{PDF}(x)\, dx \tag{4.12}$$

The process to generate the tail functions is summarized in Figure 4-3.

Figure 4-4 shows an example of tail generation, for a normal benchmark distribution

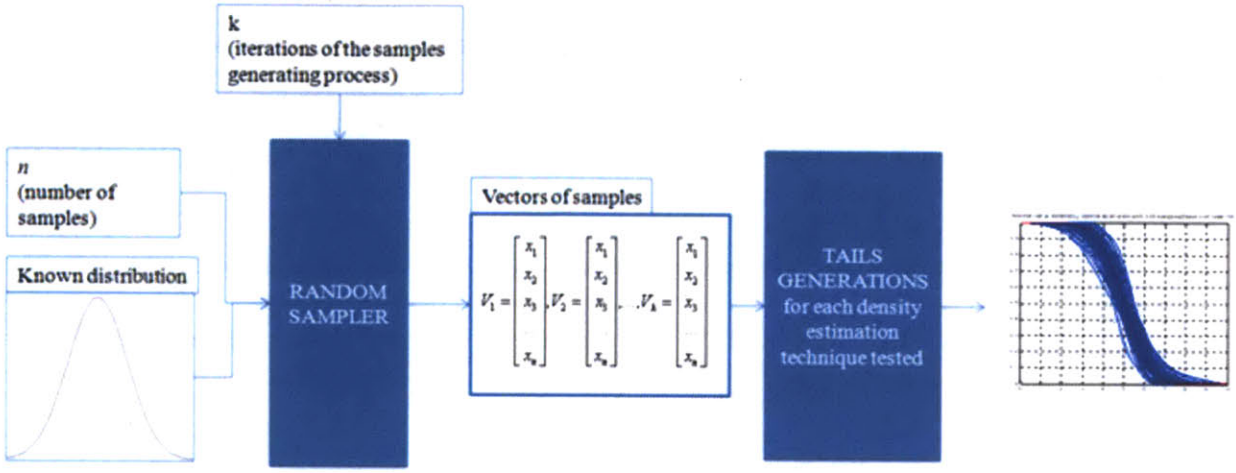

Figure 4-3: Block diagram of the process followed to generate tail functions. The known (benchmark) distribution is sampled n times, and samples are stored in a vector. The operation is repeated k times, and each vector is used to compute an estimated tail function using a specific density estimation technique.

with mean 5 and sigma 1. The number of samples used in this example is 20 and the density estimation technique tested is Kernel Density Estimator with normal kernel. The red line represents the real tail, mathematically computed by using the benchmark. Figure 4-5 presents a similar example with the same normal benchmark, and the same density estimation technique, but with a greater number of samples ($n = 90$). As expected, increasing the number of samples increases the precision of the estimator; the estimated tails are now closer to the real one with respect to Figure 4-4.

Two quantities are measured to assess the quality of each density estimation technique:

- The tails' divergence ($\Delta(x)$) across the k trials between real tail ($T(x)$) and estimated tail($\hat{T}(x)$) for each point of the distribution (x). $\hat{T}(x)$ is computed by applying the techniques previously described, for different values of the number

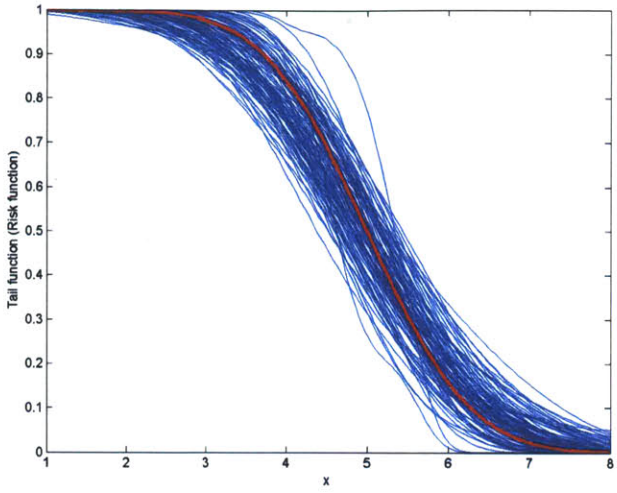

Figure 4-4: Output of the tail function generation process: the blue lines are the estimated tail functions, while the red line is the tail function for the normal distribution ($n = 20$).

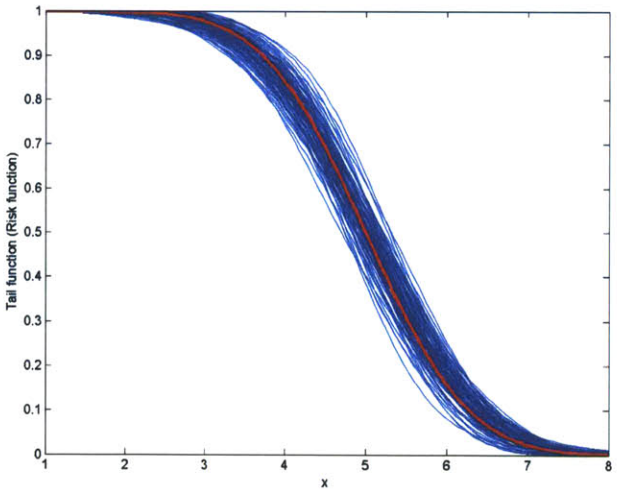

Figure 4-5: Output of the tail function generation process: the blue lines are the estimated tail functions, while the red line is the tail function for the normal distribution ($n = 90$).

of samples n. Hence, $\hat{T}(x)$ varies as the number of samples varies.

$$\Delta(x) = \frac{1}{k} \sum_{j=1}^{k} \left| T(x) - \hat{T}_j(x) \right| \tag{4.13}$$

- The estimation error ($\overline{\Delta}$)across the distribution for a certain amount of samples n (L1 metric).

$$\overline{\Delta} = \int x \cdot \Delta(x) dx \tag{4.14}$$

The results of the comparison are discussed in the next section.

4.2.3 Results

Figure 4-6 shows an example of the tails' divergence across k trials. For each point of the support, the difference between the real value and the estimated value of the tail function is calculated. This value is averaged across the k trials. The curve shows the results for different density estimation techniques applied to the same benchmark (a normal distribution with mean 5 and sigma 1), using 30 samples. Histogram is the

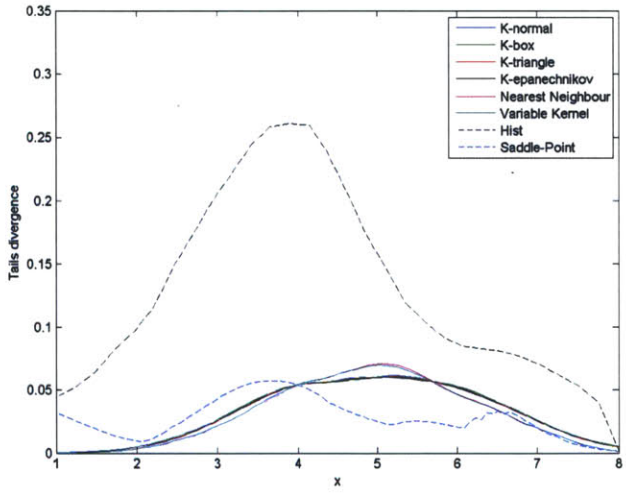

Figure 4-6: Tails' divergence computed using a normal benchmark distribution for different density estimation techniques ($n = 30$). Histogram produces the highest divergence; hence, it is the least satisfactory techniques.

least satisfactory technique (the peak of the divergence is very high), while the other estimators present better performance.

Tails' divergence changes as n increases. In Figures 4-7 and 4-8, examples of tails' divergence, computed using the same previous benchmark but with n equal to 60 and

90 respectively, are shown. As expected, tails' divergence tends to decrease as the number of samples (n) increases.

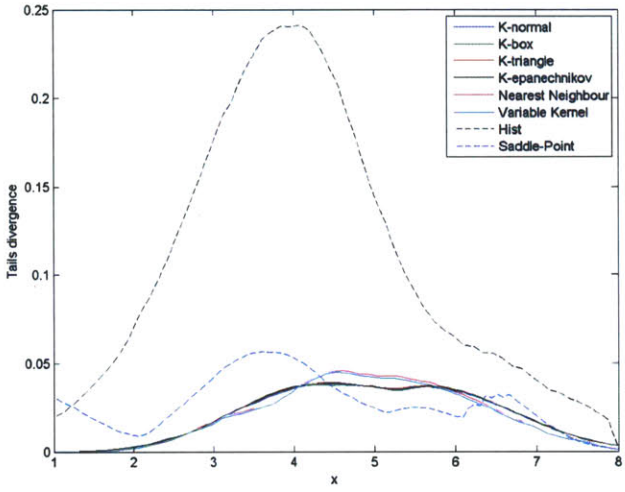

Figure 4-7: Tails' divergence computed using a normal benchmark distribution for different density estimation techniques ($n = 60$). The divergence decreases with respect to Figure 4-6.

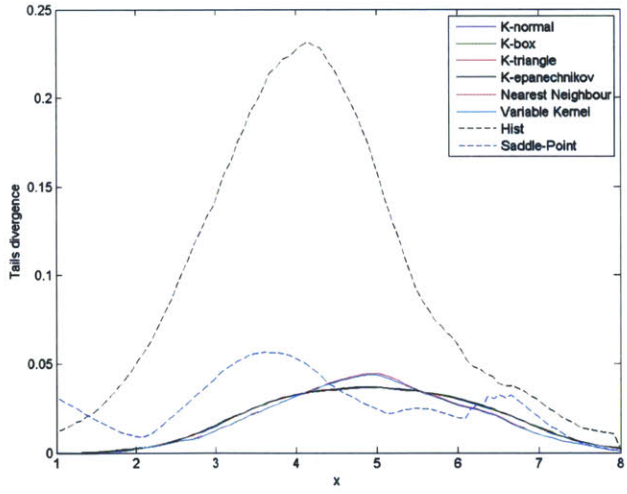

Figure 4-8: Tails' divergence computed using a normal benchmark distribution for different density estimation techniques ($n = 90$). The divergence decreases with respect to Figure 4-6 and Figure 4-7.

Finally, to characterize the effect of the number of samples, the second metric

(estimation error) is applied. This second metric is an integral of the first metric (tails' divergence) over the support of the distribution (x). Figure 4-9 shows the estimation error (Δ) for different values of n for normal benchmark.

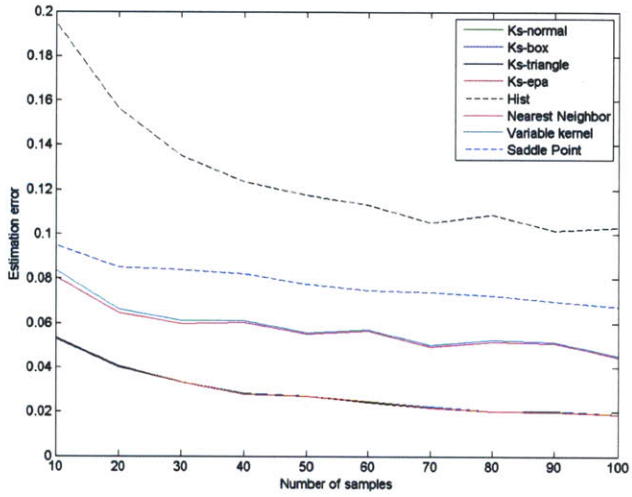

Figure 4-9: Estimation error for different values of the number of samples (n) computed using normal benchmark distributions. Histogram achieves the poorest performance, and Kernel Density Estimators achieve the best performance.

Also according to the second metric (estimation error), Histogram is the least satisfactory technique, followed by Saddle-Point Estimator. The kernel techniques obtain the best results, since an estimation error less than 4% can be achieved with only 20 samples. With different benchmarks, we expect slightly different performance due to the fact that density estimation techniques behave better when the original benchmark distribution is a normal [115]. Hence, if we change the benchmark, we expect a poorest behavior. For exponential distribution, the estimation error is plotted in Figure 4-10. It is possible to verify that the situation is less satisfying than the one of Figure 4-9. In fact, in this case 40 samples and not 20 are required to not exceed a 4% estimation error. Actually, when exponential distributions and lognormal distributions are used as benchmarks, the performance of the density estimation techniques is generally the poorest in terms of the estimation error. This result indicates that these distributions are generally more difficult to predict than others.

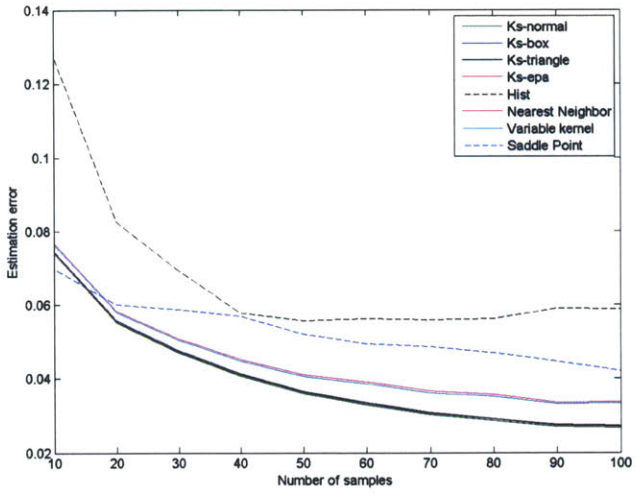

Figure 4-10: Estimation error for different values of the number of samples (n) computed with exponential benchmark distributions. Histogram achieves the poorest performance, and Kernel Density Estimators achieve the best performance. The error is greater than in the case when normal benchmark is used.

However, even if performance decreases, Kernel Density Estimators are still the techniques that achieve the best performance, especially Kernel Density Estimator with normal kernel. This result is significant, since the main motivation for this analysis is to identify the best technique to model historical data regardless of their a priori distribution (which is unknown, and for that reason this analysis has been developed with multiple benchmarks).

The calculation of the estimation error performed for normal and exponential benchmarks has been repeated for lognormal, beta and gamma benchmarks. Data are in Appendix A. Also in these cases, Kernel Density Estimators have achieved better results than Histogram, Saddle-Point, Variable Kernel, and Nearest Neighbor.

The conclusions for the analysis are the following:

- **Histogram:** It generally achieves the poorest performance compared to any other technique.

- **Variable Kernel Estimator and Nearest Neighbor Estimator:** The performance vary strongly according to the benchmark distribution used (good

with normal, bad with beta). These techniques can be helpful in some contexts, but without any previous knowledge of the "real" distributions, they can behave in an unpredictable way.

- **Saddle-Point Estimator:** It achieves good performance for very small samples size (10-20), compared to the kernel density estimation; however it does not converge as fast as kernel, as the number of samples increase.

- **Kernel Density Estimators and Naïve Estimator:** They achieve the minimum estimation error in the cases analyzed.

These results lead to an important conclusion:

> The numerical analysis performed shows that the different density estimation techniques assume the same behavior regardless of the benchmark distribution used. Specifically, the Kernel Density Estimator achieves the lower bound in the estimation error for all the cases analyzed.

Hence, the Kernel Density Estimator has been selected as the technique to elaborate the data. The size of the database is discussed in the following section.

4.3 Identification of the Required Size for the Database

When the estimation technique is selected, the required size of the database can be estimated with respect to the desired precision. The assessment of the required size of the database can be done by using results of the numerical analysis previously discussed.

Since the Kernel Density Estimator is selected as the technique to use, its ability to predict the distributions is analyzed across different benchmarks. The benchmarks used strongly influence the quality of the estimation. It is generally easier to predict normal behavior (Central Limit Theorem) than any other kind of distribution. However, regardless of the benchmark, it is possible to identify the sample sizes for which the estimation error lies in a certain interval, which is shown in Figure 4-11.

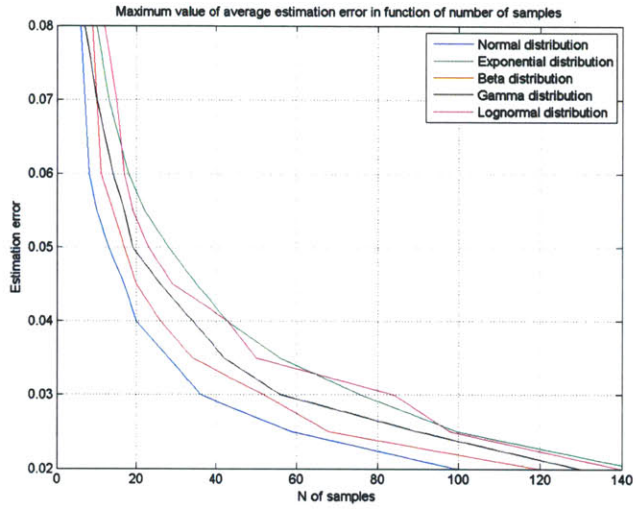

Figure 4-11: Samples required to obtain a given precision for different benchmarks using Kernel Density Estimator (normal kernel).

Figure 4-11 illustrates the number of samples required, using Kernel Density Estimator, to achieve a certain precision according to the benchmark distribution used. For example, if the designer wants to use historical data, and wants to achieve an estimation error less than 4%, no less than 20 samples are required.

4.4 Database Organization and Population

The population of the database represents a challenge due to the tendency of agencies and companies to not publish their data. To partially solve this problem, another analysis on the size of the database has been performed. From the previous section it is clear that a certain number of samples are required to obtain a certain precision. However, another interesting question is: what is the desired precision? Ideally, we would like to be as precise as possible, but since, as it can be inferred from Figure 4-11, this precision would often require too many samples, another approach to this problem is to ask what is the number of samples required by the Kernel Density Estimator to improve the statistical estimation with respect to the traditional approach. To answer this question, a simple calculation can be developed.

Any communication system can be simplistically modelled as a combination of different components. The number ranges from a minimum of two (an antenna and a transceiver) up to n. Design risk estimation is the process of estimating the total mass fluctuation which is the sum of the fluctuations of different components, under the hypothesis of independence. If fluctuations are computed using Kernel Density Estimator, the estimation error for a different number of samples is estimated in Figure 4-11. A pessimistic assumption is that the database contains the same number of samples k for each type of component. Hence, from Figure 4-11 each type of component will have the same ϵ error. This choice is a worst case: in fact in the real database the number of samples for different types of components can be different and in some cases greater than k. Finally, under the assumption that components' masses are considered independent from each other, the total estimation error can be calculated using the following expression (which derives from the property of variance composition for independent variables [18]):

$$E_{tot} = \sqrt{\epsilon_1^2 + \epsilon_2^2 + \epsilon_3^2 + ... + \epsilon_n^2} \tag{4.15}$$

When the error is computed, this approach can be compared to the traditional contingency approach which quantifies PDR's fluctuations as a 30% fixed margin value [135]. In the risk estimation process followed in this research, the objective is to be as precise as possible, but every improvement with respect to the traditional approach is valuable. Hence, the estimation error of each component needs to be such that the total error (on the sum of the component) is less than equal to the contingency value of 30%. This estimation corresponds (defining the contingency as C) to:

$$\sqrt{\epsilon_1^2 + \epsilon_2^2 + \epsilon_3^2 + \ldots + \epsilon_n^2} \leq C \tag{4.16}$$

Figure 4-12 shows the result of this computation. On the x axis, the number of components is represented, and on the y axis the total error is computed. The red line indicates the 30% contingency. The other lines show the estimation error in function of the number of components for different values of ϵ. It is possible to verify that a

precision of 4% is not always required. Specifically, when the system is composed of few components, an higher ϵ for each of them will still allow this analysis to improve the traditional approach.

Specifically, in the test cases for this research (Chapter 6) the communication systems considered have less than 5 components. Hence, from Figure 4-12 even assuming an error of 9% (0.09 in the Figure) for each single component, the statistical approach produces a better estimate than the traditional contingency method. From Figure 4-11, it can be noticed that an error no greater than 9% can be obtained with less than 20 samples, which is a reasonable amount of data.

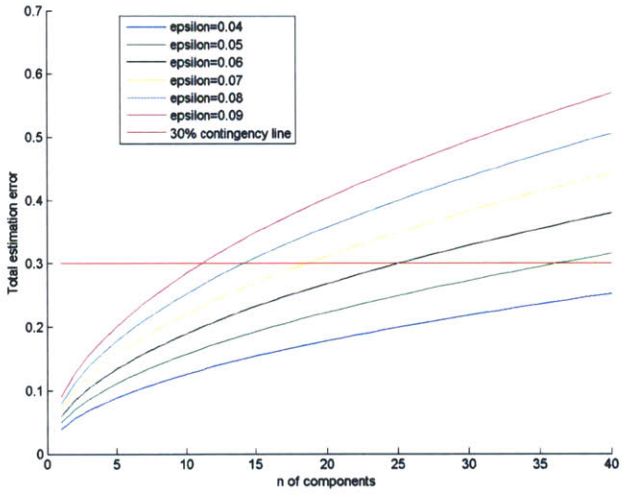

Figure 4-12: Estimation error using Kernel Density Estimator in function of the number of components. For communication systems with very few components even assuming an error of 9% for each single component, the statistical approach produces a better estimate than the traditional contingency method.

The data included in the database have been collected through different literature sources. Specifically, articles on the design of specific components for communication systems have been used ([24], [105], [55], [118], [21], [126], [39] , [63]). Another source of data is JPL Descanso publications: a collection of design documents for the communication systems of different missions (Voyager, Galileo, Cassini,...) designed at JPL. Each document([64], [125], [76], [123], [73], [120], [122], [86], [124], [119], [110], [121], [62], [13]

) contains a detailed description of the communication system with data on mass and power for components.

Data have been organized for different components, and categories of components. The components include: antennas (categories: Low Gain Antenna-LGA, Medium Gain Antenna-MGA, High Gain Antenna-HGA), transceivers, amplifiers (categories: Solid State Power Amplifier-SSPA and Traveling Waves Tube Amplifier-TWTA), switches, and diplexers. The motivation for which a type of components is sub-divided into categories is clarified in the next section. The antennas are also categorized by frequencies (included in the model are UHF, S-Band, X-Band, and Ka-Band). However, due to the lack of samples in UHF and Ka-Band, test cases have been selected to be in S-Band (for LGA) and in X-Band (for MGA and HGA). A very limited amount of data is also available for switches and diplexers; hence these components have not been analyzed in the test cases.

The next section describes the process to generate probability density distributions by using the samples and the Kernel Density Estimator.

4.5 Use of Database to Model Distributions

The statistical model uses sample data to compute the probability that a certain component will exceed caps in mass and power. As already mentioned in previous sections, these risks cannot be described with a physical model, because the fluctuations depend on several causes difficult to model. For this reason, the model uses past historical data. However, a challenge is that these data depend on different parameters. For example, the mass of an antenna depends on multiple factors including the gain and the frequency. Hence, if the database includes the mass of the antenna x, this data can be applied to estimate masses of antennas with the same gain and frequency. However, this approach will lead to the collection of an enormous amount of data. To solve this problem, a different approach is proposed.

Each sample is converted into a coefficient which represents a relation between two parameters. For example, in the case of the antennas, data are collected as fractions

of mass per unit of gain. In this way, samples of different antennas can be used to calculate the statistics. This process is preferable due to the limited data available. The generation of probability densities is summarized in Figure 4-13.

The model receives as inputs the value of a certain design metric. In Figure 4-13,

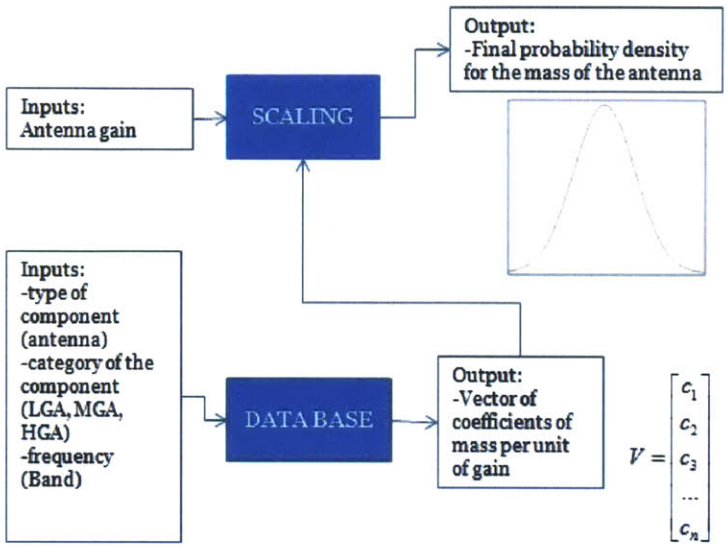

Figure 4-13: Block diagram of the probability function generation process. Type and category of component, band, and performance metric are used to identify and to scale the correct set of sample data.

the probability distribution of the mass of an antenna, given its gain as input, is estimated. Other inputs include the type of component (if it is an antenna, a transceiver, an amplifier), its band, and its category (LGA, MGA, HGA). Type, category, and band are used to identify the appropriate set of coefficients. As the coefficients are extracted from the database, the value of the design metric (gain) is used to scale the distribution. At that point, a probability distribution is generated using the Kernel Density Estimator. The mathematical formalization of the approach is presented in the following equation, where $\bar{a}$ is a vector of coefficients of mass per unit of gain (G) which contains n elements:

$$f(x) = \frac{1}{G} f(x, \bar{a} \cdot G) \tag{4.17}$$

Recalling the expression of Kernel Density Estimator, $k\left(x, \bar{a} \cdot G\right)$ is:

$$k\left(x, a\bar{\cdot}G\right) = \frac{1}{n \cdot h} \sum_{i=1}^{n} K\left(\frac{x - a_i \cdot G}{h}\right) \tag{4.18}$$

$K\left(\cdot\right)$ is a symmetric kernel function, in this case a normal kernel.

The only caveat in this approach is that the mathematical operation performed is a transformation of a probability density, which holds only if the two variables considered (here mass and gain) have a linear relation with each other. For some components, the relation is linear: Brown [24] shows (Figure 4-14) that for the Traveling Wave Tube (TWT) amplifiers there is a linear relation between mass and output power, and between input power and output power. However, in some cases, the

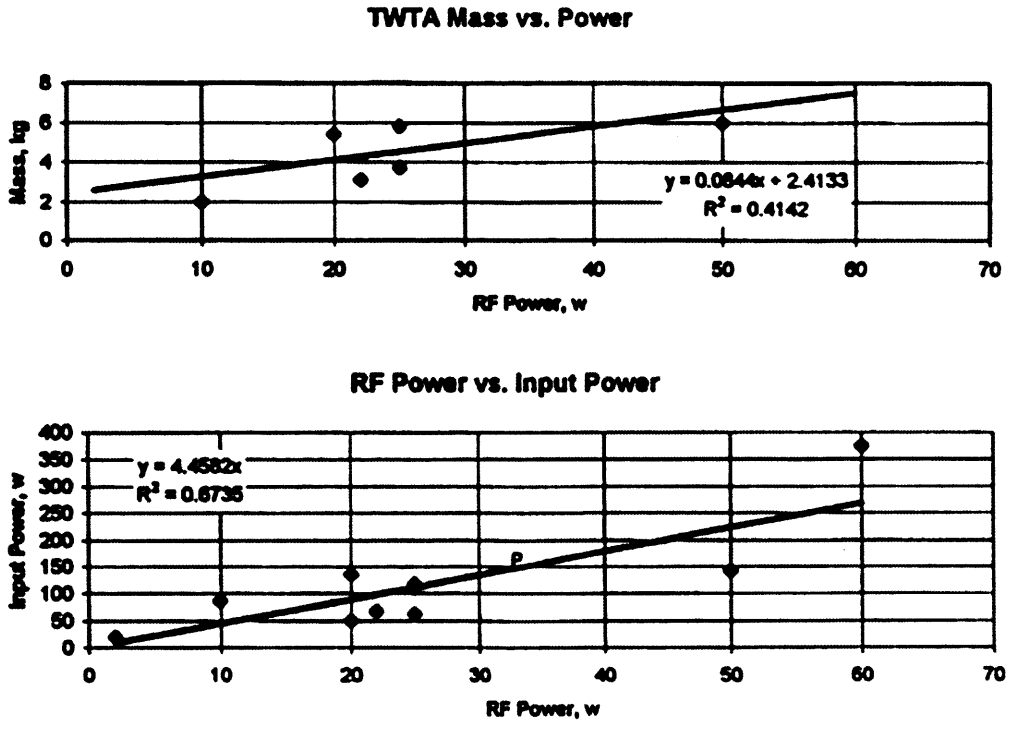

Figure 4-14: Power and mass estimations for TWTA. The plots show linear relations. Source: [24].

relation between the two quantities is not necessarily linear, as in the case of antennas (Figure 4-15). To solve this problem the non-linear function is decomposed into

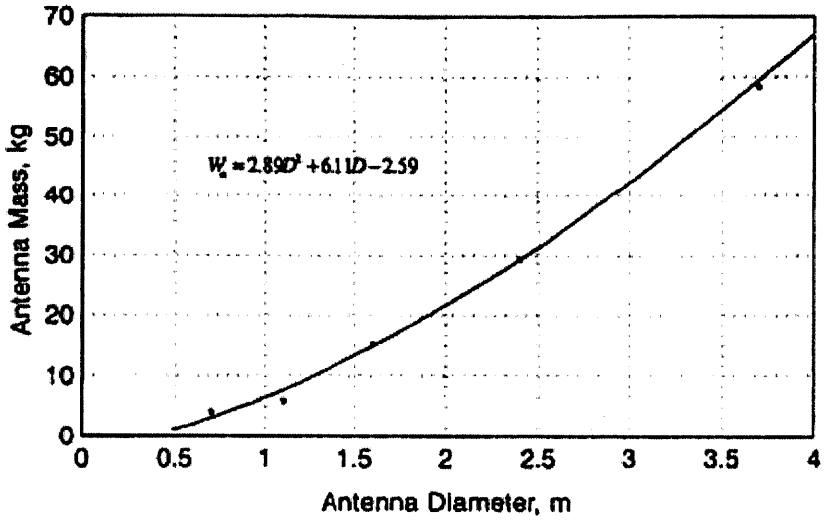

Figure 4-15: Mass estimations for antennas. The relation between diameter and mass is not linear. Source: [24].

a set of linear approximations through a convex piecewise linear approximation of the non-linear graph. This approximation breaks the non-linear curve into multiple pieces, which in the case of the antenna corresponds to different antenna gains: Low Gain, Medium Gain, and High Gain. This approximation is the reason for which one of the inputs of the database model is the component category. When we search in the database for the set of coefficients required to construct the distribution, the software needs to know to which category the antenna belongs in order to select the coefficients corresponding to the correct linear piecewise approximation. More details can be found in [6] and in [8].

An example of the computation of a density estimation probability distribution using the database and the Kernel Density Estimator is described in the following section.

4.6 Example of Data Statistics

This section presents an example of how the database and the Kernel Density Estimator are used to generate a probability distribution to describe mass fluctuations for

a component. This example is based on the antenna used for the CASTOR mission. More details on CASTOR mission and on the communication system will be given in Chapter 6.

The antenna is a 6 dB patch antenna which has passed through different design phases. Initially, it was supposed to be a Commercial-Off-The-Shelf (COTS) patch antenna, but then due to some vibration concern the team resorted to a custom-made antenna. The material selected changed across the different versions until the final configuration was reached after the CDR. The initial mass of the antenna was estimated to be around 0.5 Kg, while the final value became 0.1 Kg, showing that the initial estimate was significantly overestimated. The data statistic is used to compute a probability density which should indicate that the typical value for the mass of this kind of patch is in the order of the final CDR value. This expected result would show that data statistics are able to indicate whether or not the initial design value is an overestimation.

The computation of the probability density for the antenna is performed in the following way. The inputs for the database are:

- Type of component: in this case an antenna;
- Category of the component: in this case LGA;
- Performance metric of the component: in this case the gain of the antenna (6 dB);
- Frequency of the component: in this case S-Band.

This information is used to select sets of data in the database. Probability density and tail functions are shown in Figure 4-16 and 4-17. The peak of the probability density is around the final (CDR) value of the antenna mass: the data statistic is able to indicate that the initial estimation for the antenna's mass was an overestimation. However, this method presents some limitations due to the unavailability of data. It is true that 20 samples are enough to improve the estimation with respect to the traditional approach; however, the tail function exhibits a stretched-out shape with a

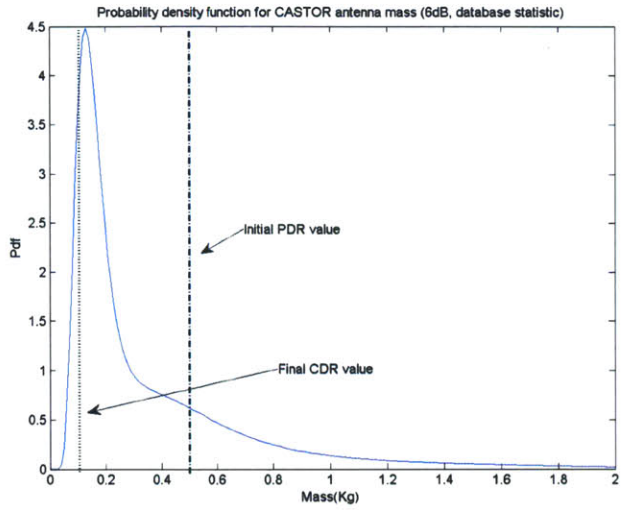

Figure 4-16: Probability density function for CASTOR antenna mass calculated using data statistic. The peak of the distribution is close to the final CDR value of the mass of the antenna.

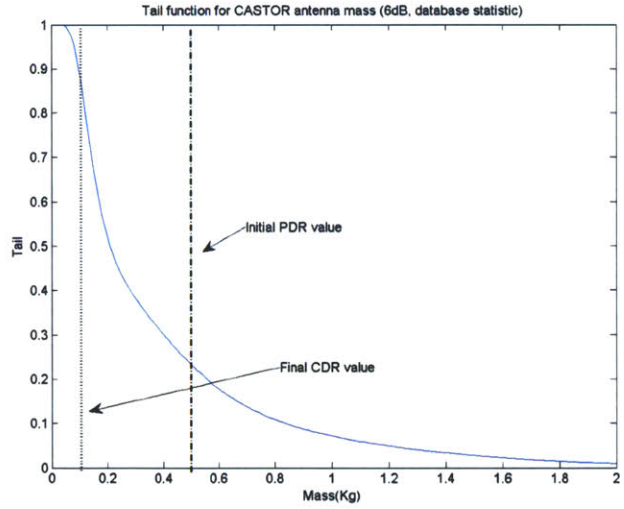

Figure 4-17: Tail function for CASTOR antenna mass calculated using data statistic. The sparsity of the data generates the heavy tail.

heavy tail as shown in Figure 4-17, indicating an overestimation of the mass margin required to retire a given design risk. For example, if the risk requirement of not exceeding the mass allocation is 0.1 Kg, the mass margin that would be required using the database method is much bigger than the one that would result from using

the expert approach or the Bayesian composed approach as it will be shown in the following chapters. This heavy tail is due to the limited number of samples of the database. Hence, an integrated approach which includes expert opinions is strongly suggested.

4.7 Summary

This chapter describes the historical data approach implemented in the thesis. The limited data availability and the necessity of using non-parametric density estimation techniques are underlined in the introduction of the chapter. As a consequence, a numerical comparison to identify the best density estimation technique for small sample sizes is performed. This analysis uses known probability distributions as benchmarks to measure the ability of the different techniques to estimate the distributions. The results of this analysis show that Kernel Density Estimator achieves the minimum error (divergence). Hence, this technique has been selected to model the distributions. The rest of the chapter discusses the minimum required sample size for the database to improve the current state of the art (30% margin applied at PDR). The analysis shows that for a limited number of components even a small number of samples (like 20), and the use of Kernel Density Estimator can improve the estimation of fluctuations. The last part of the chapter describes database organization, database sources, and generation of probability distributions. Finally, an example of the approach using the antenna of the CASTOR mission (one of the test cases discussed in Chapter 6) is presented. The following chapter is dedicated to expert elicitation approach.

Chapter 5

Expert Elicitation Statistics

This chapter presents a statistical approach to apply expert opinion in risk estimation. Expert opinion have been used in every field of science since the development of the scientific research method (18th century). However, until the 1950s the knowledge of the expert was applied as a suggestion or a hypothesis. It was only after the Second World War that expert opinion started to be modelled in probabilistic terms, and it became introduced on a large scale in risk estimation problems [34]. The concepts of expert elicitation, biases, heuristics, expert calibration, and Delphi methods described in this chapter were formulated for the first time in the 1960s ([34], [100]). Today the field of expert elicitation constitutes a very active branch of research.

Even though multiple approaches have been implemented, the development of methods to quantify experts' biases and calibration represents a challenge. As a result, the integration of multiple and often conflicting opinions can be problematic, due to the complexity of properly weighting experts' contributions.

In this chapter, an approach to address the above problem is proposed. This approach is especially useful when probability densities for seed calibration variables are not available. The methodology proposed generates an expert score that is employed to aggregate multiple expert assessments. The approach has been experimentally tested with students and professional engineers. Results indicate that the approach improves the quality of the estimations: the weighted aggregations of experts' estimates based on the experts' scores achieve better results than the correspondent aggrega-

tions based on experts equally weighted. Moreover, expert elicitation improves the statistical assessment of the risk with respect to the same assessment computed using only historical data.

This chapter covers all the aspects of expert elicitation related to the thesis. Section 5.1 motivates the use of experts in this research and gives some definitions in the field of elicitation. Section 5.2 describes the methodology and defines the quantitative metrics. Section 5.3 shows the result of the experimental tests. Section 5.4 discusses the integration of expert elicitation and historical data in a Bayesian framework. Section 5.5 provides concluding remarks.

5.1 Overview

Expert elicitation has been increasingly applied to different areas of research and engineering. Complex systems and problems not easily modelled through governing equations, hence requiring empirical analysis, are the driving forces for the spread of new approaches. In fact, in some cases, the best source of information for a problem are experts in the field, people who have developed experience in a certain area and who can provide insights of immediate applicability.

This field of research is known as expert elicitation or subjective probability [100]. According to this definition, probability represents someone's degree of belief in an uncertain proposition. Subjective probability is an emerging field of research, and different authors have explored multiple aspects of the problem.

In Chapter 2, the principal research on expert elicitation, biases and heuristics have been described. Previous research and experiments have shown how experts' biases and poor probabilistic thinking affect the quality of estimation; however, the development of a quantitative approach to weight experts' performance is still a challenge. Fundamental work in this field has been done by Cooke [34] who defines a scoring system based on two quantities: the scoring of each calibration question and the level of certainty (or information) associated with that quantity. This scoring system is part of the Cooke Classical Model, which uses seed variables to perform calibration.

The term *seed variables* indicates variables for which the real probability distribution is known (mathematically or empirically with the use of observations), and hence can be applied to calibrate experts' assessments.

Implementing this approach, however, can be challenging when no empirical data exist to compute the probability densities for the seed variables, as in the specific case of this research. In similar cases, authors Cooke and Hagan ([34], [100]) propose the use of previous expert elicitation information as probability densities for the seed variables. To the best knowledge of the author, no previous expert elicitations in the field of design risk analysis for space communication systems have been performed. Hence, as a consequence of unavailable data to compute the probability densities for seed variables, a modified approach to measure experts' sensitivity to biases and to calibrate their assessments is proposed in this thesis. The approach aims to develop a combined score to quantify experts' performance in terms of how experts model probabilities, which biases they are sensitive to, and how overconfident or underconfident they are in their estimations. Specifically, since no probability densities for the seed variables are available, the scoring system has been divided into two parts: a quality score that measures the sensitivity of experts to biases, and a calibration score that measures the ability of the expert to assess quantities in his/her field of expertise. The combined expert score is employed to aggregate probability densities from different experts into a unique estimate.

Before discussing the expert approach, it is important to recall motivation for applying expert elicitation and to clarify definitions in the field of expert elicitation. Specifically, expert elicitation is required in statistical risk estimation for the following reasons:

- Size of the database: From Chapter 4 it is clear that to avoid the problem of heavy tail distributions a database with many samples is required. Space missions have some historical tradition, but the size of databases for space components is still limited. As a consequence, an expert elicitation approach can compensate for the limited sample size.

- New components: If a component is unique, obviously no historical statistic exists. In this case, an expert elicitation approach is required.

- Cases in which the expert opinion is more reliable than historical data: Historical data are useful to create statistics, but in the world of technology, they do not always represent the evolution of the technology. For example, the size of processing units for commercial applications and for space applications tends to reduce quickly as the technology evolves. Hence, the calculation of the mass fluctuations for this kind of component through a historical database generates inevitably strong overestimation, while an expert in the field is likely to more accurately estimate the value and the variance of the final mass of the component.

A generic definition of expert from Merriam and Webster [90] is as follows:

> "An expert is a person who displays special skills or knowledge derived from training or experience."

However, as Hagan [100] points out, expertise also involves how the person organizes and uses certain knowledge. In fact, to have a good elicitation, the expert has to be able to express uncertainty accurately. The entire process of elicitation can be defined as [4]:

> "The synthesis of opinions of experts of a subject where there is uncertainty due to insufficient data, when such data is unattainable because of physical constraints or lack of resources."

Additional definitions in the field of expert elicitations are as follows:

- **Bias:** in expert elicitation process a bias (also called cognitive bias) is the "human tendency to make systematic errors in certain circumstances based on cognitive factors rather than on evidence [25]."

- **Heuristics:** simple, efficient rules, which are applied to make decisions, to express judgments, to solve problems, typically when facing incomplete information. These rules can lead to systematic errors or biases.

- **Calibration:** a measurement of the agreement between expert opinion (subjective probability) and observed relative frequency. For example, if a weather expert is asked to guess rain probability for a year on the basis of certain data, and we record the real rain frequency for that year, we can assess how well our expert is calibrated. Generally, calibration is described using the *calibration curve*, shown in Figure 5-1. Perfect calibration is represented by the points falling on the identity line between expert probability and relative frequency. The black curve is an example of a well calibrated judgment. In this thesis, calibration is measured through point estimations and not through probability densities. Hence, in the specific context of this research calibration can be more precisely defined as a measurement of the agreement between expert opinions and observed estimated data.

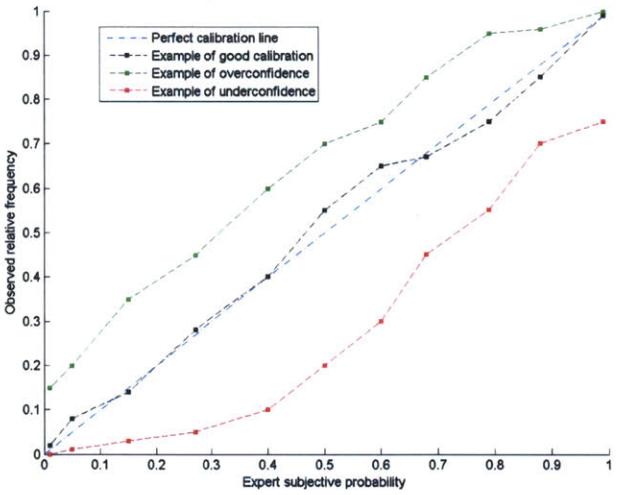

Figure 5-1: Example of calibration curve. The identity line indicates perfect calibration. The black line is an example of good calibration, while the green line and red line indicate overconfidence and underconfidence respectively (Graphic generated using data from [100]).

- **Overconfidence/Underconfidence:** These are the two opposite cases that can be identified when there is not good calibration. Specifically, when more than the average number of points falls above the identity line, the expert is said to exhibit an overconfident behavior. Conversely, when more than the

average number of points falls below the identity line, the expert exhibits an underconfident behavior. These phenomena are illustrated in Figure 5-1.

- **Experts' composition:** It is the process used to aggregate multiple experts' judgments into a unique probability curve.

The next section details the expert elicitation methodology designed for this research.

5.2 Methodology: Expert Triangle and Four-Part Process

To perform a reliable expert elicitation, it is necessary to measure the ability of experts to perform probabilistic thinking, and their tendency to over- or underestimate the parameters of interest (when seed variables with related probability densities are not available). This process should help to construct probability models that are unbiased and faithful to the experts' true understanding. To accomplish these goals, a four-part methodology has been developed, and it is summarized in the expert triangle graph (Figure 5-2). The expert triangle shows the four parts of the methodology –Probabilistic Thinking, Calibration, Elicitation, and Expert Aggregation– and their relationship. The first three parts reflect the interview, while the last part refers to the method of aggregating multiple assessments into a unique estimate. To obtain good performance in Elicitation, the expert must be able to model phenomena probabilistically, an ability assessed by the first part of the methodology, Probabilistic Thinking. The expert must also be able to estimate the quantities of interest in his/her field of expertise with a certain degree of accuracy, an ability measured by Part 2, Calibration. The results of these two parts generate two scores: a quality score from Part 1 and a calibration score from Part 2. Calibration also produces a coefficient that constitutes an input for the third part, Elicitation. Finally, Part 4 Expert Aggregation, takes as input the single probability densities modelled by each of the experts in Elicitation, and aggregates them through the quality score and calibration score.

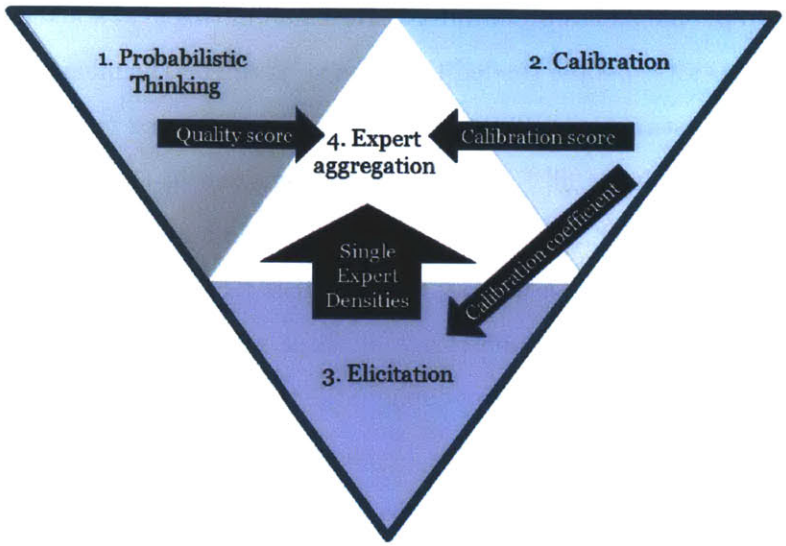

Figure 5-2: Expert triangle: overview. The methodology includes: Probabilistic Thinking, Calibration, Elicitation, Expert Aggregation. Probabilistic Thinking and Calibration generate scores that are used to compose the estimates of different experts into a unique estimate. Calibration also generates a coefficient that is used to shift the distributions in Elicitation.

Before detailing the methodology, the term "quality score," for the purpose of this thesis, should be clarified. The term "quality score" does NOT indicate any judgment of the experts and or their expertise. The term simply refers to a measure of their sensitivity to biases, and the results of the interview process. Hence, this thesis does not aim to measure the quality of an expert, but rather how much he/she is sensitive to the biases mapped by the interview process. The next four subsections detail each part of the methodology.

5.2.1 Part 1: Probabilistic Thinking (How the Brain Works: Principal Biases and Heuristics in Expert Elicitation)

Probabilistic Thinking focuses on developing quantitative measurements to characterize biases and heuristics that affect experts' judgment. Relevant literature in this field belongs to the area of psychology and cognitive science.

The first question that a researcher should ask before initiating an elicitation process

is: "how well do people assess probability?" Research in this field started in the 1950s with conflicting results: people tended to assess very well on the mean, but performed poorly in variance estimation. They had difficulties estimating the degree of association of binary variables [100]. Also, as the dimensions of the problem increased, performance decreased [99]. A series of studies to understand these phenomena was performed by Tversky and Kahneman in the 1970s ([127], [128]). Through sets of experiments with specific questions (some of those applied in the interview for this research), they discovered that when people judge probabilities, they often use heuristics, which causes systematic deviations (biases) from the optimal solutions.
Part 1 of the proposed methodology aims to map expert sensitivity to biases and heuristics. A questionnaire has been developed for Part 1 (Appendix B), and a quantitative scoring system has been generated. The following subsections present some of the biases and heuristics studied in the literature.

Judgment by Availability

It is observed that when an expert judges the probability of an event according to the ease with which similar events are called to mind [128], events that are easier to recall are the ones judged more probable and vice versa. This heuristic is effective in some cases, because on occasion if an event is more likely to happen, more instances of it can also be recalled. However, some biases can be created if other factors influence the ease with which events are recalled. For example, personal experiences can affect the judgment: if a person or her/his family has been recently victim of a car accident, the person will be more inclined to recall the event in the assessment of probability of car accidents. Another example is shown as follows ([53], [128]):

If we choose a random English word, it is more likely to:

1. Start with an r

2. Have r as its third letter

People will generally indicate the first answer, since it is easier to recall words by their starting letter (e.g., red, rank, rogue, road, rope, . . .) than by their third letter (e.g. park, bird, wire, . . .). However, the correct answer is the second.
In the specific case of engineering design risk analysis, this bias is relevant, since some experts may have more experience in the design of some components than others. Hence, availability plays an important role in expert elicitation for this research.

Judgment by Representativeness

This heuristic is generated when an expert is assessing the probability that A belongs to some class of B [100]. In this case, instead of evaluating the conditional probability, a person will tend to judge on the level of similarity between A and B [53]. Tversky and Kahneman [128] define representativeness as:

> "an assessment of the degree of correspondence between a sample and a population, an instance and a category, an act and an actor or, more generally, between an outcome and a model."

The problem with this heuristic occurs when additional data that should influence probability assessment are not taken into account due to representativeness. For example, a common error made by using this heuristic is that little attention is paid to the unconditional probability of the events. An example is shown as follows [53]

> Mr. X is described as "meticulous, introverted, meek and solemn." Assess the probabilities of Mr. X being engaged in one of the following occupations:
>
> 1. Farmer
> 2. Salesman
> 3. Pilot
> 4. Librarian

5. Physician

Most people (as well as those in our interview) assign a high probability to Mr. X being a librarian based on stereotype. However, they completely ignore base rates, such as the relative number of salesmen to librarians [128].
An error that is generally connected to representativeness is illustrated by the following example, which is also defined as *conjunction fallacy.*

> Linda is 31 years old, single, outspoken and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice and also participated in anti-nuclear demonstrations. Please tick the most likely alternative:
>
> 1. Linda is a bank teller
>
> 2. Linda is a bank teller and is active in the feminist movement.

Most people assess the second alternative as most likely [100]; however, it cannot be more likely since the conjunction of the two events can never be more probable than each event separately. This bias is known as conjunction fallacy, but is also an error of representativeness. Given that design uncertainties in communication systems depends on many factors (i.e., frequency, material, type of components, process of fabrication, etc.), it is important to understand whether experts are influenced by the bias of representativeness, and whether they can distinguish subsets from conditional probabilities.

Judgments by Anchoring and Adjustments

This bias is observed when people who are asked to estimate a quantity start with an initial estimate ("anchor"), and then adjust up or down. This strategy is useful in cases of repeated judgments for the same kind of problem. However, this heuristic can be a problem if people remain too close to the initial assessment [100]. An example is shown as follows:

> The mean IQ of the population of eighth graders in a city is known to be 100. You have a random sample of 50 children to test. The first one tested has an IQ of 150. What do you expect for the whole sample? [128]

The correct answer is 101. People tend to believe that the knowledge of the first value does not change the result, and therefore tend to say 100, remaining anchored to the initial value.
In the case of this research, experts tend to anchor to conservative estimates. Hence, measuring anchoring is relevant to this analysis.

Hindsight Bias

This bias is observed when people are asked to assess their a priori probability of an event that has actually occurred. An example is shown as follows [53]:

> In 1988, what was the probability that the Berlin wall would come down within the next five years? [53]

In 1988, it may have appeared unlikely that the Berlin wall would have come down; however, today knowledge of what occurred affects memory such that people tend to increase their a priori probability assessment for the event. This bias is probably the less relevant with respect to our research context. However, we have included a question to map this bias. Future research will most likely focus more on the other biases.

Awareness of Underlying Conditional Probabilities

This bias is observed when people do not take into account the underlying probabilities of events [53]. An example is shown as follows:

> Mark is 30 years old, and has a college degree in music with a minor in economics. Since his early years, music has been his true passion. He started playing and studying guitar when he was only 4 years old. He also sings and composes his own songs. During high school and in college, he played basketball. Express (in percentage) your likelihood that Mark is:

1. Famous (world-known) artist/song writer
2. Bank teller
3. Music teacher
4. Basketball coach

In this case, people tend to assign high probabilities to the possibility of Mark being a famous song writer since the description matches the profile of a famous artist. However, the underlying probability of being a famous artist is extremely low compared to the other professions.

As in the case of representativeness, this bias is relevant to this research as multiple factors can influence mass and power fluctuations for components. Hence, it is important to model experts' awareness of underlying conditional probabilities.

Law of Small Numbers

Sampling theory indicates that the size of a sample affects the probability of obtaining certain results. However, people generally forget that larger samples are less likely to deviate from the mean than smaller samples. Garthwaite defines the law of small numbers as the following:

> people expect a sample from a population to represent all the essential characteristics of that population, even if the sample is small [53].

An example of a typical question to study the law of small numbers is shown as follows:

> Suppose that a doctoral student has completed a difficult and time-consuming experiment on 40 animals. He has scored and analyzed a large number of variables. His results are generally inconclusive, but one before-after comparison yields a highly significant t = 2.70, which is surprising and could be of major theoretical significance. Considering the importance of the result, its surprising value, and the number of analyses that

> your student has performed, would you recommend that he replicate the study before publishing? If you recommend replication, how many animals would you urge him to run [127]?

In this case, people recommended replicating the experiment and they generally suggested at least 20 samples [127]. However, if mean and variance in the new set of samples are identical to the old set, the final value of t will be $t = 1.88$. Hence, the student chance of obtaining the same results testing a new sample set would be very small, less than 5% probability [127].
Due to the limited amount of data in engineering design risk analysis for communication systems, we believe that this bias is relevant.

Coherence

Finally, it is important to check the *coherence* in probability assessment. Besides heuristics and biases, one of the most important ways to assess how the brain models probability is to check whether it respects the basic laws:

1. The sum of a disjoint set of events that includes all possible events in the sample space has to sum to 1.

2. The probability of an intersection of independent events is the product of the respective probabilities.

Simple questions can verify whether a person has an interior understanding of these concepts. For example:

> Given the following possible places to spend a weekend, where do you most likely think that a person will go? Give percentages to the following options:
>
> 1. Hiking
>
> 2. Beach
>
> 3. Disneyland

4. Vegas

5. Museum of Arts

In this case the analyst is not interested in any of the probability numbers the expert will give per option. The only check performed is whether the sum reaches exactly 100%. Similar examples can be created to verify the expert's awareness of the second law. One example is as follows:

> Given that only 30% of the population is affected by disease x, and only 20% of the population is affected by disease y, what is the total percentage of population affected by both diseases?

If the expert's answer is 6% , the interviewer will verify that she/he has an understanding of law 2.

Coherence should be the base of any probabilistic model assessments. Hence, we consider coherence a key aspect to map in the elicitation process for this research. The next section presents the structure of the interview and the quantitative metrics used to model Part 1.

Interview Structure for Part 1 and Quantitative Metric

Part 1 of the interview includes 16 questions (partially taken from the literature and partially developed by the author), which monitor experts' sensitivity to biases and heuristics. The result of this first part of the methodology is a quality score, which is the sum of the scores for the single questions (p total) as shown:

$$Q = \sum_{i=1}^{p} s_i \tag{5.1}$$

Q is a number between 0 and 100, and can be considered a cumulative index that measures the contributions of the different biases. However, since each of the questions explores one or more of the biases at a time, it is possible to develop subindices, which show how the expert is affected by a specific bias with respect to the others. These indices are not practically used in the model, but they can be analyzed to understand

if an expert is specifically affected by one of the heuristics. Strong evidence [100] suggests that experts can improve their assessment as they become aware of their inner biases. The subindices mapped in the interview process are the following:

- Coherence index (CI)
- Availability index (AI)
- Representativeness index (RI)
- Awareness of underlying conditional probabilities (UI)
- Hindsight bias index (HI)
- Small sample index (SI)
- Anchoring index (NI)

The results of these questions are elaborated by the Matlab software to calculate the scores. The complete set of questions is reported in Appendix B. Table 5.1 contains a summary of the questions: for each of them the heuristics that are checked by the question are marked. The first part of the interview lasts approximately 30 minutes, and does not necessarily require a background in communication system design. For this reason, it is suitable to test different possible subjects. Hence, undergraduate students, graduate students, and professional engineers performed this part of the interview. Section 5.3 discusses the results.

5.2.2 Part 2: Calibration (Overestimation and Underestimation)

According to the Cooke Classical Model, calibration is a measurement of the agreement between expert opinion (subjective probability) and observed relative frequency. According to Cooke's theory [34], calibration can be measured using *seed variables*, quantities for which the realization is known to the analyst but not to the experts.

Table 5.1: Subindices analyzed by each of the questions for Probabilistic Thinking. Currently, the analysis explores all the subindices. Future work may be focused only on a subset of them.

Question	CI	RI	UI	AI	HI	NI	SI
1	x						
2	x						
3		x					
4	x						
5	x		x				
6	x		x				
7				x			
8							x
9					x		
10		x	x				
11							x
12	x						x
13		x		x		x	x
14	x					x	
15	x						x
16	x						x

The seed variables are used to determine two quantities for each expert: a *calibration* score and an *information* score. Then the scores are used to compute experts' weights according to an asymptotically proper score rule, for which the expert who wants to maximize his/her weights, should provide quantities that he/she thinks are genuinely correct. *Calibration* is generally modelled as a chi-square statistic of the deviation of observed frequencies from expected frequencies [137]. *Information* measures the degree to which an expert gives narrow uncertainty assessments, and it is generally measured as the relative information between the expert data and a background distribution over the intrinsic range. The product of *calibration* score and *information* score is then scaled and normalized to one.

The calibration methodology implemented in this research is slightly different from the Cooke Classical Model, due to the special situation required by the application–engineering design risk analysis for space communication systems. In fact, the elicitation process proposed in this research is composed of questions of this nature:

> The communication system for satellite X is equipped with an antenna of y dB gain in the S-Band. The initial estimate for the mass is m kg. Please model the probability density that expresses the probability of the final mass value.

The uncertain variables of interest are probability functions for mass and power of components, for which we do not have known values. This research is new and the "probability densities" for seed variables required to apply the Cooke Classical Model are not available.

As a consequence, the calibration process has been redefined for this specific application. Seed variables (mass and power consumption for similar components used in space missions) were available, but only one realization for each of them was available. Hence, calibration is redefined in this research as a measurement of the tendency of the expert to over- or underestimate quantities in his/her field of expertise, and it is performed as point estimation.

Ideally, to perform elicitation for this research, experts should be able to model the probability densities for the quantities of interest. This elicitation process implies two abilities: the ability to model phenomena probabilistically and the ability to estimate quantities in a certain area of expertise. In the Cooke Classical Model, these abilities are modelled at the same time in the calibration process. In this research, due to the absence of seed variables, the "traditional" calibration assessment is split into two components: Probabilistic Thinking (which measures ability 1) and Calibration (which measures ability 2). This approach is suited for this investigation and presents some advantages:

- The first part of the assessment can be developed independently, and can be applied to different elicitation processes.
- The second part of the assessment can be performed using point estimations instead of probabilities.
- The two phenomena are modelled separately, discovering information about experts that can improve future estimations. In fact, if an expert is very good

in point estimation (due to his/her experience) but not as much in probability estimations, the analyst can train him/her, and consequently improve the estimation.

Calibration in this research is performed as point estimation. The following is an example of a calibration question:

> A monolithic parabolic antenna in S-Band has a gain of 32 dB. What is the expected mass value?

The reference value for the antenna mass is 8 kg, since the antenna has been used in a previous space mission. The knowledge of the real value helps the analyst to quantify an expert's over- or underestimation. As an output of calibration, two quantities are computed:

- Calibration coefficient: It indicates how distant the expert is from the real estimate, and it is applied in the elicitation to "correct" the expert estimate. If e_i is the expert's estimated value, and r_i is the real value for question i, the coefficient is:

$$c_i = \frac{(e_i - r_i)}{r_i} \tag{5.2}$$

  A positive value indicates overestimation, a negative value indicates underestimation.

- Calibration score: It is used to aggregate expert opinions. Defining e_i the expert estimated value, r_i the real value for any i questions, and m the total number of questions for Part 2, the expression is:

$$S(e,r) = 100 \cdot \left(1 - \frac{1}{m} \cdot \sum_{i=1}^{m} \left(\frac{e_i - r_i}{r_i}\right)^2\right) \tag{5.3}$$

e_i are always defined as positive (this definition is specific to the application considered, since mass and power consumption are positive quantities), and limited up to $2r_i$, to obtain a reasonable estimate and to allow the mathematical tractability of

the coefficient and of the score. In this way, it is possible to verify that the score is always limited between 0 and 100. Also in the case of calibration, subindices are possible. They represent specific cases of underconfidence or overconfidence that the expert can show in sizing specific components. In the specific case of this research and of communication system design, calibration subindices are:

- Antenna mass calibration index (ANC)
- Transceiver mass calibration index (TMC)
- Transceiver power calibration index (TPC)
- Amplifier mass calibration index (AMC)
- Amplifier power calibration index (APC)
- Other components (diplexers, switches, cables) calibration index (OC)

Subindices are used to shift the distributions as explained in the next section. There are 17 questions in Part 2. The mapping between questions and calibration subindices is shown in Table 5.2. The calibration part of the interview lasts approximately 20 minutes, and it is specific to communication system design, requiring real expertise in the field. For this reason, Part 2 was performed only by professional engineers.

5.2.3 Part 3: Expert Elicitation (Statistical Link Analysis and Quantile Method)

The third part of the methodology is the elicitation of the uncertainty quantities of interest. In the literature, approaches to elicitation can be classified as Bayesian or non-Bayesian. The Bayesian approach is particularly useful whenever an a priori distribution is already known, or whenever new information is introduced into the system to update statistics. In this research, due to the absence of the a priori distribution, the non-Bayesian approach is applied.

Non-Bayesian elicitation approaches differ on the basis of the summaries assessed.

Table 5.2: Subindices analyzed by each of the questions for Calibration.

Question	ANC	TMC	TPC	AMC	APC	OC
1	x					
2	x					
3	x					
4	x					
5	x					
6	x					
7	x					
8	x					
9	x					
10		x	x			
11		x	x			
12		x	x			
13				x	x	
14				x	x	
15						x
16						x
17						x

Summaries are sets of parameters provided by the experts to model the distributions. There is no unique way to model the same distribution [34]. Hence, results of the elicitation vary for the different summaries used: some experts are more familiar with mean or variance, while others prefer quantile probabilities. However, if summaries are adequate and well chosen [100], they should identify the expert's distribution. Previous research in selecting summaries suitable for elicitation showed conflicting results ([100], [34]). Hence, in this research two techniques are tested simultaneously. In the interview, the expert provides two sets of summaries, and the correspondent distributions are calculated. Finally, the expert is asked to select which distribution most closely reflects his/her belief. The following subsections describe the sets of summaries and the interview structure.

First Set of Summaries: Bounds and Shapes (Method Derived from Statistical Link Analysis)

In this case, the summaries assessed are bounds and shapes of the distributions. This approach is derived from statistical link analysis, which was developed at Jet Propulsion Laboratory by J. Yuen [75] in 1975. The motivation for the development of the technique was uncertainty in telecommunication link budget. Telecommunication link budget is essentially a multiplication and division (which becomes summation in the logarithmic domain) of terms that quantify signal-to-noise ratio at the receiver. This calculation is affected by the fact that all the terms (i.e., power of the transceiver, antenna gain, noise sources, etc.) have fluctuations affecting the final value of signal-to-noise ratio. Traditionally, to take into account these fluctuations, static margins are applied. However, to develop a more precise statistical characterization of fluctuations, statistical link analysis was developed. The technique is a simple characterization of the fluctuations in which each of the experts makes assumptions on the most adverse and favorable case and on the shape of the distribution. Then, each parameter is described by a probability function and the mean and variance are derived. Finally, making the assumption that if many probability distributions are added together, the cumulative tends to be a normal distribution (Central Limit Theorem), the mean and the variances are summed and the final result gives not only the signal-to-noise-ratio but also the variance around the mean value previously calculated.

It is interesting to notice that there is a parallel between the problem of the statistical link analysis and the problem of this research. In both cases:

- The problem is of uncertainty characterization: in the case of statistical link analysis the problem was quantifying the fluctuations in the different link budget metrics, while the goal of this research is quantifying the fluctuations in mass and power of different components.

- The state of the art solution before the implementation of statistical link analysis was the application of static margins on the link budget, and in this research

the state of the art solution is the application of contingencies at the component level and at the system level.

- The amount of data to characterize fluctuation is limited in link analysis as well as in the case of this research.

Given the similarity between the two problems, the link analysis methodology has been selected as an expert elicitation approach for statistical risk estimation.
According to the statistical link analysis approach, the expert assigns:

- Design value: a priori estimate of the parameter, the center of the probability distribution.
- Favorable tolerance: lower bound of the probability distribution.
- Adverse tolerance: upper bound of the probability distribution.
- Shape of the probability distribution: uniform, triangular or normal.

Unlike in statistical link analysis, the design value is already given to the expert during the elicitation in this research. This choice is because elicitation aims to identify whether the expert can assess fluctuations around an initial estimate, allowing identification of too conservative or too risky initial estimations.
Regarding distribution shapes, "uniform" represents the classical non-informative distribution, since every possible value on the interval is assumed to be equally likely. Jacobson [67] proved that this distribution achieves the maximum variance across the bounded unimodal distributions, hence being a preferred choice whenever the expert is confident in the adverse and in the favorable case, but does not have any other additional information on the design value.
The mathematical formulation (L is lower bound, U is upper bound) and an example (Figure 5-3) are as follows:

$$f(x) = \begin{cases} \frac{1}{U-L} & \text{if } L < x < U \\ 0 & \text{otherwise} \end{cases} \tag{5.4}$$

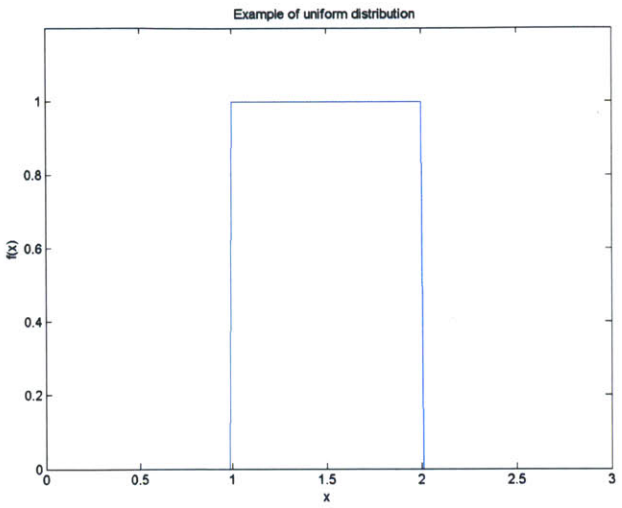

Figure 5-3: Example of uniform distribution. Uniform distribution achieves the maximum variance across the bounded distributions [67]. Hence, it is well suited in case of lack of confidence in the estimate.

Triangular distribution is a bounded distribution and is useful whenever the expert has a certain confidence in the design value.

The expression (D is the design value, L is lower bound, U is upper bound) and an example (Figure 5-4) are as follows:

$$f(x) = \begin{cases} \frac{L-(D-x)}{L^2} & \text{if } L < x < D \\ \frac{U-(x-D)}{L \cdot U} & \text{if } D < x < U \\ 0 & \text{otherwise} \end{cases} \tag{5.5}$$

Normal distribution is selected whenever the expert has great confidence in the design value. It requires a quantification of the confidence to generate the variance. For that purpose, the author defines the parameter y (to bound the distribution around the adverse and favorable values $y > 8 - 10$). The mathematical formulation (D is the design value, L is the lower bound, U is the upper bound) and an example (Figure 5-5) are as follows:

$$f(x) = \frac{1}{\sqrt{2 \cdot \pi \cdot \sigma^2}} \cdot e^{-\frac{(x-\mu)^2}{2 \cdot \sigma^2}} \tag{5.6}$$

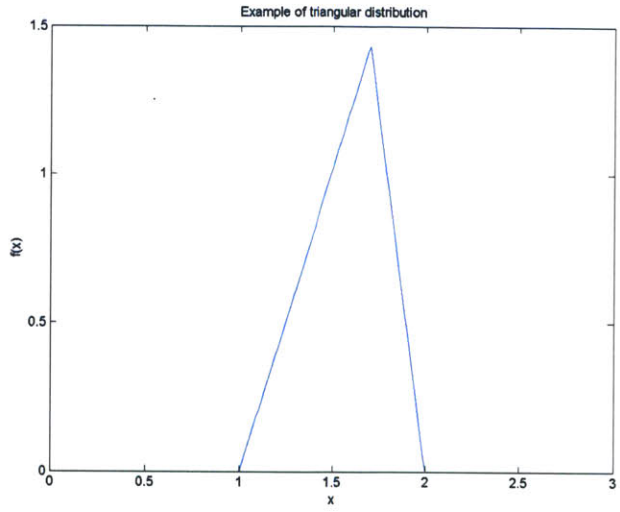

Figure 5-4: Example of triangular distribution. This distribution represents a case in which the expert has a certain confidence in the design value.

where

$$\mu = D \tag{5.7}$$

$$\sigma^2 = \frac{(U-L)^2}{y} \tag{5.8}$$

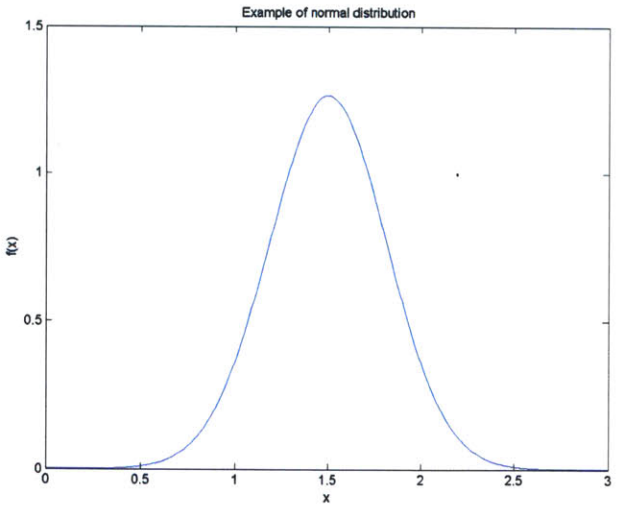

Figure 5-5: Example of normal distribution. This distribution is selected whenever an expert has high confidence in the design value.

The following section describes the second set of summaries.

Second Set of Summaries: Quantiles

When quantile probabilities are applied, the analyst asks the expert to determine the value x such that:

$$P(X < x) = p \tag{5.9}$$

where X is the quantity for which the expert estimates probabilities, and p represents the known probability values. The values of p generally used are: 25%, 50%, 75%. They are selected because for an expert it is easier to think about the distributions in term of bisection problems. However, the quantile approach used in this research is different, since the author wanted to use the same shapes of the distributions (uniform, triangular, normal) applied for statistical link analysis. Hence, the quantiles used are: 50% and 16% (1σ interval from the mean). As quantiles are assessed, the expert selects the shape of the distribution across the three alternatives used for link analysis. In the case of triangular distribution the expert has to provide an additional parameter (lower bound or upper bound) to estimate the distribution.
Once the elicitation process is completed, the statistician is able to model the distribution in the following way:

- Uniform distribution: The 50% quantile becomes the center of the distribution, and the 16% quantile is used to identify where the lower bound (L) and upper bound (U) are located. The process is done through the following calculations (x_{50} indicates the 50% quantile and x_{16} indicates the 16% quantile):

$$L = x_{50} - \frac{50 * (x_{50} - x_{16})}{34} \tag{5.10}$$

$$U = x_{50} + \frac{50 * (x_{50} - x_{16})}{34} \tag{5.11}$$

- Triangular distribution: The 50% quantile becomes the center of the distribution, and the 16% quantile is used to identify the slope of the triangle together with the lower bound.

- Normal distribution: The 50% quantile becomes the mean. Since the 16% quantile is at 1σ distance from the mean, sigma becomes:

$$\sigma = x_{50} - x_{16} \tag{5.12}$$

The 16% area quantile in the normal distribution is illustrated in Figure 5-6.

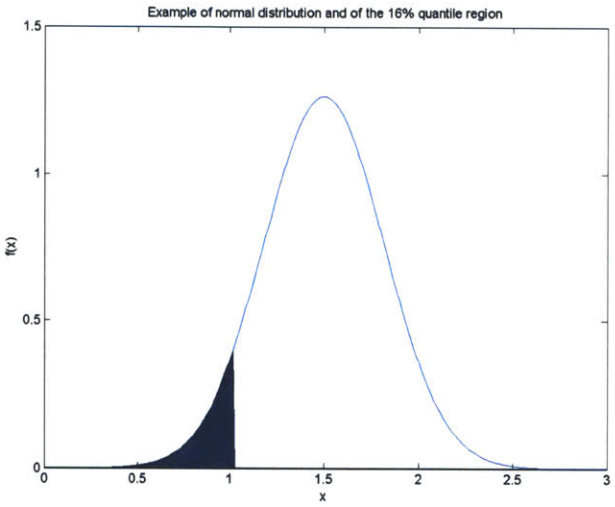

Figure 5-6: Example of normal distribution with the 16% quantile.

The next section describes the structure of the elicitation process and the software used.

Expert Elicitation Interview

In this part of the interview, the expert analyzes test cases by providing his/her opinion on the probability that initial estimations for mass and power consumption will fluctuate during the temporal evolution of the design. Initially, the test case is briefly illustrated to the expert, as shown in the following example:

> Mission description: CASTOR is a small ESPA class satellite for LEO orbit. To perform the mission required, it needs to be equipped with a communication system in the S-Band with 6 dBW EIRP. The selected

components are a 6 dB patch antenna and a COTS transceiver with an estimated transmitting power of 1 W.

The analyst will ask the expert to model the probabilities related to the quantities of interest. For example, a question regarding CASTOR antenna mass would be the following:

> Question: The antenna is a 6 dB patch with an initial estimated mass (at PDR) of 0.5 kg. What do you think are the lower and upper bounds for the mass distribution? Which shape of the probability distribution would you choose to describe the mass distribution? What are the 50% and the 16% quantile probabilities of the mass probability distribution?

A Matlab script has been developed to implement the interview. Figure 5-7, shows the graphical user interface and the summaries required from the expert. The

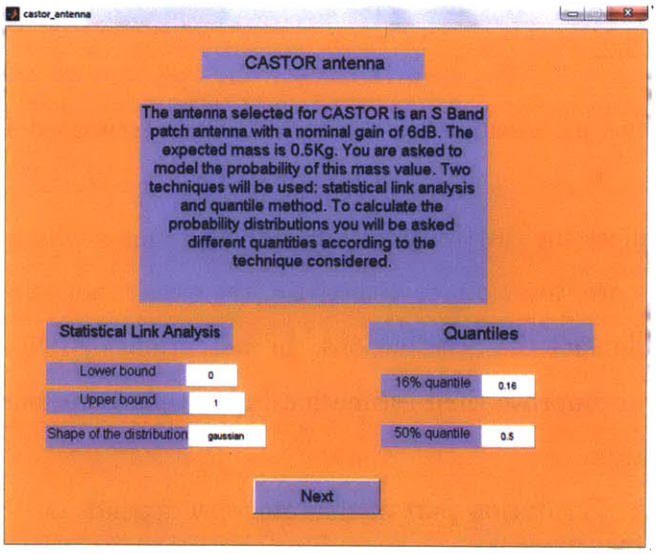

Figure 5-7: Summaries required for the antenna expert elicitation process. The summaries are used to compute the distribution.

software elaborates the expert's inputs and four plots are created:

1. Distribution calculated through the statistical link analysis technique (no calibration): The distribution is computed using the summaries assessed by the expert through statistical link analysis.

2. Distribution calculated through the quantile technique (no calibration): The distribution is computed using summaries assessed by the expert through the quantile probability elicitation technique.

3. Distribution calculated through the statistical link analysis technique with calibration: The distribution is computed using the summaries assessed by the expert through statistical link analysis. However, the bounds of the distribution are shifted by an amount which is a function of the calibration coefficient (c_i) computed in Section 5.2.2. In this way, if an expert has a tendency to overestimate quantities, the distributions are scaled to compensate for that effect.

4. Distribution calculated through the quantile technique with calibration: The distribution is computed using summaries assessed by the expert through the quantile probability technique. Similar to the previous case, the quantiles are modified by an amount indicated by the calibration coefficients (c_i) calculated in Section 5.2.2.

Figure 5-8 shows an example of the distributions estimated for the CASTOR antenna.

The expert visualizes the distributions and he/she declares whether one or more of these distributions are satisfactory. Otherwise, the expert can select to re-input the summaries to recalculate the distributions. In fact, studies ([100], [34]) have shown that experts tend to improve their estimation in multiple iterations with the help of graphical visualization.

The duration of the Elicitation part of the interview depends on the number of test cases and their complexity. In this specific interview, two space missions developed at the Massachusetts Institute of Technology, CASTOR and HETE-2, have been studied for a total of six components. Table 5.3 shows a summary of the test cases.

The duration of this part of the interview is approximately 30 -40 minutes. The duration of the total interview (all the three parts) is approximately 1 hour and 10 minutes. The following section describes the approaches applied to aggregate multiple expert opinions into a unique estimate.

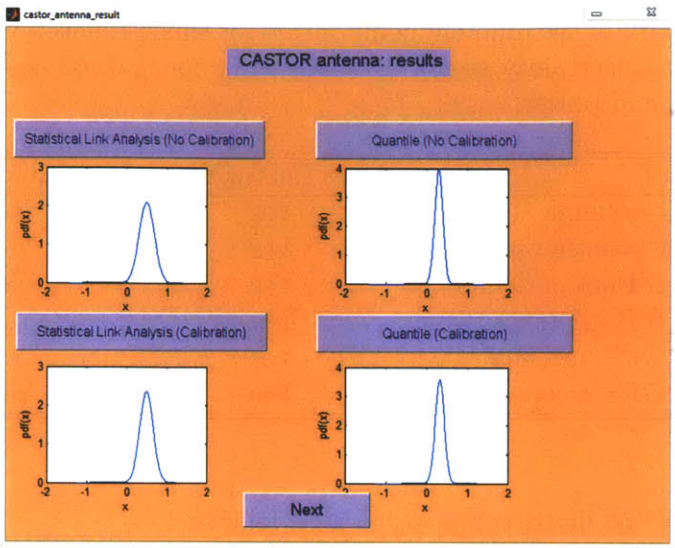

Figure 5-8: Example of the probability distributions generated in the expert elicitation process for the CASTOR antenna. The four plots show the four possible distributions.

5.2.4 Part 4: Aggregation of Probability Densities from Multiple Experts

Expert aggregation allows the analyst to aggregate information from multiple experts. In terms of aggregation techniques, the literature categorizes them into two main categories: group interaction and mathematical composition. Group interaction requires face-to-face meetings among the experts; this class of techniques can be challenging for the logistical reasons of time availability and geographic diversity of the experts. Moreover, the undesirable effects of personalities and reputations are unavoidable in a group interaction setting. Finally, different authors [100], [112] have not observed improvement in performance using group interaction with respect to mathematical aggregation. For these reasons, this research is focused mostly on mathematical aggregation.

Regarding mathematical aggregation, different techniques have been developed in the literature [100], including linear composition, exponential composition, and Bayesian composition. In the specific case of this research, the method selected is linear composition due to its marginalization property [100].

Table 5.3: Summary of the different components for which Elicitation was performed. For antennas the uncertainty metric is mass, while for transceivers the uncertainty metrics are mass and power.

Test Cases	Mass fluctuations	Power fluctuations
CASTOR antenna	Yes	No
CASTOR transceiver	Yes	Yes
HETE-2 S-Band antenna	Yes	No
HETE-2 S-Band transceiver	Yes	Yes
HETE-2 GPS antenna	Yes	No
HETE-2 GPS receiver	Yes	Yes

The expression of the linear composition is as follows:

$$f_{tot}(x) = \sum_{i=1}^{n} w_i \cdot f_i(x) \tag{5.13}$$

where $f_{tot}(x)$ represents the aggregate distribution, $f_i(x)$ corresponds to each single expert distribution, w_i is the weight of the distribution for expert i, and n is the total number of experts considered.

The generation of weights can be a challenge, since different weights generate very different aggregate distributions. To this end, the initial part of the methodology is focused on identifying ways to score expert performance. In this final part of the methodology, scores are applied to aggregate the distributions.

Four possible weighting schemes are considered:

- Equal weights: All experts are weighted equally regardless of their sensitivity to biases and regardless of their calibration.
- Quality weights: Experts are weighted on the basis of their ability to think probabilistically. Scores are computed through Part 1 of the interview. (The coefficient generation process is described in the following part of this section).
- Calibration weights: Experts are weighted on the basis of their calibration. Scores are computed through Part 2 of the interview. (The coefficient generation process is described in the following part of this section).

- Quality and calibration weights: Experts are weighted on the basis of their ability to think probabilistically and on the basis of their calibration. The scores used are a combination of the results of Part 1 and Part 2 of the interview. (The coefficient generation process is described in the following part of this section).

The weights used for expert composition need to sum to 1 in order to obtain an aggregate distribution that is a proper density. The determination of the weights is straightforward in the case of equally weighted experts. Each distribution receives a weight that is $\frac{1}{n}$ where n is the number of experts. In the other cases, the process is different. In the case of the quality-weight aggregation, the weights are computed using scores obtained from Part 1 of the interview. The conversion of scores into aggregation weights follows these steps:

- The scores are used to determine a ranking across the experts: the expert with the highest score from Part 1 becomes the first ranked expert, the expert with the second highest score becomes the second ranked expert, and so forth.

- A weight is assigned to each expert according to a bisection rule: the first-ranked expert receives a weight of 0.5, the second expert receives a weight of 0.25, the third expert receives a weight of 0.125, and the last two ranked experts receive a weight that is equal to $\frac{1}{2^{n-1}}$.

The determination of weights for the calibration composition follows the same approach, but the scores to rank the experts are computed in Part 2 of the methodology. For the quality and calibration composition, the approach is the same but the scores are the average between the quality and calibration scores.

The assignment of weights through a bi-section method is different from assigning weights proportional to the experts' scores. The motivation for this choice is the necessity to significantly differentiate the contribution of the experts across the estimation. In fact, as the number of experts increases, if the weights are assigned proportionally to the interview's scores, they tend to not appropriately reward the top-performing experts. To better explain this concept, consider an example with only three experts. Expert 1 scores 100 in the interview, while Expert 2 scores 50

and Expert 3 scores 49. If the analyst assigns weights as a proportion of the interview's scores, the weights are approximately 1/2 for Expert 1 and 1/4 each for Experts 2 and 3. If the analyst uses a bisection rule, the ranking results in Expert 1, followed by Expert 2, followed by Expert 3. The correspondent weights are 1/2, 1/4, and 1/4 respectively. In this particular case, both assignments give the same results. However, if we add a fourth expert with a score of 25, a proportional assignment of weights results in Expert 1 receiving 4/9, Experts 2 and 3 each receiving 2/9, and Expert 4 receiving 1/9. In this case, it is possible to notice that the weight of the top expert (Expert 1) decreases from 1/2 to 4/9. As the number of experts increases, the weight of Expert 1 will further decrease moving closer to the scores of the other experts, and the elicitation will lose the initial information of Expert 1 being a top performer among his/her peers. On the other hand, the bisection rule will keep the score of expert one at 1/2, and the impact of increasing the number of experts will be reflected on the experts who achieve the lower performance. This approach is empirical and not derived from the Cooke Classical Model. However, as argued by Cooke [34]:

> The classical model satisfies necessary conditions for rational consensus, but is not derived from first principles, and other weighting schemes may perform as well or better.

Moreover, the approach presented in this thesis seems to be a compromise between a proportional weight assignment [100] in which all the experts are considered in the final estimate, but the top experts are not properly rewarded, and a binary (0 or 1) weight assignment based on hypothesis testing [34] in which only the top experts are considered.

A last consideration on composition is related to the Delphi method. The Delphi [34] method is an aggregation approach developed in 1944, which implements an iterative elicitation across a group of experts. In its basic formulation, the Delphi method is composed of a questionnaire sent to the experts. The set of responses is elaborated by the analyst who computes the mean, median, and lower and upper 25% interquan-

tile ranges for any of the questions. Results are sent back to the experts, and the analyst asks them if they wish to revise their initial prediction. The experts whose answers are outside the interquantile range are asked to provide motivation for their choices. The revised responses are processed in the same way and the information is sent back to the experts for another round of revision. The process iterates for three or four rounds. Generally, answers on the final round show a reduction of uncertainties with respect to the first. This method was refined across the years until some researchers (Sackman, Gustafson, Seaver) started questioning the validity of this approach. According to Sackman [109] experts are not treated equally: the experts whose predictions fall inside the interquantile range are "rewarded" while the others are punished. Gusstafson [58] compared the Delphi aggregation to different group techniques and discovered that the Delphi technique was the worst-performing technique. Seaver [112] compared group interaction and mathematical aggregation and found that group interaction generally increases confidence, but not necessarily the correctness of the assessments.

For these reasons Delphi is almost never applied in current elicitation processes, and it was not selected as an approach to compose expert assessments in this research.

5.2.5 Summary

Figure 5-9 provides a summary of the methodology. The four parts can be summarized as follows:

- Part 1 (Probabilistic Thinking): This part analyzes biases and heuristics in expert elicitation generating the quality score applied to expert aggregation.

- Part 2 (Calibration): This second part quantifies an expert's tendency to over- or underestimate quantities in his/her technical area of expertise. Unlike the Cooke Classical Model, it is performed as point estimation. This part of the approach generates the calibration coefficient used to shift the distributions in the elicitation part, and the calibration score applied to expert aggregation.

- Part 3 (Elicitation): Experts estimate the quantities of interest using two possible approaches: quantile and statistical link analysis.
- Part 4 (Expert aggregation): Different densities estimated by experts are aggregated using scores from Probabilistic Thinking and Calibration.

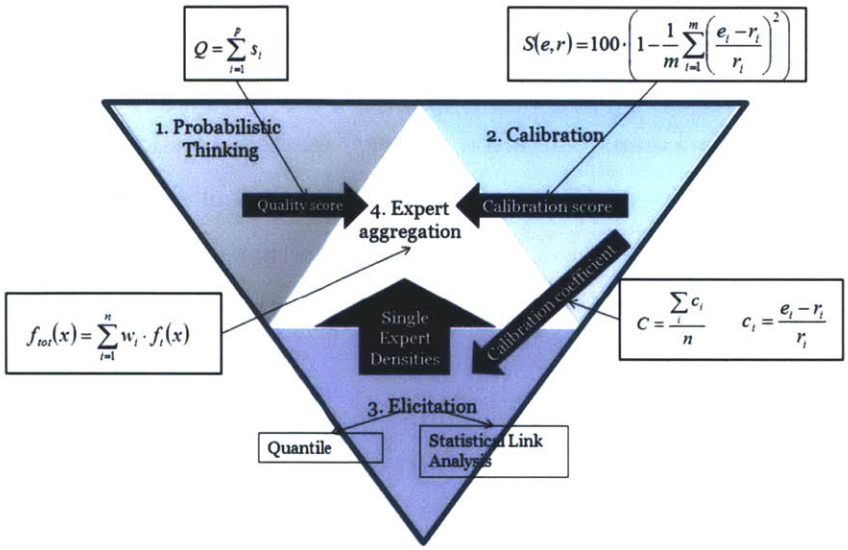

Figure 5-9: Expert triangle: summary of the methodology. Probabilistic Thinking and Calibration generate scores which are used to compose the distributions estimated by each of the experts in the Elicitation part of the interview.

The next section presents the experimental results.

5.3 Experimental Results

The expert elicitation approach proposed was tested through interviews conducted between March and September 2011. The interview structure is articulated in three parts reflecting the first three parts of the methodology. The fourth part of the methodology is applied a posteriori on the basis of the information collected. The interviews involved three populations of subjects: undergraduate students (17 subjects), graduate students (17 subjects), and communication system engineers (8 subjects). The use of students as subjects was limited to Part 1. In fact, due to its generality, Part 1 is suitable to test on different categories of subjects, who may have or not have

experience in communication system design. The testing done on Part 1 on different subjects provided several hints on: sensitivity to biases and heuristics for different categories of subjects, and correlation between biases and training/education. Moreover, testing Part 1 on students provided an initial baseline to identify average scores, above-average scores, and below-average scores. Also, testing Part 1 on different populations and observing similar results across the populations confirmed the validity of this part of the interview. Finally, testing Part 1 on students allowed the author to identify unclear questions that needed to be restated.

Parts 2 and 3 of the interview required experience in communication system design. Hence, 8 engineers with experience in the field were selected for the study.

The interviews lasted between 20-30 minutes for the subjects who participated only in Part 1, and between 70-100 minutes for the subjects who participated in the whole interview.

The interview is compliant with the regulations imposed by the Committee on the Use of Humans as Experimental Subjects (COUHES, Massachusetts Institute of Technology). The analyst who performed the interviews with subjects was properly trained. Moreover, before the beginning of the interview, each subject was briefed by the analyst on the motivation of the interview, and he/she was asked to sign a consent form. Additionally, the interview followed the five steps of the Stanford Research Institute (SRI) assessment protocol (motivating, structuring, conditioning, encoding and verifying) [89], with the conditioning stage being performed as part of the interview. Finally, the privacy of the subjects is preserved since only the interviewer knows the identity of the subjects. The results are described in the following subsections organized according to the four parts of the methodology.

5.3.1 Results for Part 1: Probabilistic Thinking

As mentioned before, Part 1 was tested across the three populations of subjects (undergraduate students, graduate students and engineers), due to the generality of the questions. Figure 5-10 shows the general score. The score was calculated as described in Part 1 of the methodology. The most important results obtained through

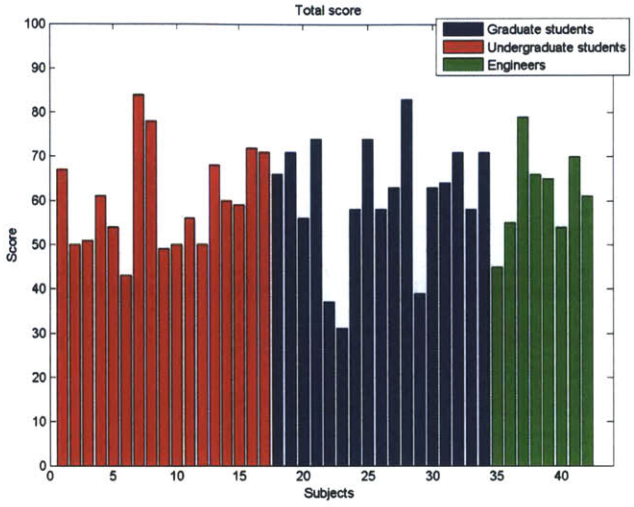

Figure 5-10: Quality score (Part 1) for the three populations. The average score is around 60-61 in each of the three populations.

testing Part 1 across the three populations are as follows:

- **The results are consistent across the three populations:** The average quality score is around 60-61 in each of the three populations. This result provides a baseline for future applications of this part of the interview, because it shows that the interview can capture inner biases regardless of the level of instruction. Hence, for future subject tests, the analysts know that, regardless of the population, a score greater than 60 tends to be above the average.

- **The results provide a method to identify top-performing experts:** Very few subjects obtained a score greater than 80. This result indicates for future reference that a score greater than 80 represents a strong ability to model phenomena probabilistically and a limited sensitivity to the biases mapped.

The sensitivity of the experts to the subindices has been mapped. Figure 5-11 shows the scores for coherence, representativeness, awareness of underlying conditional probabilities, and availability. Figure 5-12 shows the scores for anchoring, hindsight bias, and small sample bias. The color red indicates graduate students, blue indicates undergraduate students, and green indicates engineers.

To interpret these graphs, it is necessary to remember that the scores are "reward-

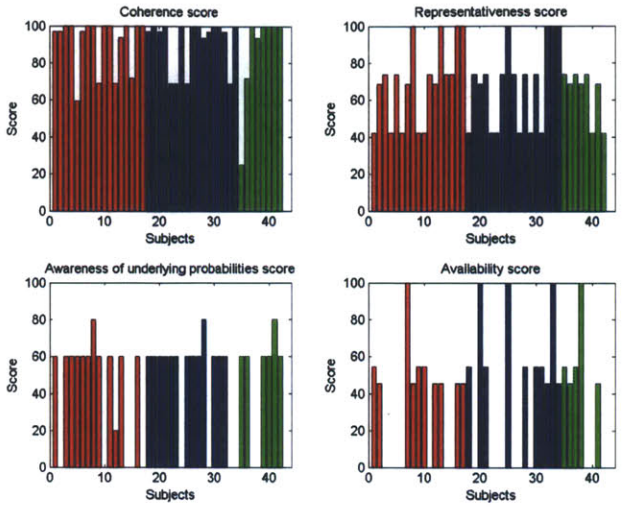

Figure 5-11: Scores for coherence, representativeness, awareness of underlying conditional probabilities, and availability. Subjects exhibit very good coherence and representativeness, while the performance for the other two biases are less promising.

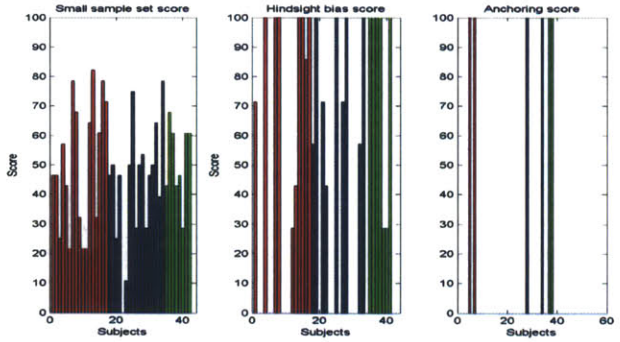

Figure 5-12: Scores for small sample bias, hindsight bias, and anchoring. Performance varies on a case by case basis.

ing scores" meaning that the subjects who showed less sensitivity to a bias received a higher score. Observations on subjects' sensitivities are listed as follows:

- Coherence score is generally very high: Students and engineers show a good understanding of the basic probability laws.

- Representativeness, small sample, and hindsight bias scores show different performance across the subjects. In some cases, like hindsight bias, the performance of the engineers is generally better than that of the students. The representativeness index indicates whether subjects have an understanding that the probability of a certain subset is less than or equal to the probability of the set in which the subset is included. The results for this index are conflicting: some subjects performed very well, others did not. In the latter case, subjects have been biased by additional information which has damaged their judgment. Typical examples of errors in representativeness have been discussed in Section 5.2.

- Awareness of underlying conditional probabilities, availability, hindsight bias, and anchoring scores suffer from limited granularity. Only few questions were used to test the sensitivity to these biases. Hence, subjects who answered to all the few questions correctly received a score of 100, and subjects who answered all incorrectly received a score of zero. Future research will focus on expanding this part of the interview, including more questions to increase the granularity/resolution of the analysis.

During this part of the interview, we asked each subject to list his/her experience/training in probability. We then quantified this exposure to probability with a number indicating the number of classes and/or research projects developed involving probability. For example, if a subject took one probability class, and he/she developed only one research project involving probability, then his/her experience score was equal to two. A scatter plot of quality score vs. experience in probability is given in Figure 5-13. The results show that quality score and training are uncorrelated at least for the students. A limited trend can be observed in the engineers. One of the problems with this analysis is that it weighted all experience equally, while certain classes/research can impact how people develop their probabilistic skills differently. Future research will focus on improving the way in which we quantify the experience in probability.

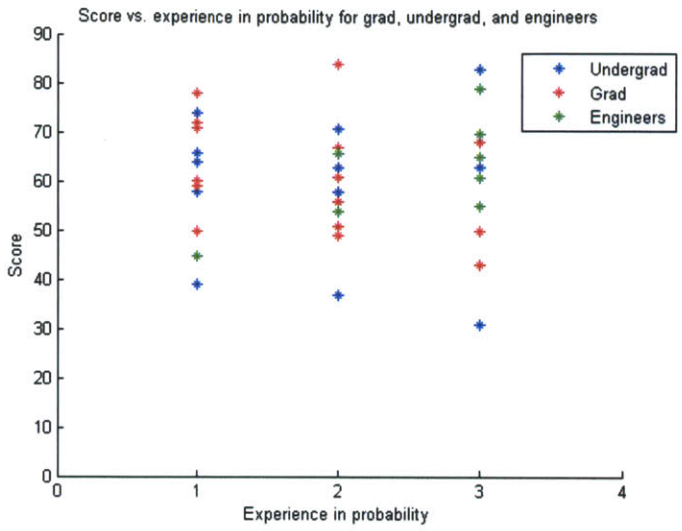

Figure 5-13: Quality score (Part 1) with respect to experience in probability: the two aspects seem to be uncorrelated.

Some other findings from Part 1 of the interview are the following:

- Overthinking: Some subjects tend to overthink about each question by trying hard to identify what the interviewer was looking for. In some cases, this phenomenon created more stress and led to wrong answers.

- Lack of courage in the answers: Some subjects successfully identified the biases and mentioned them during the interview, but then they picked the wrong answers. When interrogated on the motivation for their final selections, they replied that they did not have enough proof/confidence to support their initial intuition. In other words, they trusted their biases more than their intuition. A typical example is question 13 (see Appendix B), in which a list of singers is shown and the person has to identify if the list is composed of more male singers or more female singers. On purpose, the author select female singers who were more famous than their male counterparts, but they were less in number. Some subjects recognized that the female singers were more famous, but they felt like they did not have enough proof to argue that there were more males than females, hence they picked the wrong answer.

- Coaching: In some cases coaching helps while in some others it did not improve the final score.

In conclusion, results from Part 1 reveal that the interview works uniformly across the populations, showing that it is a valid tool to assess Probabilistic Thinking. For this reason, the analyst can trust the scores assigned to the real experts (engineers), which are used later in the aggregation part of the methodology. Moreover, Part 1 is independent of the type of elicitation performed. Hence, the results obtained from Part 1 can be used as a reference for other expert elicitation processes that use Part 1 as their tool to quantify Probabilistic Thinking.
The next section describes the results of the Calibration part of the interview.

5.3.2 Results for Part 2: Calibration

The second part of the interview aims to identify whether the experts tend to over- or underestimate quantities of interest. Due to the unavailability of seed variables with measured probability densities, this part of the interview is structured as a point estimation, under the assumption that the expert's ability to model probabilistic phenomena has already been measured in Part 1. In this part of the interview, expertise in communication system design is required. Hence, only engineers have been used as subjects for Part 2. The results of Part 2 include two types of measurements:

- Calibration coefficient: This number can be positive or negative and it expresses the tendency of the expert to over- or underestimate quantities of interest. It is used to shift probability distributions in the Elicitation part of the interview.

- Calibration score: This number is a positive quantity used to aggregate experts' distributions.

The interview included 17 questions that asked the experts to estimate mass and power consumptions of communication system components used in space missions. The calibration coefficient is calculated for all the 17 questions, and Figure 5-14 shows the average for each expert.

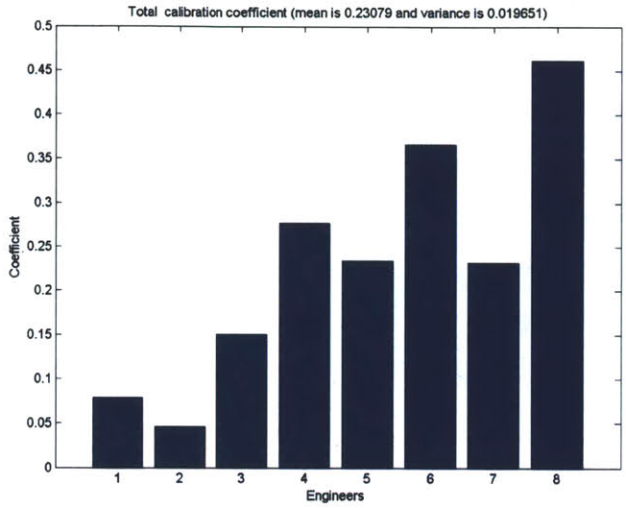

Figure 5-14: Average calibration coefficient for each of the 8 experts interviewed. Experts tend to overestimate more than to underestimate.

All coefficients were positive. On average, experts tend to overestimate more than to underestimate. This result is reasonable, since engineers in spacecraft design are generally conservative in their estimations.

Figure 5-15 plots the calibration score. In this part of the interview the experts were

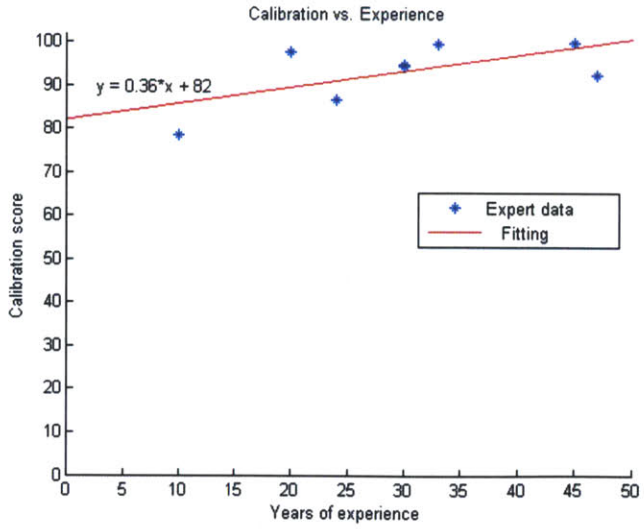

Figure 5-15: Calibration score vs. experience in communication system design. The slope of the trend is smaller than expected.

asked to quantify their years of experience in communication system design. Figure 5-15 shows calibration score vs. experience in design. Experience in communication system design appears to increase the calibration score, but the slope is smaller than expected. This result can be caused by experts having more experience in designing a specific component with respect to the others. For example, while they may exhibit good performance in antenna mass estimation, they may exhibit poor performance in transceiver mass estimation.

Calibration score is used to compose multiple probability densities, which are generated in the Elicitation part. The next sections present these results.

5.3.3 Results for Part 3: Elicitation

In this part of the interview, engineers worked on test cases (space missions) to estimate the probabilities of interest. In the case of this research, probabilities express the likelihood that estimates for mass and power consumptions of different components will fluctuate during the evolution of the design. Due to the technical expertise required, this part of the interview has also been performed only with engineers.

The test cases used (Table 5.3) are missions in the final stage of design or that have already launched. Hence, for these missions data of the initial stage of the design (Preliminary Design Review [PDR]) and of the final stage of the design (Critical Design Review [CDR]) were available. The missions were developed in university, and the experts interviewed had never worked on these missions nor had any information on the final design, guaranteeing the estimates were not biased by any previous knowledge.

The test cases provide a tool to validate the performance of the expert elicitation approach: the analyst can compare expert densities with the PDR design value, and assess whether the experts were able to identify if the initial estimates were over- or underestimated. In this way, the expert elicitation for design risk analysis can become a valuable tool for future proposed space missions, as engineers have to quantify design risks.

The results of the Elicitation are split between this section and the next. This section

summarizes general results of the distribution types selected by the experts, and the number of iterations. The next section discusses the probability densities obtained and the overall performance of the approach.

Regarding the types of distributions selected, experts were asked to model the probability densities using three possible shapes: uniform, triangular, and normal. Figure 5-16 shows the average percentage of the shapes selected by the different experts.

Triangular is the preferred shape (45%). Also normal distribution is selected in

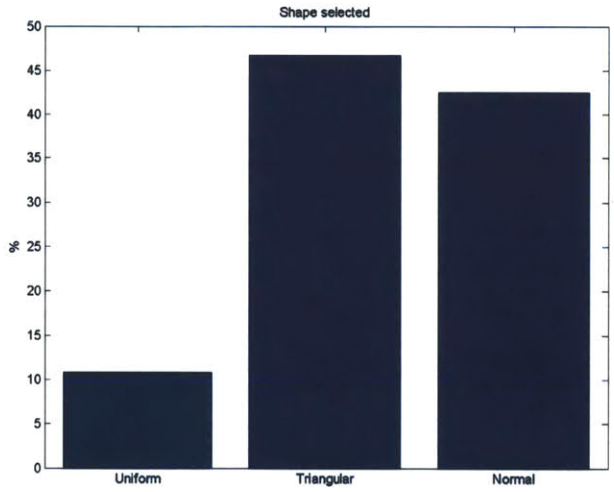

Figure 5-16: Shape selected for the probability densities estimated. Triangular is the preferred shape.

many cases. Interactions with experts show that experts tend to prefer asymmetric distributions. In fact, in most cases of engineering design risk analysis, the uncertain quantities of interest (mass and power consumption) tend to increase more than decrease. Hence, an asymmetric distribution is the preferred choice. Figure 5-17 shows a breakdown of the data for the different experts. Expert 2 is the only subject who selected uniform, while most of the other experts alternated between triangular and normal.

Another analysis of experts' data identifies their preferences with respect to the elicitation techniques proposed. As described in the methodology section, as experts estimate the summaries for the uncertain quantities of interest, the graphical

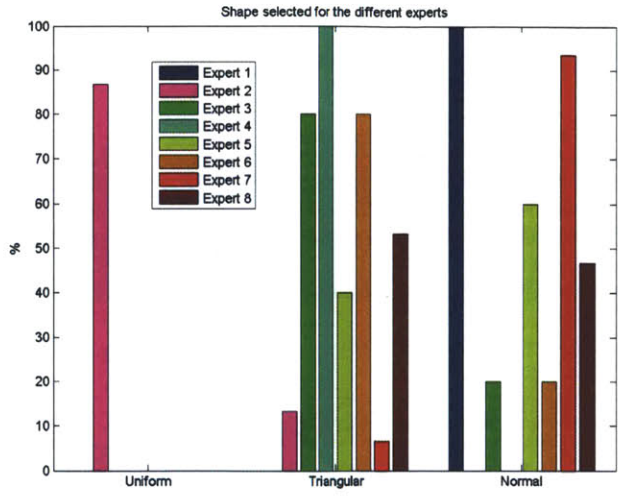

Figure 5-17: Shape selected for the probability densities estimated: breakdown for different experts.

tool shows them four possible statistics (statistical link analysis without calibration, statistical link analysis with calibration, quantile without calibration, quantile with calibration). Experts must select the distribution that most closely represents their belief, or they can reiterate the process by changing the parameters and the shape of the distribution. Figure 5-18 shows their preferences in terms of the four elicitation techniques considered.

Generally, they seem to prefer the quantile approach, and the calibrated (shifted) distributions are rarely selected. Experts appear to prefer their own estimate with respect to the ones corrected by the software through the calibration coefficient. Figure 5-19 shows the same data analyzed for each expert. Some experts such as Expert 8 seem to select different techniques in different cases, while others such as Expert 2 always select the same technique.

The last analysis for this section focuses on the average number of iterations. Iterations are the times in which experts decide to re-iterate the process eliciting new summaries, when the densities shown in the graphical tool do not represent their belief. Figure 5-20 shows the average number of iterations.

Since 1 means no iterations, it is possible to conclude that the number of iterations

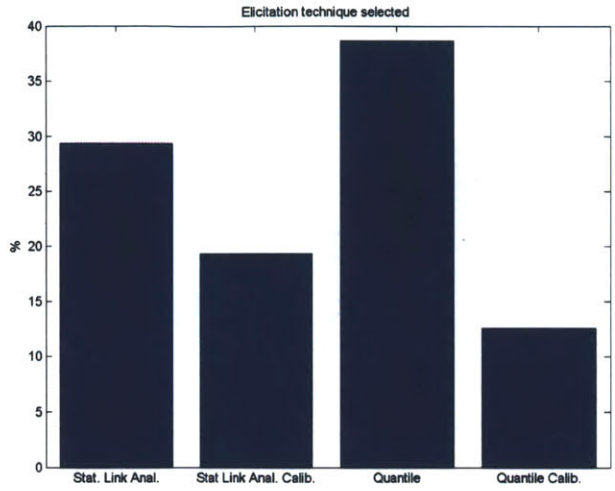

Figure 5-18: Elicitation techniques selected (average across the different test cases). Experts seem to prefer the non-calibrated quantile approach.

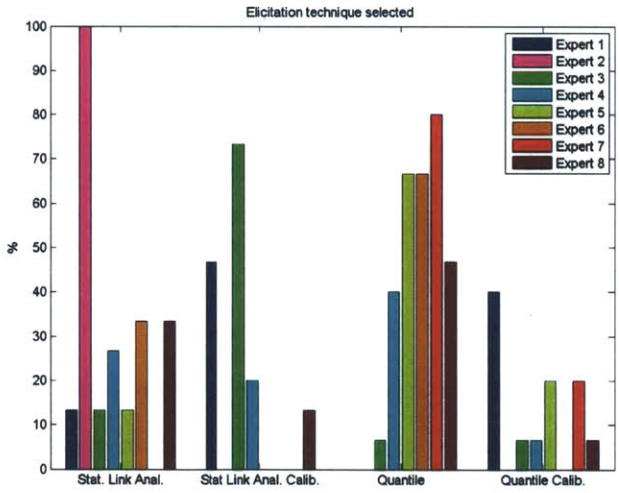

Figure 5-19: Elicitation techniques selected (average across the different test cases): breakdown for the different experts. Some experts tend to select always the same technique.

experienced is generally pretty low. Experts tend to pick a distribution at the first iteration and only in a few cases do they modify their initial selection.

The next section discusses the results for aggregation.

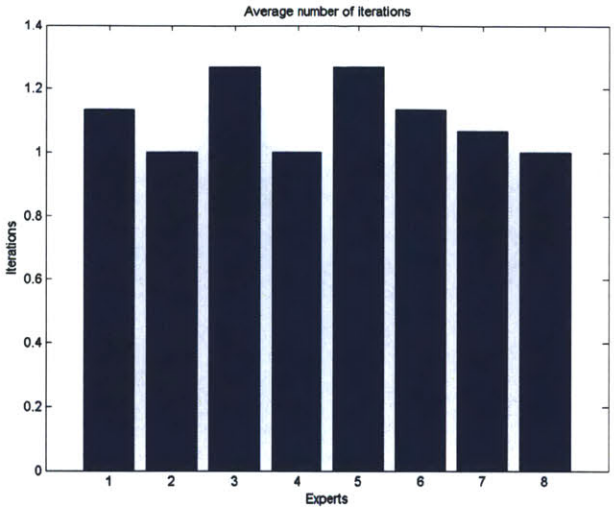

Figure 5-20: Average number of iterations for each expert across the different test cases. The number is pretty low.

5.3.4 Results for Part 4: Aggregation

This section describes the results obtained by aggregating probability densities of different experts. Examples are first used to show the performance of the aggregated densities with respect to the initial (PDR) and final (CDR) values for specific test cases. A comparative analysis is then presented to illustrate performance across the four aggregation techniques.

Probability densities in this research are also represented in terms of tail (risk) functions. Tail functions are the reverse of the more commonly used S curves (also known as cumulative distribution functions), and are applied in risk assessment problems because they provide an immediate quantification of the risk of exceeding a certain value. The advantage of tail functions is that they provide a graphical tool to quantify risk. If the uncertainty metric is on the x axis, the risk of exceeding value x_i is simply the tail function evaluated in x_i ($T(x_i)$).

Figure 5-22 shows an example of a tail function. In this case, the uncertain metric is the mass of a small antenna. This antenna has been used in the CASTOR mission. The initial estimation of the antenna mass performed by the team before PDR was 0.5 kg. The final value of the mass at CDR is known to be 0.1 kg. The test

case was therefore meaningful to validate the expert elicitation approach. During the Elicitation phase, experts were asked the following question:

> The antenna is a 6 dB patch antenna operating in the S-Band with an expected mass value of 0.5 kg. What do you think are the lower and upper bounds for this mass? Which shape of the probability distribution would you choose to describe the mass distribution? What are the 50% and the 16% quantiles of the mass probability distribution?

Figure 5-21 and 5-22 show the results of the different experts in terms of PDF and tails.

The information given by the experts is conflicting. Some (Experts 3, 5, and 7)

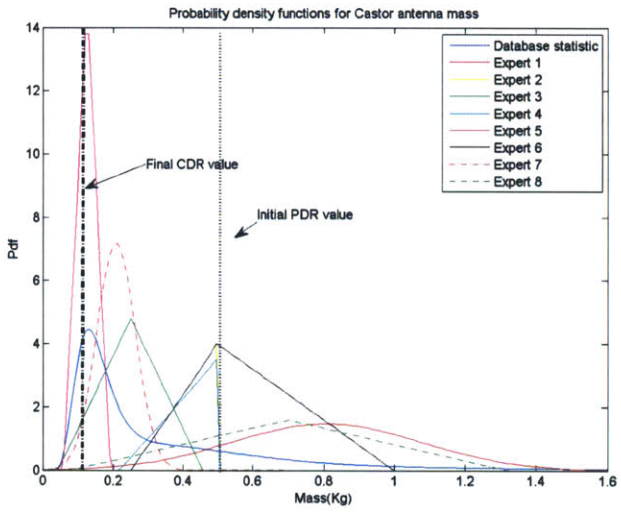

Figure 5-21: Probability density functions for CASTOR antenna mass (different experts). Experts 3, 5, and 7 believe that the initial estimate is too conservative.

believe the initial estimate to be too conservative; their distributions are centered in values closer to the final (real) value of the antenna mass. Also, their estimation of the risk (tail) of exceeding the PDR value is almost zero, confirming that they consider the initial value overestimated. On the other hand, the other experts (Experts 1, 6, 8) are more conservative; their estimates are closer to the initial PDR value, and their tail functions evaluated for 0.5 kg are relatively high.

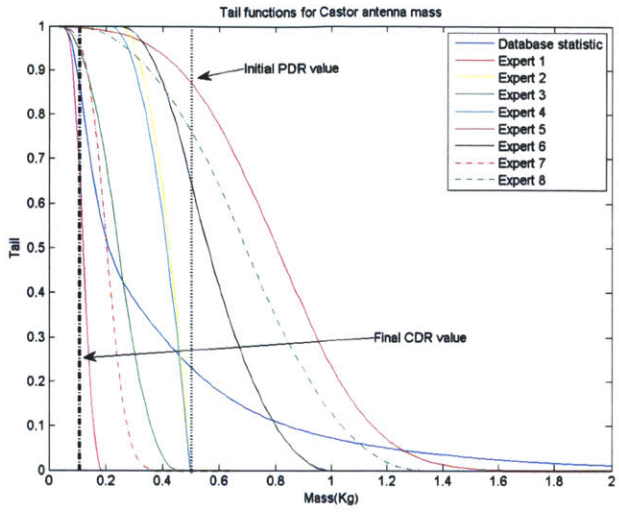

Figure 5-22: Tail density functions for CASTOR antenna mass (different experts). Experts 3, 5, and 7 believe that the initial estimate is too conservative.

In order to resolve these conflicts the densities were composed using the linear aggregation (Equation 5.13) and the 4 weighting schemes discussed in Section 5.3.4. Figure 5-23 and 5-24 show the results.

Since this case is a case of overestimation, the quality of expert assessments can be

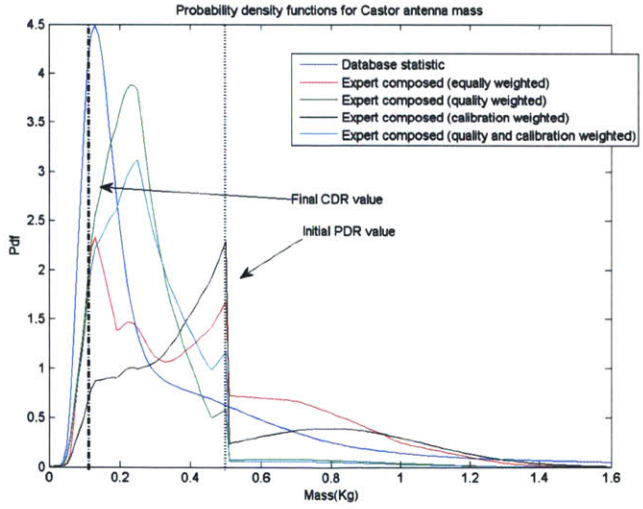

Figure 5-23: Probability density functions for CASTOR antenna mass (experts composed). The composition using the scores from Part 1 and Part 2 improves the risk estimation with respect to the data statistic.

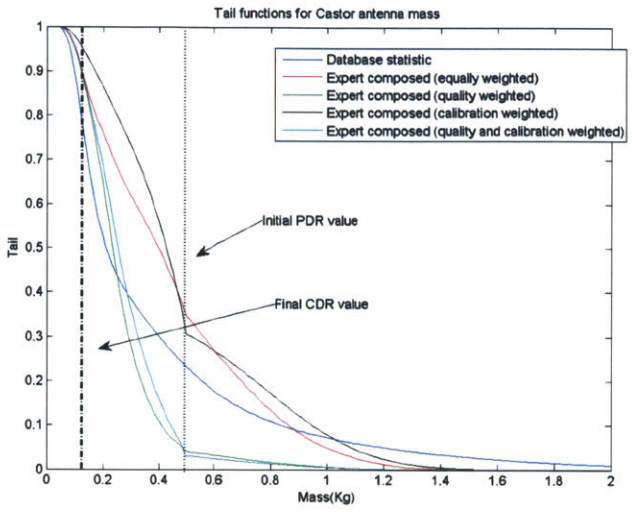

Figure 5-24: Tail functions for CASTOR antenna mass (experts composed). The composition using the scores from Part 1 and Part 2 improves the risk estimation with respect to the data statistic.

measured as follows. For any aggregation technique, the tail function is evaluated at the PDR value ($T(x_{PDR})$): the better prediction corresponds to the smallest value of the tail, hence the smallest risk. In fact, the composed densities should reveal that the initial quantity was overestimated, which means a very limited risk of exceeding the initial value of 0.5 kg.

If we apply this metric, it can be observed that the best estimation is the one given by the aggregation of experts' densities with the scores derived from Part 1 plus Part 2. The worst result is given by the aggregation of experts equally weighted. This result shows that the analysis of biases and calibration improves the estimation. It also shows that the estimation obtained by applying the experts aggregation is consistently better than the one performed through the data statistic.

The performance of the aggregated densities obtained using only one of the two scores varies from case to case across the different test cases analyzed. In the case of the CASTOR antenna, the density generated using scores from Part 1 is better than that generated using only scores from Part 2. In other cases the situation may be different. Figure 5-25 shows another example with a different test case using the HETE-2

mission (GPS receiver, initial mass 2 kg, final mass 1 kg).

Looking at the HETE-2 example and using the same metric described previously

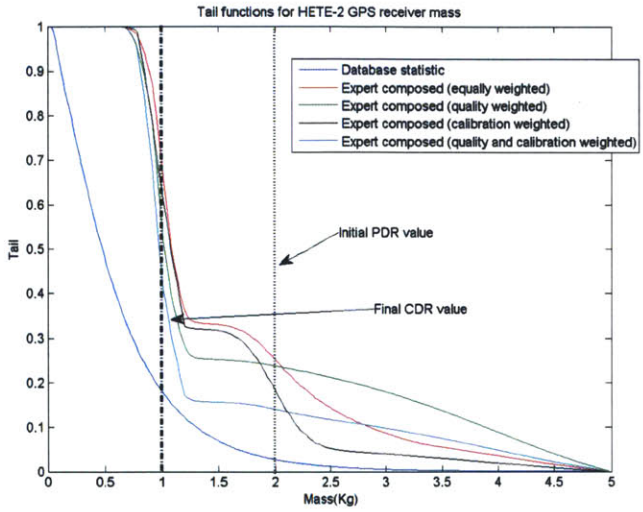

Figure 5-25: Tail functions for HETE-2 GPS receiver mass (experts composed). In this particular case data statistic performs better than expert statistic.

for the CASTOR antenna, the best estimation across the experts is still the one obtained using scores from Part 1 and Part 2. The worst estimation is still the one with the equally weighted experts. However, in this case, the statistic based on data performs much better than all the others by identifying that the initial value was an overestimation (the quantification of risk at PDR using this statistic is much smaller than the ones computed using expert opinion). To elaborate a cumulative analysis of the four aggregation techniques in the different test cases, the following procedure has been applied:

- Separation of over- and underestimation cases.
- Definition of performance metrics for the two cases:
  - Overestimation: The metric is the tail function evaluated at the PDR value ($T(x_{PDR})$). The best estimation is the one with the smallest value of the tail.

- Underestimation: The metric is still the tail function at the PDR value ($T(x_P DR)$). However, since the initial value is underestimated, the experts should be able to identify a very high risk in this solution. Hence, in this case the best estimation is the one that presents the highest value of the tail.

- Assignment of performance scores: On a scale of 1 to 5, the best estimation receives a score of 5, the second a score of 4, and so forth.

Figure 5-26 (cases of overestimation) shows an example of score attributions. The numbers indicate the risks measured at the PDR value, while the colors indicate the score assigned.

In most cases the combined (Part 1 and Part 2) score obtains the best result. A

Case	Risk computed using data statistic	Risk at PDR (equal weights) $T_{Equal\_weights}(x_{PDR})$	Risk at PDR (Part 2-calibration weights) $T_{Calib\_weights}(x_{PDR})$	Risk at PDR (Part 1-Probabilistic thinking weights) $T_{Prob\_weights}(x_{PDR})$	Risk at PDR (Part 1+ Part 2 weights) $T_{Calib\_and\_Prob\_weights}(x_{PDR})$
Castor antenna mass	0.2417	0.3725	0.3506	0.041	0.031
Castor transceiver mass	0.865	0.81	0.81	0.505	0.504
HETE GPS receiver mass	0.029	0.2657	0.19	0.2379	0.14
HETE GPS receiver power consumption	0.39	0.51	0.7	0.17	0.42
HETE S antenna	0.491	0.485	0.484	0.21	0.209

Legend: Best performance (score 5); 2nd best performance (score 4); 3rd best performance (score 3); 4th best performance (score 2); Worst performance (score 1)

Figure 5-26: Comparison of different techniques: risks and performance score. The combination of the scores from Part 1 and Part 2 achieves the best performance.

similar analysis has been performed for the underestimation cases.

Figure 5-27 shows the total performance score for the four techniques, averaged across the nine test cases (overestimation and underestimation together). As previously observed, the technique that achieves on average the best performance is the one that uses a combination of the scores computed in Part 1 and in Part 2. This technique performs better than the ones that use the scores of Part 1 only or Part 2 only, or than the one in which experts are equally weighted. The next section describes the Bayesian integration of data statistic and expert statistic.

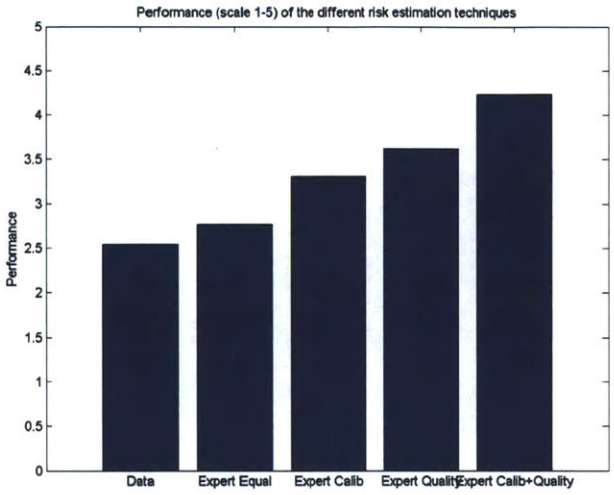

Figure 5-27: Comparison of different techniques using the performance score previously defined. Expert statistic (with experts' probability densities composed using scores from Part 1 and Part 2) performs better than data statistic and than any other expert aggregation techniques.

5.4 Bayesian Integration Between Expert Opinion and Historical Data

The third possible way to assess risks for mass and power fluctuations is to combine the sources of information given by data and by expert. In this thesis, a Bayesian approach to integrate the two statistical sources is proposed.

In the Bayesian framework data statistics are considered the prior distributions; and the distribution that results from the expert elicitation process is considered the likelihood function. The resulting a posteriori distribution represents the combined estimate. The mathematical expression is as follows:

$$f_{Bayesian}(x \mid \vartheta) = \frac{f_{Data}(x) \cdot f_{Expert}(\vartheta \mid x)}{\int f_{Data}(x) \cdot f_{Expert}(\vartheta \mid x)\, dx} \tag{5.14}$$

An example for CASTOR antenna is shown in Figure 5-28 and in Figure 5-29.

It is possible to notice that the Bayesian combination reflects contributions from both distributions: the peak tends to reflect the data statistic, while the tail tends

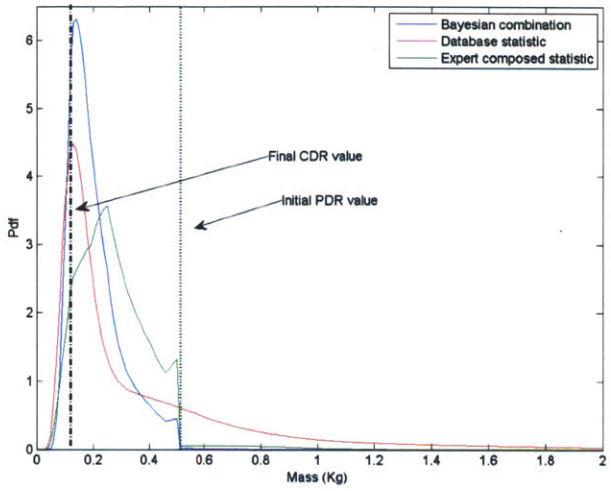

Figure 5-28: Probability density function for CASTOR antenna. Bayesian estimation is compared to the database estimation and the expert estimation.

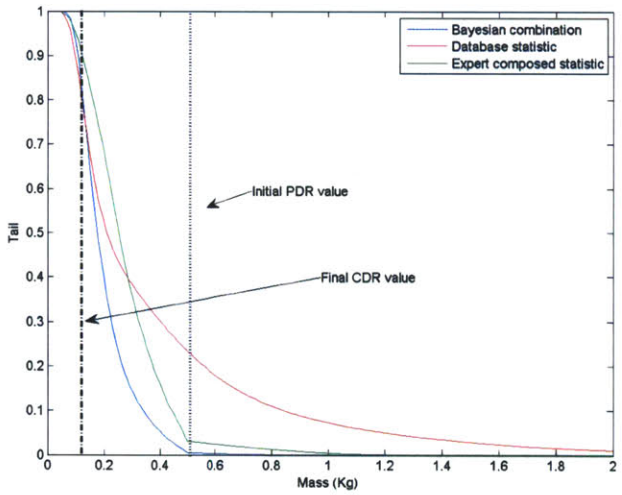

Figure 5-29: Tail function for CASTOR antenna. Bayesian estimation shows a less heavy tail.

to be more skewed similarly to the one computed through expert elicitation. The Bayesian approach generally helps to avoid the problem of the heavy tail functions which has been described for the database in Chapter 4. This feature makes the Bayesian combination approach very useful when the initial estimations of mass and power are overestimated. More discussion on this aspect is included in Chapter 6, as

the test cases are discussed.

5.5 Conclusions

This chapter proposes an approach to perform expert elicitation that differs from the classical model. One of the main motivations to develop a customized model was the unavailability of empirical data to model probability densities for the seed variables. The proposed approach is organized in four parts:

- Probabilistic Thinking: It maps experts' sensitivity to biases and heuristics, generating a quantitative score.
- Calibration: It assesses the tendency of experts to over- or underestimate quantities, using point estimation questions. It generates a calibration score and a calibration coefficient.
- Elicitation: It models the probability densities of interest using the quantile method or statistical link analysis method. The calibration coefficient is applied to shift the distributions appropriately. A graphical tool allows experts to visualize different densities and to select the one that represents their belief.
- Expert Aggregation: It composes experts' densities according to four possible weighting schemes, which use scores computed in Probabilistic Thinking and in Calibration.

This methodology was implemented through an interview performed with students (Part 1 only), and engineers (total interview). The results are summarized as follows:

- Probabilistic Thinking: The interview works uniformly across the three populations (average quality score measured is 60-61). This result provides a baseline to assess the quality of Probabilistic Thinking for experts. In addition to the quality score, different biases were mapped. Students and engineers showed a good understanding of coherence, while performance for other biases varies from case to case. Quality scores were analyzed with respect to experience/training

in probability revealing lack of correlation, which is also a consequence of the fact that the interview is designed to detect biases which may not necessarily be correlated with the training in probability.

- Calibration: Experts tended to overestimate the quantities of interest. The result is not surprising given the tendency of engineers to produce conservative estimates for the design metrics (mass and power consumption). Calibration scores were compared with years of experience in communication system design, revealing a limited correlation.

- Elicitation: Experts generally selected asymmetric triangular probability distributions. They tended to prefer the quantile elicitation method with respect to the statistical link analysis elicitation method. Finally, experts were generally inclined in selecting the non-calibrated distributions and not the calibrated distributions. This result indicates that experts do not always agree with the shift of the distributions performed by the calibration coefficient.

- Expert Aggregation: The composition of densities based on the computed scores improves the estimation with respect to the case of densities equally weighted. To show the improvement, test cases of missions already developed were used. For these missions, initial and final estimations of the metrics of interest (mass and power consumption) were available. Hence, test cases were applied to quantify the ability of the expert to recognize whether the initial estimations were over- or underestimated. A performance metric to compare the performance of the different aggregation strategies was introduced, which showed that the computed scores improve the estimation with respect to the case of experts opinions equally weighted. The composition of densities also improves the estimate with respect to data statistics in the great majority of the cases.

As a result of the methodology and of the experiment presented in this chapter, expert elicitation seems to improve the assessment of risk in communication system design with respect to using data statistics only.

However, to exploit both sources of information (expert and data) a Bayesian approach which integrates expert elicitation and historical data, is proposed. The results indicate that this approach improves the risk estimation, especially in the overestimation cases. This method seems to be a good combination of the two information sources.

Chapter 6

Risk Model

This chapter presents the development of the risk model. This model aims to quantify the risk of exceeding mission constraints for different communication architectures. To properly compute risks, the model needs a baseline, which represents an initial estimate of the design. In some cases, such as the test cases here discussed (CASTOR and HETE-2), the baseline design at PDR is already known. However, if this initial estimation is not available (as described in Chapter 7), a model to generate an initial estimate is required. In this thesis, a baseline model is developed to accomplish this purpose: it computes one or more possible baseline architectures which can accomplish a specific mission. The output of the baseline model includes sets of possible antennas, transceivers, amplifiers and switches which can provide the adequate EIRP (Equivalent Isotropic Radiated Power) to transmit the amount of data required. The baseline model has been validated using previously developed NASA missions. For any proposed system architecture, the baseline model calculates the corresponding average mass and power consumption estimations. These outputs then become input for the risk model.

The risk model computes the risk of exceeding constraints in mass and power for any component of the communication system. It integrates the data statistics and expert elicitation statistics, as described in Chapters 4 and 5, to calculate tail functions for any component. The risk model is corroborated through the use of test cases. Test case verification helps to infer guidelines and eventual limitations in the process of

quantifying design risks. In this thesis, two test cases have been considered: CASTOR and HETE-2. An additional test case is the CASSINI-HUYGENS mission. Unfortunately, the unavailability of experts' data for this mission limited the analysis. For this reason, the CASSINI-HUYGENS mission is not described in this chapter, but it is briefly reported in Appendix D.

The chapter is structured as follows: the baseline model is presented, the risk model is described, and test cases are analyzed. Finally, guidelines and limitations of the risk model are discussed.

6.1 Baseline Design

The baseline design provides an initial estimation of the PDR design for the metrics of interest (mass and power in this case). It generates a point estimation which constitutes the input of the risk model. In the specific case of communication systems, the baseline design uses state of the art equations to model coverage, to perform link analysis, and to generate a preliminary estimate of mass and power for each of the components. Each submodule of the baseline design is an analytical model based on current knowledge in communication systems.

Baseline design is a parametric model; hence, different space communication systems can be modelled through an opportune combination of communication input parameters. The baseline design models the communication system as a set of k channels, where each channel is an independent communication system with its own central frequency (VHF, S-Band, X-Band, and Ka-Band). For any channel, the communication system can be decomposed in multiple antennas and transceivers according to the architecture selected and the level of redundancy.

The output of the baseline design is the design of one or multiple communication architectures that can accomplish the mission objectives. The details of the design include specifications on design values for mass and power consumption for each component. The baseline design is composed of three sub-modules (Figure 6-1):

1. Coverage sub-module: It receives as inputs the six orbital parameters (eccen-

tricity, semi-major axis, inclination, longitude of the ascending node, argument of periapsis, and mean anomaly), the simulation time, and the location of the receiving ground stations. The output is represented by a coverage map, which indicates whether the satellite is in coverage with one or more receiving stations for any instant of time (the time scale of the model is in seconds). The coverage model is developed using STK-Matlab software.

2. Link analysis sub-module: It receives as inputs the coverage, the ground station receiver characteristics such as gain and noise, and the amount of data to be transmitted. Link-analysis uses coverage data to establish the minimum transmission rate required to download data. The required EIRP for each of the channels is computed, and it becomes an input for the next sub-module.

3. Mass and power consumption estimation sub-module: For a given EIRP and a given bandwidth, different communication architectures can be considered. If the problem is constrained by mass, a designer will generally opt for a small gain antenna with an amplifier, which drives the power consumption. On the other hand, if the problem is constrained by power consumption, a designer will select a high gain antenna without an amplifier. This solution will increase the mass of the system while limiting power consumption. The module is developed in Matlab, and it uses fitting relations available in the literature [24] to calculate the design values of mass and power for components. This output becomes the input for the risk model.

The following subsections detail inputs, outputs, and validation for the baseline design.

6.1.1 Parametric Baseline Model: Inputs and Outputs

The baseline design is parametric, and allows the designer to characterize different communication systems by defining the input parameters. The input vector of the

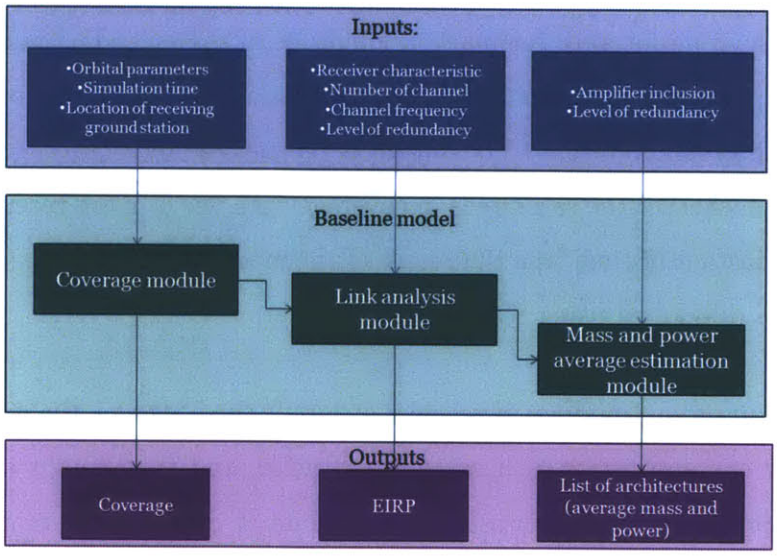

Figure 6-1: Summary of baseline model. Inputs and outputs are listed for the three submodules (Coverage, Link Analysis, Mass and Power Estimation).

model, defined as $\bar{X}$, is summarized in Equation 6.1:

$$\bar{X} = \begin{bmatrix} k & f_1 & f_2 & \dots & f_k & r_1 & r_2 & \dots & r_k & Lat & Lon & G_{rx} & \frac{G_{rx}}{T_k} & i \\ \Omega & \omega & e & d & M_0 & Eb/N0 & D_1 & D_2 & \dots & D_k & T & A & S_M & \end{bmatrix} \quad (6.1)$$

The input parameters (Equation 6.1) are described as follows:

- **Number of communication channels (k):** This input identifies the different communication systems (in different bands) that the system provides. Each channel is defined in the thesis as an independent communication system composed of at minimum a transceiver and an antenna. In some cases, tuning transceivers are adaptable for multiple frequencies; hence separate transceivers for different frequencies are not always required. However, in some cases, separate transceivers are used for redundancy purposes. For simplicity, in this thesis, separate transceivers for different channels are assumed.

- **Central frequency for each channel (f_i):** For each of the communication channels, the corresponding bandwidth (VHF, S,L Ka) is selected.

- **Level of redundancy (r_i):** For each channel, it is possible to select between

four levels of redundancy: level 0 (no redundancy), level 1 (redundancy in the antennas, not in the transceivers or in the amplifiers), level 2 (fully redundant system: redundancy in the antennas and in the transceivers), level 3 (redundancy in the transceivers and not in the antennas). This element of the baseline design is used to model the so-called inclusion constraints, i.e. the necessity of including certain components whenever a specific architecture is selected. For example, when redundancy in the antennas is selected, a switch needs to be added to the system. There are many considerations for which a designer should select any of those possible redundancy alternatives. The non-redundant system is generally selected for very constrained mass and power systems, or for multi-channel systems in which the redundancy is guaranteed by the existence of multiple channels. The redundancy in the antennas is a solution selected to increase coverage by placing antennas in multiple locations: it generally requires an active switch or a passive splitter to connect the antennas to the transceiver. Redundancy in the transceivers is mostly used to counteract transceivers' failures. The fully redundant system is the solution that counteracts the greatest number of failures, since the system is double stream. However, this option severely impacts mass and power consumption.

- **Receiver characteristics** (Lat,Lon,G_{rx},$\frac{G_{rx}}{T_k}$): location (latitude and longitude), gain, and gain over noise temperature ratio for Earth stations. The location is used by the coverage submodule, while the other data are used by the link analysis submodule to compute the required EIRP.

- **Orbital parameters:** inclination (i), longitude of the ascending node (Ω), argument of periapsis (ω), eccentricity (e), semi major axis (d), and mean anomaly (M_0). They are applied in the coverage module to simulate the spacecraft's orbit.

- **Quality requirement** ($\frac{Eb}{N0}$)**:** minimum Signal to Noise Ratio (SNR) or Energy per bit over spectral noise temperature density ($\frac{Eb}{N0}$) required to obtain a certain quality of the received signal on the ground.

- **Mission data requirement (D_i):** data to be transmitted per channel during mission lifetime.

- **Simulation time (T):** time selected to run the simulation. If this parameter is not modified by the user, the default value is 2 Earth days.

- **Amplifier inclusion (A):** a binary value that models whether or not the communication system will be equipped with amplifiers. For small satellites in LEO (Low Earth Orbit), an amplifier is generally not required since the mission can be accomplished with low power transceivers and low gain antennas. However, in deep space exploration missions or for commercial satellites, the EIRP required generally demands an amplifier.

- **System margin (S_M):** desired margin over the quality of the received signal for each channel. A good rule of thumb is that a minimum of 3 dB margin should be taken into account in the design of the communication system.

The outputs of the baseline design are the following:

- **EIRP required for each of the k channels:** the amount of energy required for each communication channel of the communication system to transmit data.

- **List of components with respective mass and power design values for any architecture:** These values are initial estimations, and they do not include risk calculation, which is handled in the risk model.

6.1.2 Model Validation

Validation of the baseline design is performed with data from past and current space missions. Specifically, the link analysis submodule is validated by comparing the EIRP computed by the model with the one obtained from technical documentation for different missions ([125], [76], [123], [73], [120], [122], [86], [124], [119], [110], [121]). The results are in Figure 6-2. The values of EIRP computed by the model are close to

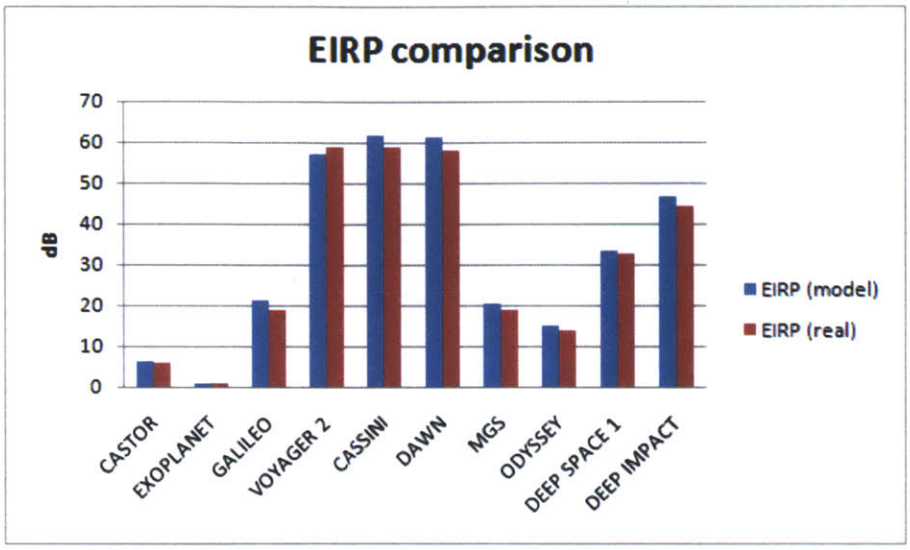

Figure 6-2: Link analysis validation. The EIRP computed by the model is very close to the real values of EIRP listed in the technical documentation.

the real ones: variations are due to slight differences in the noise temperature calculations, or in the estimation of pointing errors. A similar validation has been performed by comparing values of mass and power consumption computed by the model with the real values obtained from technical documentation. Specifically, for small satellites, data from two university missions currently developed at MIT Space System Laboratory have been used (CASTOR [1], and ExoplanetSat [2]). The first satellite is an ESPA ring class satellite designed to test a new type of electric propulsion system (more details in the next sections), while the second is a CubeSat conceived to detect Earth-like planets. The real values of mass and power consumption for both systems are compared with the model values in Figures 6-3 and 6-4. The results are very close (within 8%).

A similar analysis has been developed using data from missions developed at NASA JPL ([125], [76], [122], [124], [119]). Figures 6-5 and 6-6 show that the real values of mass and power consumption are close to the ones computed using the baseline design. Model validation shows that the baseline design can provide reliable estimates of the required EIRP, and of the design values of mass and power con-

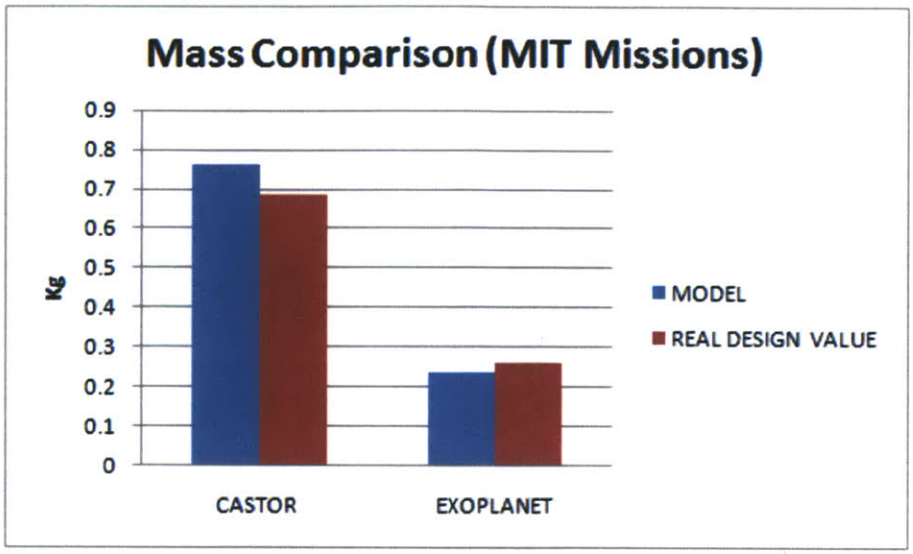

Figure 6-3: Mass comparison for CASTOR and ExoplanetSat. The baseline design values (model) correlate well to the real values.

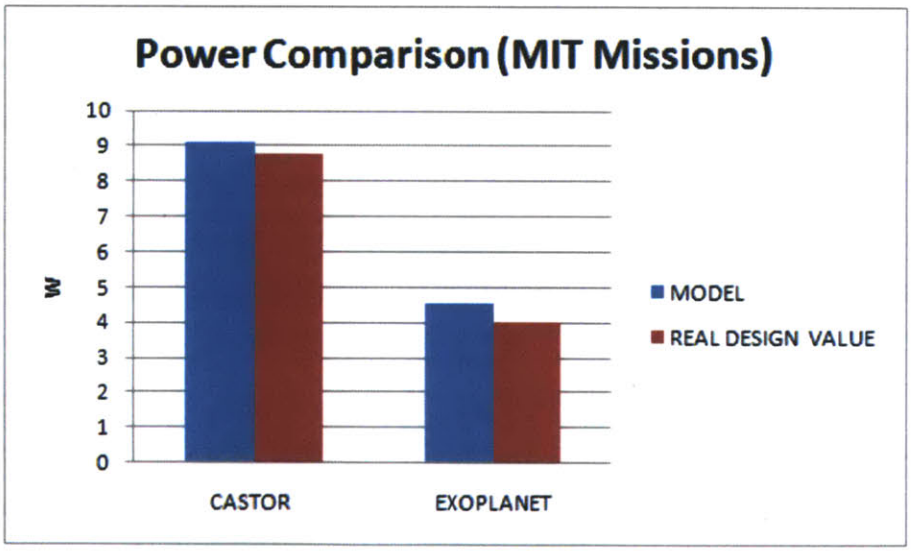

Figure 6-4: Power consumption comparison for CASTOR and ExoplanetSat. The baseline design values (model) correlate well to the real values.

sumption for specific communication system architectures. However, as mentioned in Chapter 3, the correctness of baseline design is not critical for the performance of the entire statistical risk estimation model. In fact, the goal of the risk model is to identify whether or not the initial design is likely to exceed expected values of mass and power consumption. If the estimation created by the baseline design is an

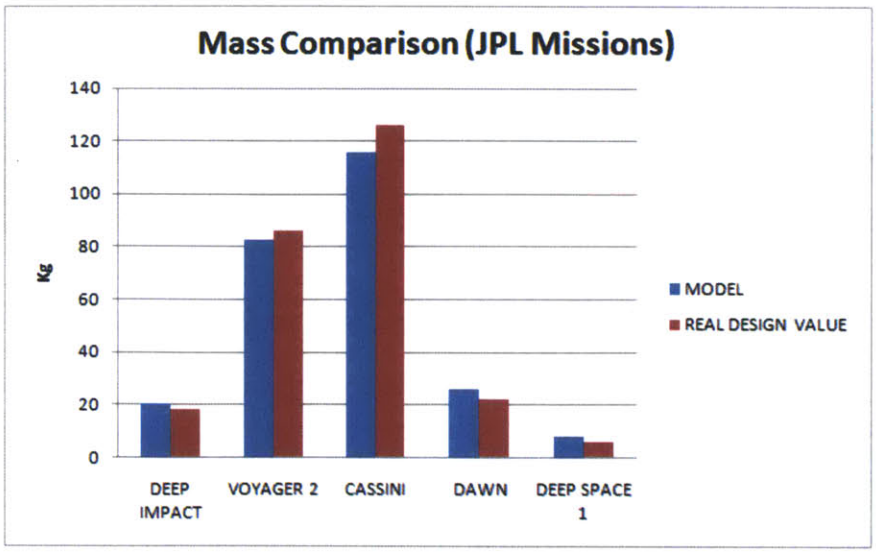

Figure 6-5: Mass comparison for NASA JPL missions. The baseline design values (model) correlate well to the real values.

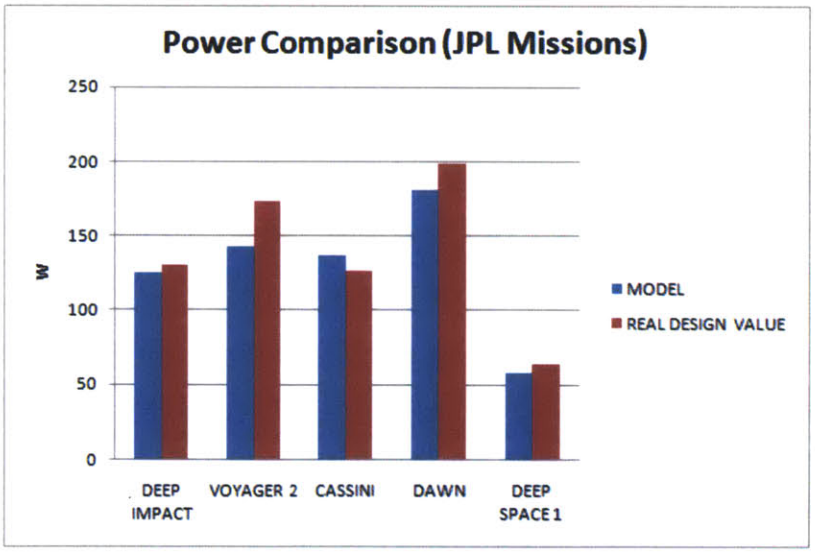

Figure 6-6: Power consumption comparison for NASA JPL missions. The baseline design values (model) correlate well to the real values.

overestimation or an underestimation, the risk model will reveal it.

6.2 Risk Model

The risk analysis model receives as input a given design, and estimates the probability that the overall mass or power would exceed the given allocations. This probability is computed using the techniques previously discussed: data statistics, expert elicitation statistics, and Bayesian combined statistics.

Regarding data statistics, density estimation is used to compute the probability distributions. The size of the database is limited for each category of components, and the samples tend to take on a wide range of values. The distribution constructed from these samples is inclined to have a larger dispersion. This problem is detailed in the application examples discussed in the next section, and it is one motivation to encourage the use of expert elicitation statistics.

Regarding expert elicitation statistics, the accuracy of experts' judgments is difficult to measure, but the three-part interview is used to quantify the expert's precision. An advantage of expert elicitation is that when an expert is confident, he or she will impose a smaller uncertainty range with respective to his or her assessment, and the elicitation helps to generate narrower probability distribution functions compared to the ones constructed using data statistics (more details in Section 6.3).

Data statistics are independent from human judgment, but they face limitations due to the size of the database. On the other hand, expert elicitation can reduce uncertainty if the expert is very confident, but the validity is dependent on human judgment. Combining the two statistics is a key aspect to improve risk estimation. Hence, the Bayesian combined statistics are used, and it appears to be the most satisfactory technique to model risks (see Section 6.3).

The three statistics are simultaneously applied to the test cases (Section 6.3), because all of them can provide complementary information to model design risks. Specifically:

- Data statistics are generally used to identify whether the initial estimated design values of mass and power consumption are an over- or an underestimation.
- Expert statistics perform better than data statistics in modelling the tail of the

distribution, which is the key metric in evaluating design risk.

- Bayesian statistics are a hybrid of the two previous statistics, and they capture information from different sources.

The next subsection details the generation of probability density functions and tail functions.

6.2.1 Probability Density Functions and Tail Functions

The baseline design (or the PDR data, if available) provides an initial estimation of design values for mass and power consumption of different components. Each component (x_i) is described by the following parameters:

$$x_i = \begin{bmatrix} m_i & p_i & g_i \end{bmatrix} \tag{6.2}$$

where m_i is the average mass, p_i is the average power consumption, and g_i is the performance metric of the component. The performance metric measures the ability of the component to perform the desired function. It is very dependent on the problem analyzed. In the specific case of statistical risk analysis for communication system design, this metric indicates the impact of the component on the required EIRP on the satellite, which is the gain in dB for the component. m_i, p_i, and g_i are used to compute the probability functions which describe the fluctuations of mass and power consumption. For passive components (like antennas), only probability density functions for mass are computed, while for active components the probability densities for both mass and power consumption are calculated. The procedures for computation vary for the different statistics and they are listed as follows:

- Data statistics: The performance metric g_i is used to scale data samples in the database. Recalling Chapter 4, the probability function for the mass of component x_i, with gain g_i results:

$$f_{Data}(x_i, x) = \frac{1}{g_i}\left(\frac{1}{n \cdot h}\sum_{j=1}^{n} K\left(\frac{x - a_j \cdot g_i}{h}\right)\right) \tag{6.3}$$

where a is a vector of n elements that collects all the samples of mass per unit of gain for that specific type and category of component. $K(\cdot)$ is a normal kernel used for the Kernel Density Estimator, and x is the support of the distribution. In the case of power consumption the expression is the same, but the vector a is different.

- Expert elicitation statistics: The performance metric g_i, the average mass m_i, and power consumption p_i are given as input to the expert. The experts model the statistics according to the approach defined in Chapter 5. Each expert observes the statistics elicited and selects the distribution that most properly expresses his/her beliefs. The distributions from each expert are aggregated as described in Chapter 5 using the score computed with Part 1 and Part 2 of the interview. Hence, expert elicitation statistics are formulated as follows:

$$f_{Expert}(x) = \sum_{i=1}^{n} w_i \cdot f_{exp_i}(x) \tag{6.4}$$

where w_i are the weights used to compose multiple opinions from the different experts (f_{exp_i}), and x is the support of the distribution.

- Bayesian combination of data and expert statistics: The two distributions are composed using the Bayesian rule. The data statistic represents the a prior distribution and the expert statistic is the likelihood function. The process is described in Chapter 5.

In the case of multiple components, the total probability densities for mass and power consumption are computed as the convolution of the single densities under the hypothesis of independence:

$$f_{tot}(x) = f_1(x) * f_2(x) * ... * f_n(x) \tag{6.5}$$

where $f_i(x)$ indicates the probability density function for mass fluctuations or power fluctuations of the *ith* component. The convolution provides the exact calculation for

the sum of independent random variables. However, if many components are considered, the convolution can become computationally expensive. Hence, when more than two components are considered (Chapter 7) a computation based on Saddle-Point approximation [38] is applied. When the distribution is computed, the tail function is calculated through integration.

6.2.2 Model Validation

Validation of risk model requires the availability of real probability densities of mass and power fluctuations for different components to compare with the ones generated by the model. Unfortunately, such probability densities do not exist. Hence, the risk model cannot be fully validated.

As an alternative, rather than validate the model, test cases are used to compare the results of the model with the real data (at PDR and at CDR). The comparison aims to show how the model proposed in this thesis can improve the state of the art in risk assessment and in sizing of the contingencies.

6.3 Test Cases

Two test cases have been selected to validate the approach: CASTOR and HETE-2. The following subsections provide details on the missions, the communication systems, and the risk analysis performed by applying the risk model.

6.3.1 CASTOR

The following subsections describe: the CASTOR mission, its communication system, risk analysis, and conclusions.

CASTOR mission

CASTOR [1] is a small ESPA ring class satellite developed at the MIT Space System Laboratory. The satellite dimensions are 50x50x60 cm with a total mass of 50 Kg. The

main goal of this spacecraft is to test in the space environment the performance of a new type of electric propulsion engine, the DCFT (Diverging Cusped-Field Thruster), which guarantees up to 1 km per second of ΔV. This type of engine is very efficient in terms of mass/impulse ratio and the whole system is very promising to perform rapid orbital transfer maneuvers. The deployed configuration of the CASTOR satellite is shown in Figure 6-7. CASTOR has been developed entirely at MIT Space System

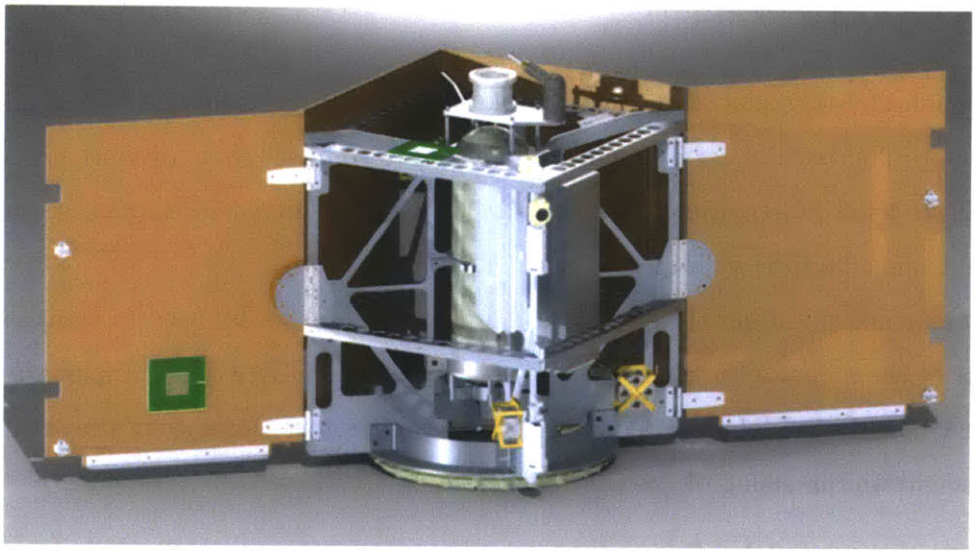

Figure 6-7: Configuration of CASTOR spacecraft with solar panels deployed. Source: [1].

Laboratory over 4 years. Hence, the test case is suitable for risk analysis, since all the data of the design are available ([1], [10]). Specifically, data of interest are:

- EIRP required: This input is important to validate the baseline model as shown in Section 6.1.

- Configuration of the communication system: type, and number of antennas used, type and number of transceivers used, and additional components.

- Initial (PDR) values of mass, power, and gains (m_i, p_i, g_i) for any component of the communication system: these values are used as substitutes for the baseline design to obtain the starting point of the risk analysis.

- Final (CDR) values of mass and power for any component of the communication system: these values are used to test the statistical estimation methodology developed.

Communication System

The CASTOR communication system is a fully redundant system composed of three antennas and two modems. The three antennas are custom built patch antennas with approximately 6 dB of gain. The modems are Microhard MHX2420, a Commercial-Off-The-Shelf (COTS) product. One antenna is connected directly to one of the modems, while the other two antennas are connected to a second modem through a passive splitter.

The operating frequency of the system is S-Band (2.442 GHz), and the ground stations are the same used for HETE-2 missions. HETE-2 mission transmitted data to a network of ground stations owned by MIT Kavli institutes located in: Singapore, Kwajalein, and Cayenne.

Risk Analysis

The risk analysis performed looks at the fluctuations of values for mass and power of transceivers and antennas. The switch has not been analyzed due to the lack of historical data and expert information on the component.

The three antennas are identical; hence we discuss the mass fluctuations for only one of them. The antenna design has passed through different phases. Initially the antennas were supposed to be COTS patches, but then due to some vibration concerns the team resorted to custom-made antennas. The material for the antennas changed a number of times before the final configuration was selected at CDR. The initial mass of the antenna at PDR was estimated to be approximately 0.5 Kg, while the final value became 0.1 Kg, which shows that the initial estimate was overestimated.

The risk analysis performed applies three statistics: data statistic, expert elicitation statistic, and Bayesian composed statistic. For data statistic, the following steps are

followed:

- Component's type and category: In this case it is a low gain antenna. This information is used to select the correct sets of data in the database. As already mentioned in Chapter 4, data samples used to generate statistics are grouped according to the type of components (antenna, transceivers, amplifiers, etc), and to the component categories (for antennas: LGA, MGA, and HGA).
- Frequency: S-Band.
- Performance metric: in this case the gain of the antenna, used to scale the sets of data.
- Estimation of the final probability density and tail function.

For the expert statistics, the experts were asked to model the probability distribution that the mass of the antenna would exceed the value of 0.5 Kg. Different experts have been tested (see Section 5.3.4). The great majority of them understood immediately that the initial value of the antenna mass was an overestimation. The experts' assessments for the probability distribution were aggregated according to the process described in Chapter 5. The Bayesian combination of expert and data is obtained by applying the process described in Chapter 5.

The results of the risk estimation performed using database approach, expert approach, and combined approach are shown in Figure 6-8 (probability densities), and 6-9 (tail functions). Specifically in Figure 6-8, the peaks of the three densities are closer to the final CDR value than to the PDR value, which shows that the methods are effective in estimating the eventual value, with the data curve predicting better than the expert curve and the composed curve. However, in the data statistic approach the density function exhibits a stretched-out shape with a heavy tail as shown in Figure 6-9, indicating an overestimation of the mass margin required to counteract a given design risk. For example, if the risk requirement of not exceeding the mass allocation is 0.1, the mass margin that would be required using the data statistic approach (0.4 Kg) is bigger than the one that would be allocated by using the expert

approach or the Bayesian composed approach (0.2 Kg for expert approach, 0.15 Kg for Bayesian approach). This effect is due to the limited number of samples of the database as described in Chapter 4.

In summary, data statistic, expert elicitation statistic, and Bayesian composed statistic generate density functions whose peaks recognize that the initial design value of 0.5 Kg is overestimated, and that the expected value for the antenna mass should be less. The peak of the data statistic is closer to the CDR value than expert statistic. However, when it comes to risk quantification (tail graph) the expert statistic offers a much smaller mass margin than data statistic for the same level of risk. The Bayesian combination seems to collect the positive features of both approaches: the peak of the density is closer to the final CDR value, and the tail function shows a smaller mass margin with respect to the other techniques.

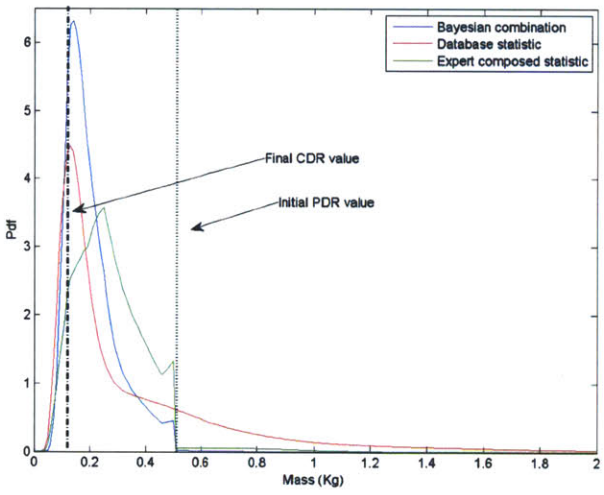

Figure 6-8: Probability density function for CASTOR antenna. The peaks of the distributions are closer to the CDR value than to the PDR value, showing that the initial PDR estimate of the antenna mass was an overestimation.

An analogous analysis has been performed with the CASTOR transceiver, and the results lead to very similar conclusions. The component is a COTS product with an output transmitting power of 1 W. For this component, the initial values for mass and power consumption at the PDR stage were 0.11 Kg and 1.5 W respectively. The final values after CDR were 0.05 Kg and 4.5 W. As observed, the initial estimate was an

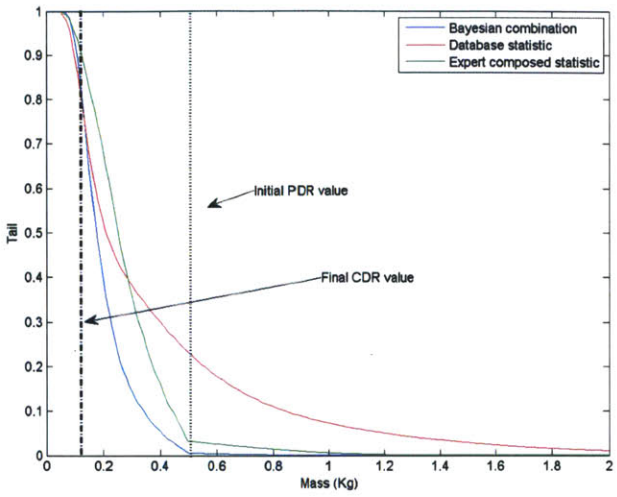

Figure 6-9: Tail function for CASTOR antenna. Data statistic shows a heavier tail with respect to expert statistic and to Bayesian statistic.

overestimation for the mass, and an underestimation for the power. The underlying reason for the fluctuations in the values is mainly due to the misunderstandings with the company. The component was already developed and built, but the data sent from the company were incorrect and the final values were assessed only as the component arrived at the laboratory.

Similar to the case of the antenna, the risk analysis for the transceiver was performed using the three statistics previously described. Regarding data statistic, the steps are very similar to the ones discussed for the antenna. Regarding expert statistic, the experts were asked to model the probability distribution that the mass of the transceiver would exceed the value of 0.1 Kg, and that the power consumption would exceed 1.5 W. Also the Bayesian composition was computed. The probability density functions and the tail functions for the transceiver mass are shown in Figure 6-10 and Figure 6-11, and for the transceiver power consumptions are given in Figure 6-12 and Figure 6-13.

Regarding the transceiver's mass, the peaks of the expert probability density and of the Bayesian combined density are higher than the one computed using data statistic. The density computed using data statistic tends to be more spread, and it does not

reach a high peak. However, the peak of the density computed using data statistic is closer to the final CDR value, while the expert opinion distribution and the Bayesian distribution overestimates the transceiver mass. As in the antenna case, the expert opinion generates a narrower distribution than the data, and the overall effect is that it requires a smaller transceiver mass margin, 0.4 Kg versus 0.7 Kg, to mitigate the risk of exceeding the mass allocation at 0.1 risk level. In the case of the power consumption (Figure 6-12), the peaks of the probability densities are close to the final CDR value, showing that the initial estimate on the power consumption was an overestimation. The tail functions (Figure 6-13) show once again that the expert statistic is helpful to reduce the power margin compared to the one computed using data statistic. The Bayesian combined statistic is also helpful in reducing the tail of the distribution.

The next subsection summarizes conclusion on the CASTOR test case.

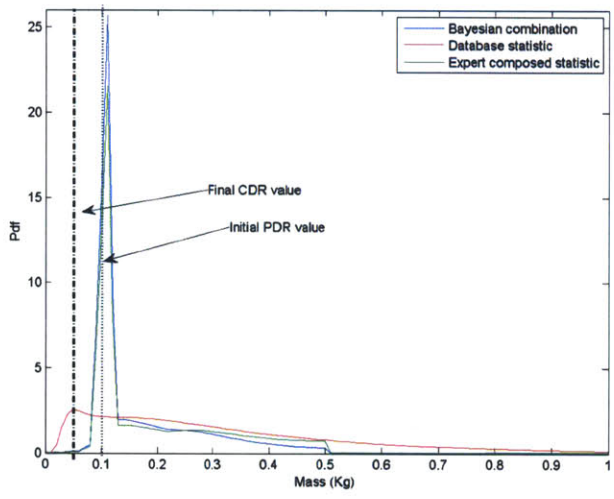

Figure 6-10: Probability density function for CASTOR transceiver mass fluctuation. The data statistic tends to be more spread and it does not reach a high peak value.

Conclusions

The risk model is applied to a university mission CASTOR, and the risk assessment is focused on the antenna and on the transceiver. The results reveal advantages and

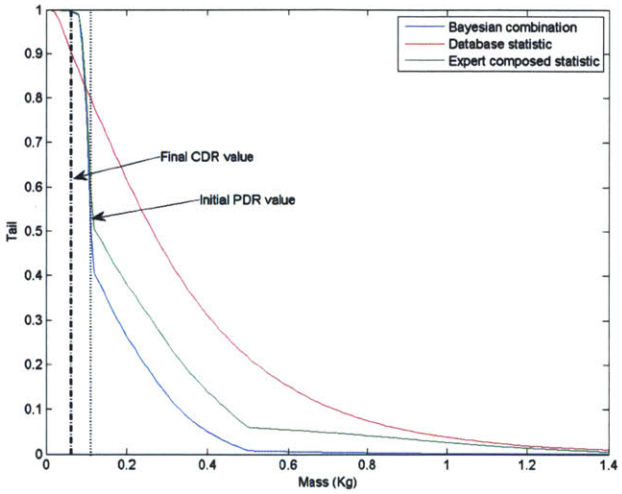

Figure 6-11: Tail function for CASTOR transceiver mass fluctuation. The risk quantification performed with expert opinion and Bayesian statistic allows the designer to reduce the mass contingency with respect to the one performed using data statistic.

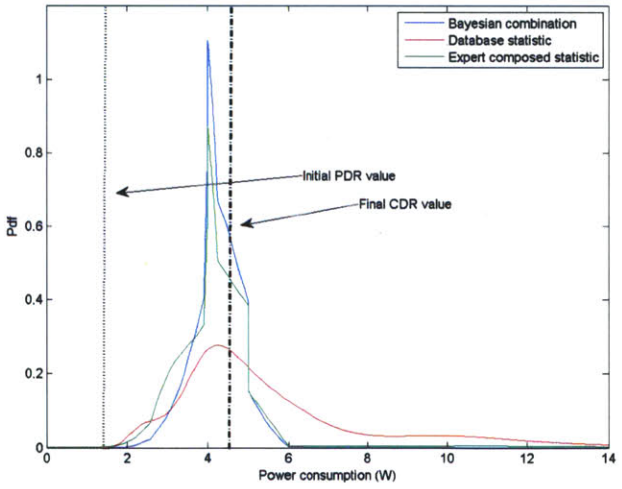

Figure 6-12: Probability density function for CASTOR transceiver power fluctuation. The peaks of the three distributions are close to the CDR value, showing that each of the statistics is able to identify that the PDR value is an underestimation.

disadvantages in the different statistics used.

Data statistic provides valuable information regarding the peak of the probability density function. The peak of the distribution is in the three cases close to the final

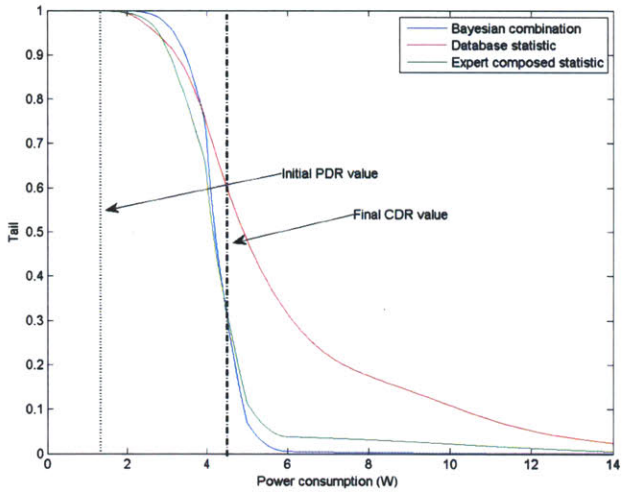

Figure 6-13: Tail function for CASTOR transceiver power fluctuation. The expert approach is helpful to reduce the power margin.

CDR values. Hence, data statistic is useful to compute an estimate of the final value of mass and power of a component. However, the tail functions computed through data statistics present heavy tails due to the limited number of samples available. The heavy tails generate very cautious estimates of risk. If data statistics are used to estimate contingency (the amount of extra mass or power that should be taken into account to maintain risk below a certain threshold) then the result will be clearly an oversizing of such contingency. As a consequence, an engineer who uses only data statistic to perform the estimation will be able to assess properly the final value of mass and power consumption of the component by looking at the peaks of the probability density functions, but the risk estimation for the component will be very high and the contingency oversized.

On the other hand, probably due to the conservative mindset of most engineers, expert opinion tends sometimes to overestimate the mass and power consumption as it is observed in the peaks of the distributions for the above CASTOR antenna and transceiver experiments. However, expert's confidence helps to limit the variance (or spread) of the distribution, thus reducing the required margin to counteract the design risk. Hence, an engineer who uses only expert statistic to perform the estimation will

be able to assess properly the risk estimation and the value of the contingency, while the assessment of the final value of mass and power consumption for the component will generally suffer from some overestimation.

Since both techniques have advantages and weaknesses, the Bayesian approach seems to combine both sources of information in a mathematically tractable manner, being a good compromise across the solutions. However, in some cases, as a result of the product rule, the peaks of the Bayesian distributions tend to be excessively close to the expert statistic (especially when database peaks are negligible compared to expert ones, as it is observed for CASTOR transceiver mass), losing some of the information contained in the data statistic.

As shown in Section 6.3.3, the CASTOR transceiver mass is the only test case in which statistical risk estimation provides a sizing of the contingency which is less satisfactory than the traditional approach based on contingencies, due mainly to a misunderstanding with the experts performing elicitation. In fact, it seems that the experts over sized the mass of the transceiver. When interrogated on this result, some of them seem to have included in their estimation the mass of a box to contain the transceiver, while the interview asked specifically for the transceiver only.

The next section discusses the results of the HETE-2 test case.

6.3.2 HETE-2

The next subsections describe: the HETE-2 mission, the HETE-2 communication system, risk analysis for HETE-2, and conclusions.

HETE-2 mission

The HETE (High Energy Transient Experiment) [131] mission was an international mission led by the Massachusetts Institute of Technology (MIT). Its prime objective was to carry out the first multi wavelength study of gamma-ray bursts (GRB) with UV, X-ray, and gamma ray instruments. A unique feature of the mission was the capability to localize bursts with several arc-seconds accuracy, in near real-time aboard

the spacecraft. The HETE spacecraft was sun-pointing with four solar panels connected to the bottom of the spacecraft bus. Spacecraft attitude was to be controlled by magnetic torque coils and a momentum wheel.

The HETE satellite was launched with the Argentine satellite SAC B. Unfortunately, HETE was trapped within the Dual Payload Attachment Fitting due to a battery failure in the Pegasus-XL rocket third stage. Due to its inability to deploy the solar panels, HETE lost power several days after launch.

The HETE-2 (High Energy Transient Explorer 2) [131] mission was designed to detect cosmic gamma-ray bursts (GRBs) to determine their origin and nature. The satellite had three instruments: a set of wide-field gamma-ray (6 - 400 keV) spectrometers (FREGATE), a wide-field X-ray (2 - 25 keV) monitor (WXM), and a set of soft X-ray (0.5 - 10 keV) cameras (SXC). These instruments covered a solid angle of 1.5-2 steradians and they were used for simultaneous, broad-band observations in the various listed energy ranges. The goal of the mission was to continuously scan the sky, to identify occurrences of GRBs, to establish precise locations, and to transmit coordinates in near real time (< 10 seconds).

The spacecraft is a rectangular cube (100x50x50 cm) with four solar panel petals protruding from the bottom (Figure 6-14). The bottom section of the spacecraft holds the power, communications, and attitude control and the upper section the science instruments.

Communication System

The communication system [131] was via S-Band in uplink (2.092 GHz) and downlink (2.272 GHz). The system was composed of a transceiver and 5 patch antennas displayed in different locations on the spacecraft for redundancy and coverage. A GPS channel was also available, and it was composed of an antenna and a GPS receiver. Finally, a VHF downlink (137.9622 MHz) was used for real-time burst alerts.

The ground station used was the MIT Kavli HETE network, the same as CASTOR. HETE-2 was successfully launched in 2000 and the mission was active until 2007.

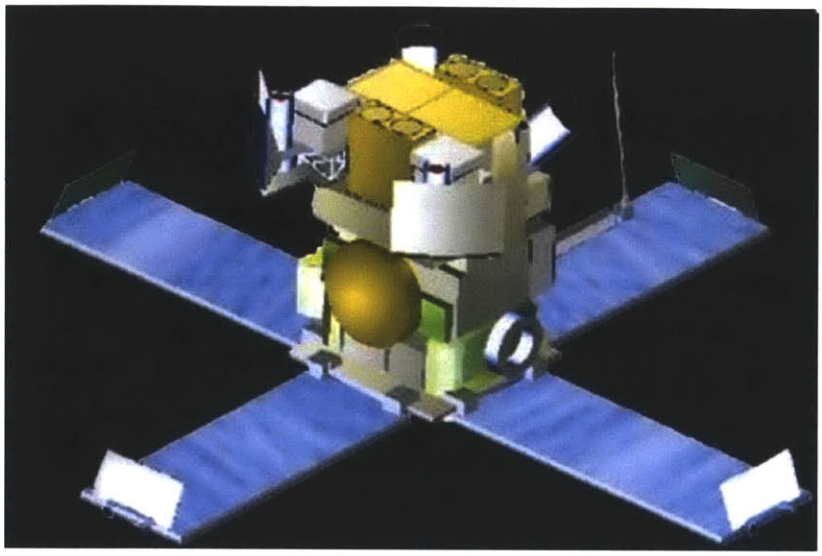

Figure 6-14: HETE-2 spacecraft overview. Source: [131].

Risk Analysis

The risk analysis performed for this research is focused on HETE-2, and particularly on the communication system for the spacecraft. Specifically, the analysis is centered on the S-Band channel and on the GPS channel, for which PDR and CDR data are available.

The S-Band system is composed of a transceiver and 5 patch antennas. The antennas have a gain of 5.5 dB each and the transceiver power is 8.5 W (total EIRP of the channel was 14.7 dBW). The initial mass estimate for each of the antenna was 0.2 Kg at PDR, while the final value was 0.18 Kg. The risk analysis is performed using data statistic, expert statistic and Bayesian combined statistic. In the case of the S-Band low gain patch antennas, the steps followed to perform risk analysis using the database are:

1. Component type and category: low gain antenna.
2. Central frequency: S-Band.
3. Performance metric: antenna gain (5.5 dB).
4. Computation of probability density function and of tail function.

In the case of experts' statistic, different experts have been interviewed and the opinions have been composed applying the approach described in Chapter 5. The same considerations hold for the Bayesian approach.

The probability density functions and tail functions for HETE-2 antenna mass are shown in Figure 6-15 and 6-16. In this case, the initial (PDR) value and the final (CDR) value are very close. Hence, it is difficult to use this test case to validate the risk model. However, it is possible to notice that the Bayesian combined probability distribution is the one that achieves the best result in this test case. Specifically, the peak of the distribution corresponds to the final (CDR) value of the mass of the component, and the tail function of the distribution (together with the one computed using expert opinion) drops very quickly to zero, showing that the PDR value is overestimated and that there is very little risk of exceeding the PDR mass value. This example (Figure 6-16) shows the advantage of introducing expert methodology in modelling risk estimation. In fact, the experts' confidence in the antenna's mass not exceeding 0.2 Kg helps to reduce the probability of exceeding the initial estimate: a designer who might use Figure 6-16 to size the antenna's contingency can be very confident in allocating no more than 0.2 Kg for this antenna, since the risk of exceeding this value, according to the experts, is minimal. In contrast, if only database information is available, the designer needs to allocate more mass to mitigate the risk, since the probability density, based on data, does not seem to provide a strong confidence in the design value.

In the case of the transceiver, the initial mass was 1.5 Kg, while the final value was 1.72 Kg. The initial value for the power consumption was 30 W and the final 39.2 W. Using a procedure similar to the one described for the antenna, statistics based on data are generated for HETE-2 transceiver mass and power consumption. In the case of experts' statistic, different experts have been interviewed and the opinions have been composed by applying the approach described in Chapter 5. The same considerations hold for the Bayesian approach.

Probability density functions and tail functions of mass fluctuations are in Figure 6-17 and 6-18. For power consumption, the probability density functions and tail functions

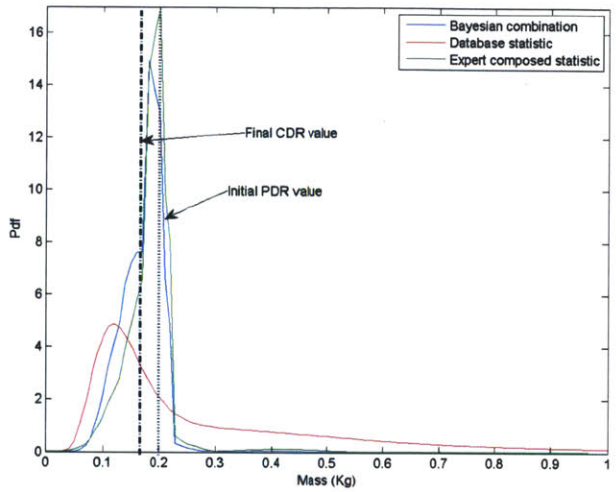

Figure 6-15: Probability density functions for HETE-2 antenna mass (S-Band channel). The peak of the Bayesian distribution is the closest to the final CDR value.

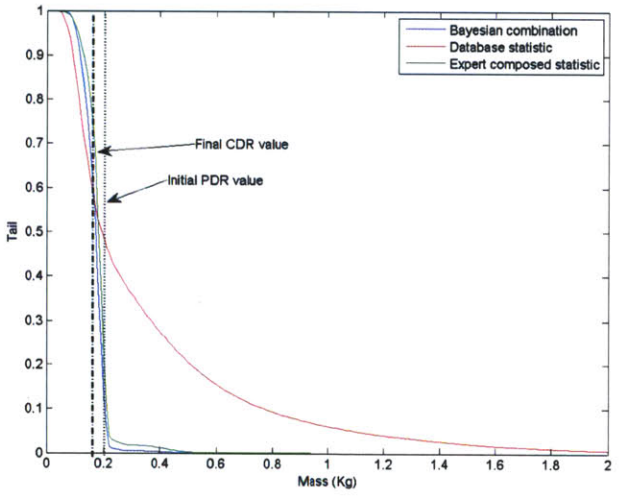

Figure 6-16: Tail function for HETE-2 antenna mass (S-Band channel). The experts' confidence helps to reduce the probability of exceeding the PDR value with respect to the statistic based on data.

are in Figure 6-19 and 6-20.

In the case of HETE-2 transceiver's mass, all three estimates fail in identifying the final value of the mass. The peak of the distribution for the data statistic probability density (Figure 6-17) is far from both the initial (PDR) and final (CDR) value.

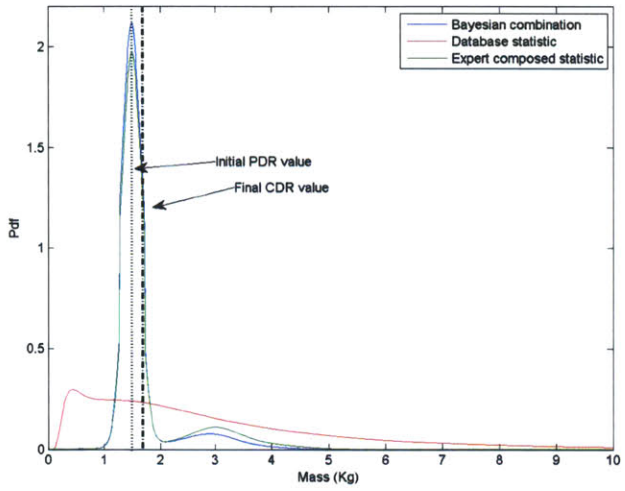

Figure 6-17: Probability density function for HETE-2 transceiver mass fluctuation (S-Band channel). The peak values for the three distributions are not centered on the final (CDR) value.

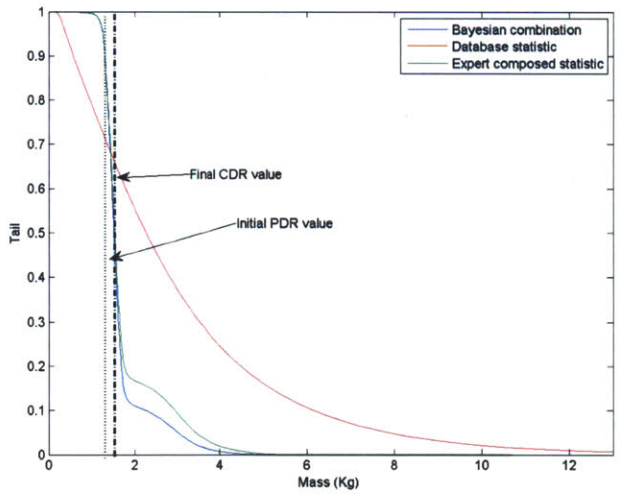

Figure 6-18: Tail function for HETE-2 transceiver mass fluctuation (S-Band channel). The risk quantification performed with expert opinion and Bayesian statistic allows the designer to reduce mass contingency with respect to the one performed using data statistic.

Differently, expert statistic and Bayesian statistic are centered in the initial (PDR) value and they do not show that this value is slightly underestimated. Similarly to the CASTOR transceiver, the data statistic probability distribution is more spread

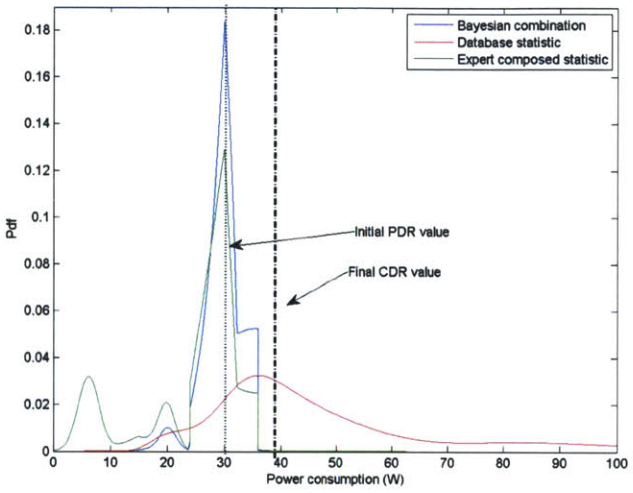

Figure 6-19: Probability density function for HETE-2 transceiver power consumption (S-Band channel). The peak of the data statistic distribution is centered on the final CDR value, while expert and Bayesian statistics are closer to the initial PDR estimate.

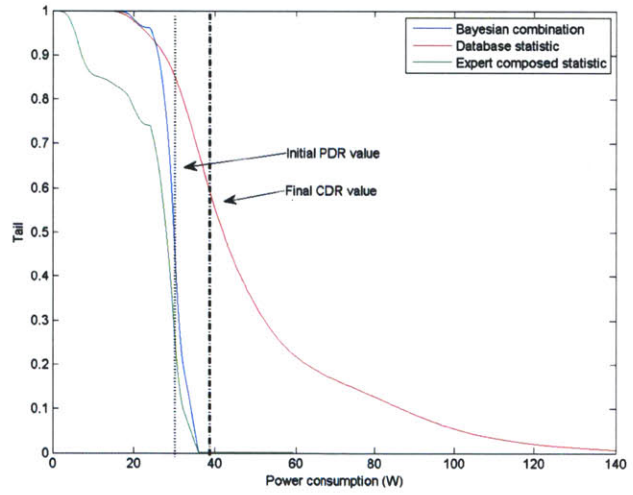

Figure 6-20: Power consumption tail function for HETE-2 transceiver (S-Band channel). The experts' confidence helps to reduce the risk of exceeding the initial PDR estimate and margin allocation.

with respect to other statistics. The tail function (Figure 6-17) shows once again (like the CASTOR test case) that the experts' confidence helps to reduce the probability of exceeding the initial PDR value. Also, in this case, if the designer had had to

size contingency only on the base of data statistic, the extra mass allocated would have been much greater than the one allocated by using expert statistic or Bayesian statistic.

In the case of HETE-2 transceiver power consumption (Figure 6-19), the database statistic correctly identifies the final value of power consumption: the peak of the distribution is centered in the CDR value. Expert statistic and data statistic underestimate power consumption. However, again the confidence of expert judgment reduces the uncertainty. In fact, the tail function (Figure 6-20) shows that expert statistic and Bayesian statistic may be inaccurate in predicting the peak of the density, but they provide a better identification for the risk of exceeding certain values. Specifically, tail functions computed through expert statistic and Bayesian statistic drop quickly just before the CDR value.

The analysis of the HETE-2 S-Band channel shows similarity with CASTOR. Probability density functions computed using data statistics tend to be more spread than the ones computed using expert statistics. Expert statistics and Bayesian statistics are not always able to identify the final value of mass and power (the peaks of the distribution do not always align with the CDR value). However, the tail functions show that experts' aggregated knowledge can provide a more reliable risk estimation: tail functions are generally less heavy than the ones computed with data statistics at CDR value. Hence, the mass or power margins that a designer would need to allocate on the basis of expert statistic would be less than the ones applied using data statistics. This aspect is important to improve the design of spacecraft communication systems, since excessively sized contingencies often lead to a very inefficient use of resources (Chapter 1).

The HETE-2 GPS channel was also analyzed. The system is composed of a 13.5 dB patch antenna and a GPS receiver. The initial value for the mass of the antenna at PDR was 0.2 Kg, while the final value was 0.33 Kg. The process of risk analysis using the three statistics has been applied to the GPS antenna and the results are shown in Figure 6-21 and 6-22.

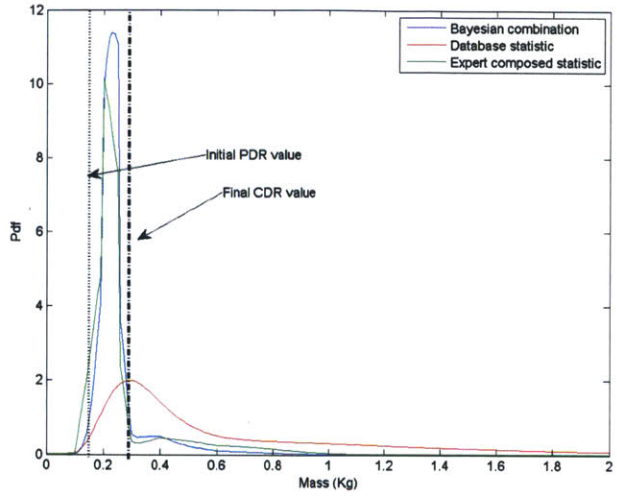

Figure 6-21: Probability density function for HETE-2 antenna mass (GPS channel). The peak of the data statistic is close to the final CDR mass value.

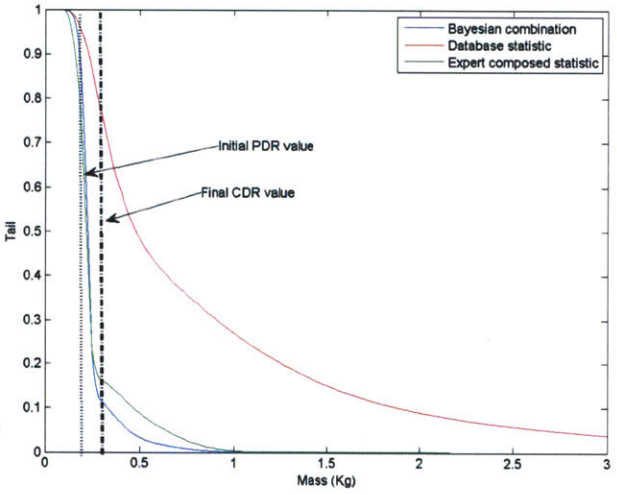

Figure 6-22: Tail function for HETE-2 antenna mass (GPS channel). Experts' confidence helps to reduce contingency allocation.

Also in this case, data statistic identifies correctly the final value of the mass (Figure 6-21). On the other hand expert statistic helps to reduce the sizing of contingency as shown in Figure 6-22.

For the GPS receiver, initial mass was 2 Kg and final mass was 1 Kg. The initial power consumption was 10 W, while the final value was 8 W. Statistics have been

computed using the same techniques previously discussed. Figure 6-23 and 6-24 show probability density function and tail function for GPS receiver mass, while Figures 6-25 and 6-26 show probability density function and tail function for GPS receiver power consumption.

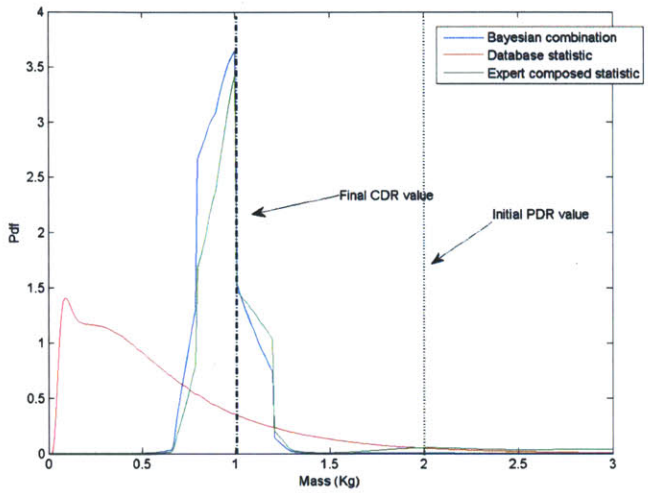

Figure 6-23: Probability density function for HETE-2 GPS receiver mass. The three statistics identify the initial overestimation of the PDR value.

In the case of the GPS receiver, the initial mass value at PDR is overestimated. Regarding the probability density functions (Figure 6-23), the three statistics identify that the PDR value is excessively conservative. Moreover, expert statistic and Bayesian statistic present a peak value which is very close to the final CDR value. Like the case of CASTOR transceiver power consumption, the model is able to identify when an initial value of mass and power is excessively conservative or excessively risky. Regarding the tail functions, (Figure 6-24), the GPS receiver mass is the only case in which the data statistic exhibits a lighter tail than the expert statistic. It may depend on the fact that the component is on a different frequency than the other test cases. The Bayesian statistic provides the best assessment for both density and tail function. This result supports the claim that Bayesian statistics seem to capture the positive aspects of both data statistics and expert statistics.

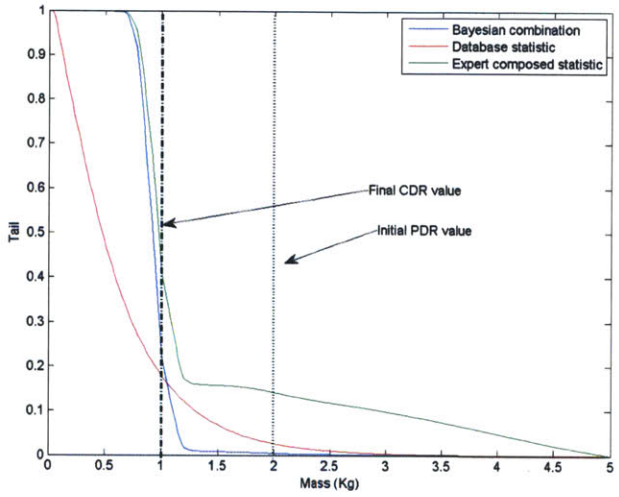

Figure 6-24: Tail function for HETE-2 GPS receiver mass. Bayesian statistic performs the best risk assessment for this case.

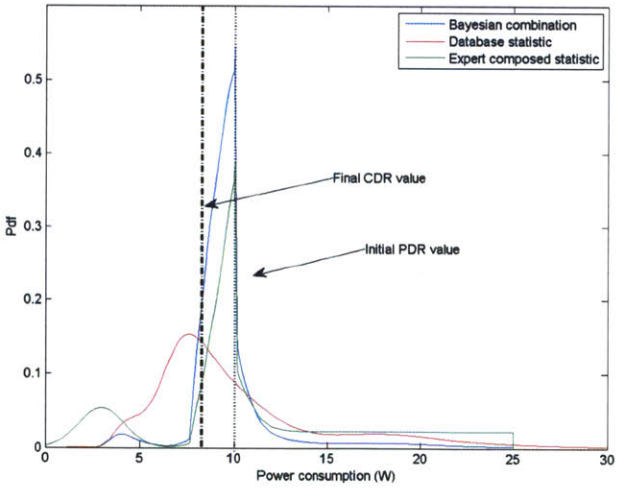

Figure 6-25: Probability density function for HETE-2 GPS receiver power consumption. Data statistic identifies the final value, while expert statistic and Bayesian statistic are anchored on the initial (PDR) estimation.

Regarding GPS receiver power consumption, the probability density function (Figure 6-25) computed using data statistic shows that the PDR value is overestimated. Differently, the other statistics are anchored on the PDR value. Regarding tail func-

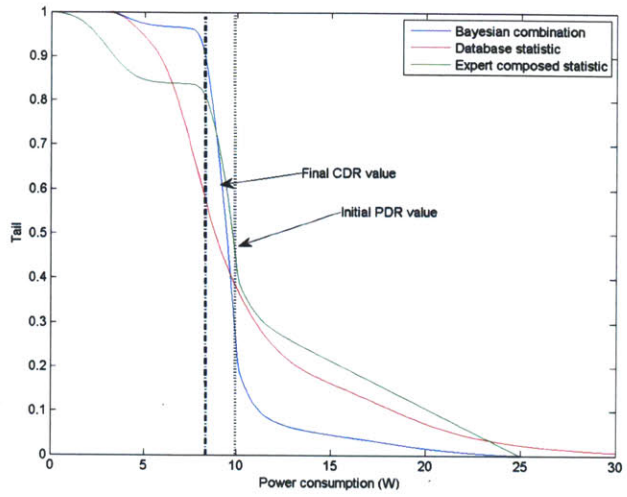

Figure 6-26: Tail function for HETE-2 GPS receiver power consumption. The three statistics exhibit heavy tails. Bayesian statistic provides the most satisfactory risk estimation.

tions (Figure 6-26), both expert statistic and data statistic exhibit heavy tails, while Bayesian statistic obtains a slightly better result. Conclusions on HETE-2 test case are discussed in the next subsection.

Conclusions

The analysis of the HETE-2 test case shows that data statistics help to identify the final value of mass and power of a component, since the peaks of those distributions generally corresponds to the CDR values. This result has been observed for the GPS antenna mass, for the S-Band transceiver power consumption, and for the GPS receiver power consumption.

On the other hand, expert statistics provide better assessments of the risk (as shown in tail functions), for almost all the test cases except for the GPS receiver mass. The tail functions estimated by using the expert approach show that if experts in the field have enough confidence in their estimations, the contingency allocation can be reduced with respect to the use of data statistics.

The Bayesian approach combines the two statistics and it improves the estimations:

tails are very light and the peaks of the distributions generally lie between the data statistics and expert statistics.

Despite some differences, all three statistics identify cases of high overestimation or underestimation (defined as cases in which the estimate is off by 100%). This result has been observed for CASTOR transceiver power consumption, and also for HETE-2 GPS antenna mass. In this last case, the PDR estimate was twice the final value of the mass, and all the three statistics identify the overestimation. The following section summarizes the analysis performed on the different test cases and provides guidelines and limitations in the application of the model to perform statistical risk estimation for communication systems.

6.3.3 Comparison with the Traditional Approach: Guidelines and Limitations

The test cases analyzed are compared with the traditional (contingency) approach. In the previous sections, each case is described and the probability density and tail function are discussed. From the previous section, it is clear that data statistics are helpful in identifying the final value of mass or power consumption. However, the most important information in terms of statistical risk estimation is carried by the tail functions. For those functions, Bayesian statistic provides the most satisfactory results.

A comparison between statistical risk analysis (using Bayesian statistics) and the traditional approach is shown in Table 6-27. The table lists:

- Name of the test case;
- CDR value (as known from technical documentation)
- Traditional approach: in this column the PDR value plus the contingency allocated is shown, together with a quantification of whether the PDR value is an over- or an underestimation with respect to the final CDR value. Finally, the risk at PDR is quantified by the model proposed in this thesis. The model

is able to identify cases of overestimation (risk of exceeding the PDR value is very low) and of underestimation (risk of exceeding the PDR value is very high, almost 100%).

- Thesis approach: in this column the mass and power allocations for a maximum risk of 10% are shown. The result column quantifies also whether the PDR allocation determined by the model is an over- or an underestimation with respect to the final CDR value. The last column indicates whether the performance of the new method is better, the same, or worse than the traditional approach.
- Observation: final comments on the test cases.

The test cases are also color-coded:

- Red: statistical risk estimation performs worse than the traditional approach in sizing contingencies.
- Yellow: statistical risk estimation slightly improves the risk assessment at PDR, but the sizing of the contingency is identical to the one obtained by applying contingency (no improvement).
- Green: statistical risk estimation improves risk assessment at PDR, and the contingency sized with the statistical method is closer to the final CDR value and/or allows the designer to save on mass and power allocations.

Nine test cases are discussed: six are green, two are yellow, and one is red. Statistical risk estimation improves risk assessment and the sizing of contingency in most of the cases, and it performs worse than the traditional approach only in one case. This result suggests that the methodology can improve the process of assessing the risk of exceeding design constraints in the communication systems, and it also allows the designer to avoid excessively sized contingencies. In fact, if the experts are very confident that a certain component will not exceed certain mass and power values, this research suggests that only limited contingency should be added.
To summarize the guidelines inferred from comparing the test cases to the state-of-the-art approach:

Test case	CDR value	Traditional approach			Thesis model (Bayesian statistis)			Observation
		PDR + Contingency	Result	Estimated risk of exceeding PDR (Thesis model)	PDR with 10% risk	Result	Comparison with traditional approach	
Castor antenna mass	0.1 Kg	0.55 Kg	Over by 900%	1%	0.35 Kg	Over by 850%	Better	Bayesian statistic shows overestimation (the risk at PDR is very low). The mass allocation using bayesian statistic is close to the final CDR value with 38-45% savings in mass allocation.
Castor transceiver mass	0.05 Kg	0.12 Kg	Over by 140%	48%	0.3 Kg	Over by 500%	Worse	Problem in recognizing the overestimation. There is no mass saving with respect to the state of the art.
Castor transceiver power consumption	4.5 W	1.65 W	Under by 63%	100%	5 W	Over by 11.1%	Better	Bayesian statistic shows underestimation (the risk of exceeding the PDR value is 100%). The final mass allocations by using bayesian statistic is close to the final value, while the state of the art approach would have generated a wrong estimation.
HETE-2 antenna mass(S-Band)	0.18 Kg	0.22 Kg	Over by 22%	10%	0.2 Kg	Over by 11.1 %	Better	Bayesian statistic shows overestimation (the risk at PDR is very low). The mass allocation using bayesian statistic is closer to a final value and allow the designer to save 30% in mass allocation.
HETE-2 transceiver mass (S-Band)	1.72 Kg	1.65 Kg	Under by 4%	94%	1.8 Kg	Over by 4%	Same	Bayesian statistics shows underestimation (the risk of exceeding the PDR value is 94%). However, the mass allocation obtained by using Bayesian statistic obtain is the same as the current practice
HETE-2 transceiver power consumption (S-Band)	39.2 W	33 W	Under by 15%	40%	37.5 W	Under by 4%	Better	Bayesian statistic recognizes the underestimation. In terms of mass allocation , bayesian statistic obtain a value close to the final CDR value which improves the statistic with respect to the state of the art.
HETE-2 antenna mass(GPS)	0.33 Kg	0.22 Kg	Under by 33%	91%	0.33 Kg	Exact value	Better	Bayesian statistic shows underestimation (the risk of exceeding the PDR value is 91%). The final mass allocations is close to the final value, while the state of the art approach generates a wrong estimation.
HETE-2 GPS receiver mass	1 Kg	2.2 Kg	Over by 120%	1%	1.2 Kg	Over by 20%	Better	Bayesian statistic shows overestimation (the risk at PDR is very low). The mass allocations using bayesian statistic is close to a final value and allow the designer to save 46-53% in mass allocation.
HETE-2 GPS receiver power consumption	8 W	11 W	Over by 37.5%	20%	11 W	Over by 37.5 %	Same	Bayesian statistic shows overestimation (the risk at PDR is very low). The mass allocation using bayesian statistic is close to a final value and to the state of the art.

Figure 6-27: Summary of test cases. Statistical risk estimation improves the mass/power allocation with respect to the traditional approach.

- Statistical risk estimation is preferable to the static contingency approach, since it performs better in most of the cases.
- Sizing contingency by using statistical risk estimation can avoid generating too risky solutions in the cases of underestimation, and it allows up to a 56% savings in mass/power allocation if the initial estimate is an overestimation.
- Across the three statistical techniques, Bayesian composition seems to exhibit the most satisfactory performance.

A limitation of this analysis is that only nine test cases have been considered, due to the limited availability of data. Future research should include more cases to confirm the results.

Another limitation of this approach is that probability density functions of mass and power are calculated independently. However, dependencies across the two metrics are captured by the baseline design. Moreover, the generation of the statistics without the hypothesis of independence would necessarily require the generation of bi-dimensional distributions. These statistics are difficult to model for the following two reasons:

- In the case of data statistics the curse of dimensionality causes an exponential increase in the number of samples required to model probability distributions. As shown by Silverman [115], the amount of data required to apply Kernel Density Estimator is in the order of hundreds of samples, which makes the application of this technique impossible.
- If expert statistics are used, it is shown that experts have difficulties in modelling multi-dimensional distributions [100]. At the actual state of the art, reliable methods to elicit bi-dimensional distributions do not exist.

For these reasons the model approaches the two aspects of mass and power consumption as independent in the statistical analysis. However, the dependencies across mass and power are still captured in the baseline design and in the optimization model described in Chapter 7. The next section summarizes conclusions for Chapter 6.

6.4 Summary

The first part of this chapter is dedicated to baseline design: parameters, modules, inputs, and outputs. A validation performed with past missions' data shows that the baseline design is able to produce a reliable estimate of a communication system design in terms of performance, average mass, and average power consumption.
The second part of the chapter focuses on the risk model. The model is described, and test cases are discussed. Two missions are used as test cases: CASTOR and HETE-2. The availability of PDR and CDR values for these missions is a key to corroborate risk estimation. Risk analysis is performed for each component of the communication systems for both missions: probability density functions and tail functions are computed. Results show that the three different statistics (data, expert and Bayesian) provide complementary information. Specifically, data statistics identify, in the great majority of cases, the proper final value of mass and power for the component. On the other hand, expert statistic improves the estimation of tail functions and shows how experts' confidence can reduce the mass margin allocated for risk. Bayesian composition combines the advantages of both techniques. The comparison with the traditional approach shows that statistical risk estimation improves the sizing of contingencies in most of the cases, saving up to 56% in mass/power allocation.
The next chapter describes the optimization model developed for the thesis.

Chapter 7

Optimization

This chapter presents the mathematical framework to perform statistical risk estimation as a constrained optimization problem. The framework allows the evaluation and comparison of different architectural solutions to identify the ones which maximize a certain objective while maintaining design risk below a certain threshold. The optimization model developed uses the statistical methods presented in Chapter 4 and in Chapter 5 to compute the risks for each of the architectures.
Chapter 7 is structured as follows: initially the mathematical framework is presented. Considerations on how the statistics developed in Chapter 4 and 5 are applied to the optimization process are also discussed. Finally, an example of the framework is presented. The results show how risk affects the number of allowable solutions, how the statistics affect the results, and how the Pareto front of the optimal solutions changes as a result of the application of risk analysis.

7.1 Mathematical Framework

The main objectives for the mathematical framework are as follows:

- Formulate statistical risk estimation as a constrained optimization problem: This objective includes the development of a framework with variables, objectives, and constraints to allow designers to compare and select solutions.

- Identify acceptable architectural solutions: Acceptable solutions are defined as the ones for which risk is below a certain threshold. The framework helps designers to distinguish between risky and non-risky solutions.

- Evaluate the effects of different statistics on the number of acceptable solutions: The number of solutions changes as the type of risk quantification changes. The results of this chapter show how the statistics impact risk assessment.

- Analyze how risk analysis affects the Pareto front of solutions: The set of optimal solutions changes as a function of the risk thresholds, and of the statistics.

To accomplish these objectives, a framework has been developed.
The framework aims to design a communication system which minimizes one or more figures of merit, given certain constraints. Risk can be treated as a constraint, but it can also be an objective in the cases in which the goal of the optimization is to minimize risk. Hence, different formulations can be proposed to approach the problem. In the specific case of this research, risk is considered a constraint.
As described in Chapter 6, the communication system is modelled as a set of channels at different frequencies. The channels are indicated with the index i (from 1 to k). The components inside each channel are indicated by the index j, which goes from 2 (a communication system includes at least an antenna and a transceiver) to n. Each component is described by mass, power, and performance metric. Hence, the vector of design variables is as follows:

$$\bar{X} = \begin{bmatrix} m_{1,1} & g_{1,1} & p_{1,1} & m_{1,2} & g_{1,2} & p_{1,2} & \cdots & m_{1,n} & g_{1,n} & p_{1,n} \\ m_{2,1} & g_{2,1} & p_{2,1} & m_{2,2} & g_{2,2} & p_{2,2} & \cdots & m_{k,n} & g_{k,n} & p_{k,n} \end{bmatrix} \tag{7.1}$$

The objective is to minimize mass and power consumption. The formulation is as follows:

$$\min k_1 \cdot M_{tot} + k_2 \cdot P_{tot} = k_1 \cdot \sum_{i=1}^{k} \sum_{j=2}^{n} m_{i,j} + k_2 \cdot \sum_{i=1}^{k} \sum_{j=2}^{n} p_{i,j} \tag{7.2}$$

k_1,k_2 are the weights of the two objectives respectively.
The constraints for the optimization are as follows:

- Constraints on total mass and power caps: Each communication system is limited in the maximum amount of available mass and power. This constraint refers to the average values of mass and power. The fluctuations around these averages are modelled by the risk constraint. The constraints on total mass and power caps are as follows:

$$\sum_{i=1}^{k}\sum_{j=2}^{n} m_{i,j} \leq M_{cap} \tag{7.3}$$

$$\sum_{i=1}^{k}\sum_{j=2}^{n} p_{i,j} \leq P_{cap} \tag{7.4}$$

where $p_{i,j} = 0$ if the component is passive.

- Constraint on single channel quality: Each channel has to guarantee a certain EIRP which is the logarithmic sum of the gains of the components. The expression, for each of the i channels, is as follows:

$$\sum_{j=2}^{n} g_{i,j} = EIRP \tag{7.5}$$

- Constraints on maximum allowed risks in mass and power: These constraints deal with mass and power fluctuations, and they represent the bulk of risk quantification. They establish that risk needs to be below a certain threshold. The formulations are as follows:

$$\int_{M_{cap}}^{\infty} f_{totmass}(x)\,dx \leq \alpha_{mass} \tag{7.6}$$

$$\int_{P_{cap}}^{\infty} f_{totpower}(x)\,dx \leq \alpha_{power} \tag{7.7}$$

α_{mass} and α_{power} are the risk thresholds for mass and power consumption. $f_{totmass}(x)$ and $f_{totpower}(x)$ indicate the probability density functions for total mass and power. They result from the convolutions of the distributions of the single components according to the hypothesis of independence. The

formulation is shown in the following equations:

$$f_{totmass}(x) = f_{1mass}(x) * f_{2mass}(x) * ... * f_{nmass}(x) \tag{7.8}$$

$$f_{totpower}(x) = f_{1power}(x) * f_{2power}(x) * ... * f_{npower}(x) \tag{7.9}$$

Each density is calculated as a function of the performance metric and of the type of statistic (data, expert opinion, or Bayesian composition). Details on the statistics are in the next section.

7.2 Statistics

To compute risk constraints, probability distributions for mass and power fluctuations are required. These distributions have been computed using the statistical approaches presented in Chapters 4 and 5.

In the case of data statistics, the densities are computed by using the same formula described in Chapters 4 and 6:

$$f_{Data}(x_i, x) = \frac{1}{g_i}\left(\frac{1}{n \cdot h}\sum_{j=1}^{n} K\left(\frac{x - a_j \cdot g_i}{h}\right)\right) \tag{7.10}$$

where a is a vector of n elements that collects the samples of mass per unit of gain for a specific type and category of components. $K(\cdot)$ is a normal kernel used for the Kernel Density Estimator, and x represents the support of the distribution.

For expert opinion, the computation of the probability densities is slightly different from the approach presented in Chapter 5 and 6, due to the fact that expert opinion collected in the previous chapters was limited to two test cases for which the experts provided estimations. In this case, however, the simulation is set to compare hundreds, or possibly thousands of solutions. It is clear that performing a separate elicitation for each solution would be impractical. Hence, in this chapter, the information from previous interviews is generalized to model expertise in communication

systems.

As the optimization model computes the probability density function for a component x_i (with performance metric g_i), the model selects a similar component for which the expert opinion is available (from CASTOR or HETE-2 analysis). The probability density for the known component is $f_{Expert}(c, x)$, where c is the component and x the support of the distribution. The component has a performance metric, defined as g_c. To compute the probability density for component x_i by using expert opinion, the density is scaled. The expression is:

$$f_{Expert}(x_i, x) = \frac{g_c}{g_i} \cdot f_{Expert}(c, x) \tag{7.11}$$

The third possible statistic used is a Bayesian combination of data and expert statistics as shown in Chapter 5.

The statistics described are based on real historical data and/or real expert opinion. In addition to *real* statistics, *simulated* statistics have also been applied. Simulated statistics are standard parametric distributions scaled by the average value of mass and power consumption of a component. Simulated statistics do not provide a risk assessment of the same quality as real statistics, but they are a valid mathematical instrument to check whether the model behavior is correct. In fact, as discussed in Appendix C, when a distribution is continuous, unimodal, bounded, symmetric around the mean and monotonically increasing on its left side, the uniform distribution represents an upper bound on the risk estimation. Hence, the simulated statistics used for this analysis are as follows:

- Uniform distributions: For each component x_i, probability densities for mass and power fluctuations are computed as a uniform distribution centered in average estimates (m_i,p_i), and bounded between $\pm 30\%$ of the average mass and power consumption (m_i,p_i).

- Triangular distributions: For each component x_i, probability densities for mass and power fluctuations are computed as a triangular distribution centered in

average estimates (m_i,p_i), and bounded between ±30% of the average mass and power consumption (m_i,p_i).

- Normal distributions: For each component x_i, probability densities for mass and power fluctuations are computed as a normal distribution centered in average estimates (m_i,p_i), and with a variance selected to bound the distribution between ±30% of the average mass and power consumption (m_i,p_i). Technically, normal distributions are unbounded. However, the model truncates the distribution at the fourth sigma interval.

Simulated statistics help to show aspects of the model. As discussed in Appendix C, uniform distributions achieve the upper bound in risk estimation with respect to distributions that are bounded in the same interval, symmetric around the mean of the uniform, continuous, unimodal and monotonically increasing on the left side of the density. Hence, since triangular and normal distributions satisfy these criteria, the risk computed using simulated uniform distributions should be higher than the one computed with any of the other two distributions. Acceptable solutions are the ones for which risk lies below a certain threshold. Hence, the number of solutions allowed in the case of uniform distribution is a lower bound with respect to the other two simulated statistics. As shown in the example described in the next section, this result has been verified in our model, providing a confirmation that the mathematical framework behaves accordingly to what is mathematically derived in Appendix C.
In the optimization model statistics are computed for each component. Then the total distributions are computed by convolving the distributions of each single component under the hypothesis of independence. The next section describes the model implementation with a test case and the results.

7.3 Test Case Implementation and Results

The mathematical framework has been implemented in Matlab. The analysis aims to compare only architectures which provide the same quality of service. These architec-

tures should provide the same EIRP; hence they are called iso-EIRP. To reduce the computational complexity and to compare only iso-EIRP solutions, a pre-processing is applied. When the pre-processing is completed, the optimization is performed. The optimization is applied, as an example, to the case of a communication system composed of two channels: one in the S-Band and one in the UHF-Band. Each channel is composed of a transceiver and an antenna. The minimum EIRP is 14 dBW. The architectures are generated by varying the transmitting power in discrete steps between 10^{-2} W and 10 W. Hence, the problem is discretized in a set of finite communication architectures. As a result, the solutions derived in this section are influenced by the granularity of the discretization: the number of acceptable solutions can increase/decrease proportionally if the granularity of the analysis is modified. The gain of the antenna is determined by imposing the iso-EIRP constraint. The mathematical framework for this example becomes:

- Variables' vector

$$X = \begin{bmatrix} m_{1,1} & g_{1,1} & p_{1,1} & m_{1,2} & g_{1,2} & p_{1,2} & m_{2,1} & g_{2,1} & p_{2,1} & m_{2,2} & g_{2,2} & p_{2,2} \end{bmatrix} \tag{7.12}$$

  The components with indexes (1,1) and (2,1) are the antennas, and their power consumption is already zero. Hence, a simplified version of vector X is:

$$X = \begin{bmatrix} m_{1,1} & g_{1,1} & 0 & m_{1,2} & g_{1,2} & p_{1,2} & m_{2,1} & g_{2,1} & 0 & m_{2,2} & g_{2,2} & p_{2,2} \end{bmatrix} \tag{7.13}$$

- Objective function: k_1 and k_2 are assumed equal to 1.

$$\min k_1 \cdot M_{tot} + k_2 \cdot P_{tot} = k_1 \cdot \sum_{i=1}^{k} \sum_{j=2}^{n} m_{i,j} + k_2 \cdot \sum_{i=1}^{k} \sum_{j=2}^{n} p_{i,j} = m_{1,1} + m_{1_2} + p_{1,2} + m_{2,1} + m_{2,2} + p_{2,2} \tag{7.14}$$

- Constraints:

  - Quality (iso-EIRP): They are computed for each of the channels (S-Band

is channel 1 and UHF is channel 2).

$$g_{1,1} + g_{1,2} = 14dBW \tag{7.15}$$

$$g_{2,1} + g_{2,2} = 14dBW \tag{7.16}$$

- Total caps in power and mass: The caps are fixed at 15 Kg of mass and 30 W of power.

$$m_{1,1} + m_{1,2} + m_{2,1} + m_{2,2} \leq 15Kg \tag{7.17}$$

$$p_{1,2} + p_{2,2} \leq 30W \tag{7.18}$$

- Total risk in mass and power below a certain threshold:

$$\int_{M_{cap}}^{\infty} f_{totmass}(x)\, dx \leq \alpha_{mass} \tag{7.19}$$

$$\int_{P_{cap}}^{\infty} f_{totpower}(x)\, dx \leq \alpha_{power} \tag{7.20}$$

where:

$$f_{totmass}(x) = f_{1,1mass}(x) * f_{1,2mass}(x) * f_{2,1mass}(x) * f_{2,2mass}(x) \tag{7.21}$$

$$f_{totpower}(x) = f_{2,1power}(x) * f_{2,2power}(x) \tag{7.22}$$

The optimization is performed for the different statistics previously described (data statistic, expert statistic, Bayesian combined statistic, simulated uniform, simulated triangular, and simulated normal), and for different thresholds of risk ($\alpha = 1\%, 5\%, 10\%, 20\%, 25\%$)). Results are discussed in the following sections.

7.3.1 Effect of the Different Statistics on the Number of Acceptable Solutions

Figure 7-1 shows the acceptable architectures for a risk threshold of 5%. In this case, expert statistics are applied. Each of the dots represents a single architecture. Any architecture is composed of the same number of components. However, the components selected to achieve the same EIRP are different, and the correspondent mass and power consumption are also different. Figure 7-1 shows the average values of mass and power consumption for any architecture. The probability of exceeding risks in mass and power are computed by using expert statistics. The colors of the dots show the results of the optimization. Specifically:

- Blue dots: They indicate the solutions which are violating the caps in mass and power consumption. These solutions are not even considered in the optimization process because if the means of the distributions are already exceeding the boundaries, the risk of exceeding those boundaries is very high (if the distributions are symmetric, greater than 50%). As a consequence, these architectures can be discarded even before running the optimization process.

- Red dots: They indicate solutions for which the total average values of mass and power consumption are in the boundaries. However, according to risk analysis, these solutions present a risk greater than 5% of exceeding the caps. Hence, these solutions are classified as *excessively risky* and they are not considered acceptable solutions.

- Green dots: They represent acceptable solutions for which risk is below the 5% threshold. In this set of solutions, two points are distinguished from the others: the yellow dot and the light blue dot.

  - Yellow dot: It belongs to acceptable solutions, and it is the one which achieves the minimum power consumption.

  - Light blue dot: It belongs to acceptable solutions, and it is the one which achieves the minimum mass.

For a low risk threshold (5%), the number of acceptable solutions according to expert statistics is very limited (24), while it increases for risk greater 5%. The limited number of acceptable solutions at low risk level indicates a conservative approach in expert estimation which has also been observed by [100] in different expert elicitation practices. However, a similar limited number of acceptable solutions is also observed when data statistics are applied to the optimization (Figure 7-3). Hence, the phenomenon can also be caused just by the specific probability distributions computed for this example. Also, the granularity of the analysis affects the number of solutions. The analysis of variance proposed in the last section of this chapter aims to identify the components which mostly influence mass and power fluctuations.

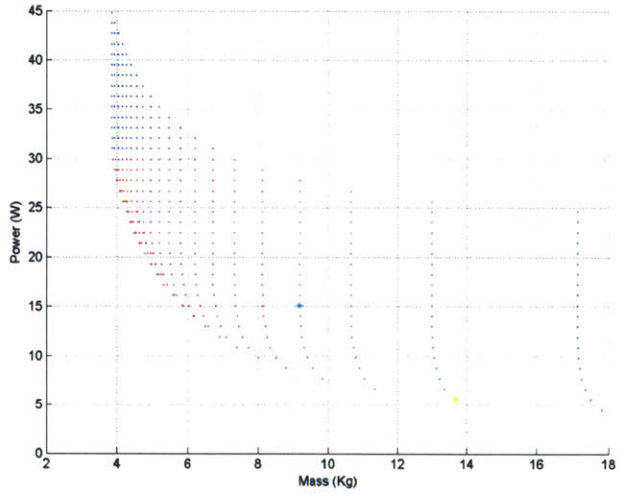

Figure 7-1: Summary of architectures at the 5% risk threshold, computed using expert statistics. The blue dots are unacceptable solutions because their average mass and power consumption violate the caps. The red solutions are also unacceptable because they have a risk of exceeding the caps of more than 5%. The green dots are the acceptable solutions (the yellow dot is the minimum power solution and the light blue dot is the minimum mass solution).

Another important aspect to study in the optimization is how the Pareto front of solutions moves by applying risk analysis. This analysis is shown in Figure 7-2. The risk threshold and the statistics are those used in Figure 7-1. The only difference is

that in Figure 7-2 the Pareto fronts for the two cases (with and without risk analysis) are isolated with respect to the other dots. Pareto fronts indicate the set of optimal solutions in the two metrics of interest (power and mass). A solution belongs to the Pareto front if a decrease in one of the two metrics causes an increase in the other one. The analysis performed for this case shows that the shape of the Pareto front moves as risk analysis is applied. This analysis is very important for the designers of communication systems, because it provides a visual way to compare solutions on the bases of risk, and to identify the optimal set of alternatives for which risk remains below a certain threshold.

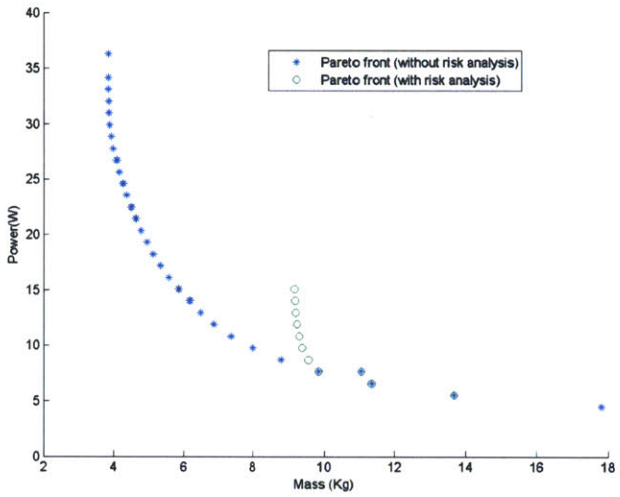

Figure 7-2: Pareto fronts comparison: the blue dots generate the Pareto front without risk analysis, while the green dots generate the Pareto front with risk analysis (5% risk threshold and expert statistics). Risk analysis moves the shape of the Pareto front.

The same analysis has been repeated for the same risk threshold (5%), but with data statistics. The set of acceptable solutions are the green dots shown in Figure 7-3. It is possible to notice the limited number (12) of acceptable solutions for very low risk thresholds. Also in this case, the Pareto fronts have been separated and they are shown in Figure 7-4.

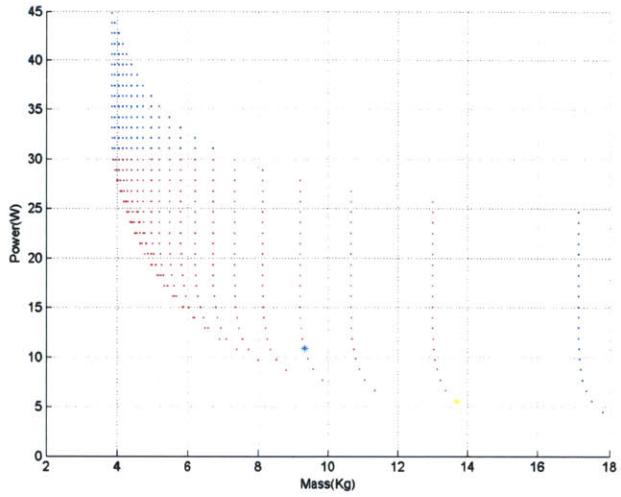

Figure 7-3: Summary of architectures at the 5% risk threshold, computed using data statistics. The blue dots are unacceptable solutions because their average mass and power consumption violate the caps. The red solutions are also unacceptable because they have a risk of exceeding the caps of more than 5%. The green dots are the acceptable solutions (the yellow dot is the minimum power solution and the light blue dot is the minimum mass solution).

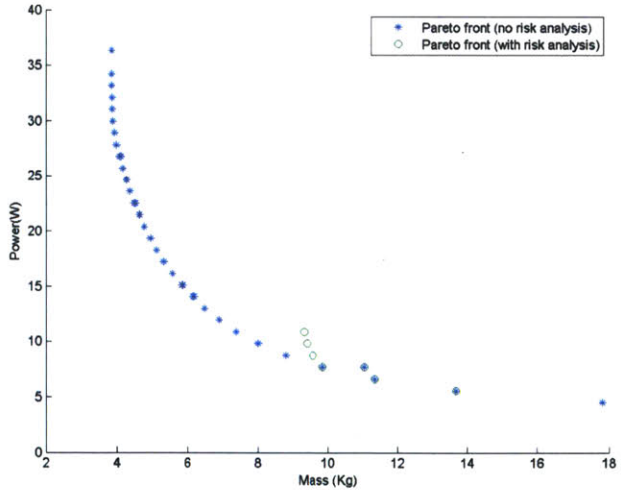

Figure 7-4: Pareto fronts comparison: the blue dots generate the Pareto front without risk analysis, while the green dots generate the Pareto front with risk analysis (5% risk threshold and data statistics). Risk analysis moves the shape of the Pareto front.

Similar results can be observed when Bayesian statistics are applied. (Figures 7-5 and 7-6).

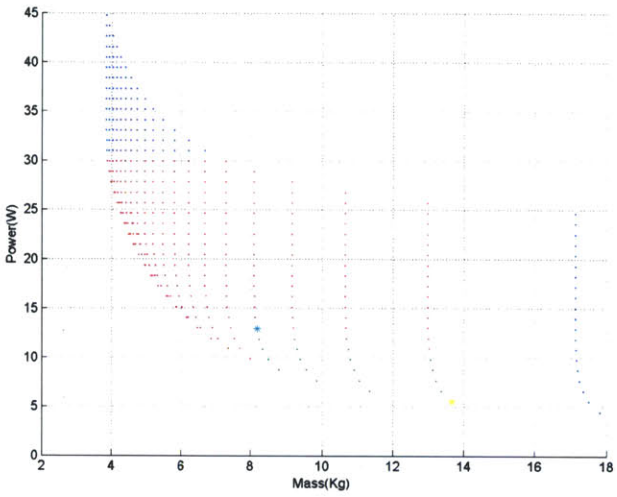

Figure 7-5: Summary of architectures at the 5% risk threshold, computed using Bayesian combined statistics. The blue dots are unacceptable solutions because their average mass and power consumption violate the caps. The red solutions are also unacceptable because they have a risk of exceeding caps of more than 5%. The green dots are the acceptable solutions (the yellow dot is the minimum power solution and the light blue dot is the minimum mass solution).

The sets of acceptable solutions are similar across the three statistics. This result is comforting since it shows that no matter which statistic is chosen, the model tends to give similar results. However, the sets present some differences. Hence, it is difficult to decide which set of solutions to select. Two approaches can be used to select the solutions. One approach is to consider acceptable only the solutions that are acceptable according to all the statistical techniques considered. In the specific case of this example, this indicates selecting only the 12 solutions suggested by data statistics, which are also acceptable according to the other statistics. The second approach is to examine each acceptable solution, maybe refining the statistical assessment by performing an ad-hoc expert elicitation similar to the one applied to CASTOR and HETE-2 (Chapter 6). The selection of the final solution has not been performed for

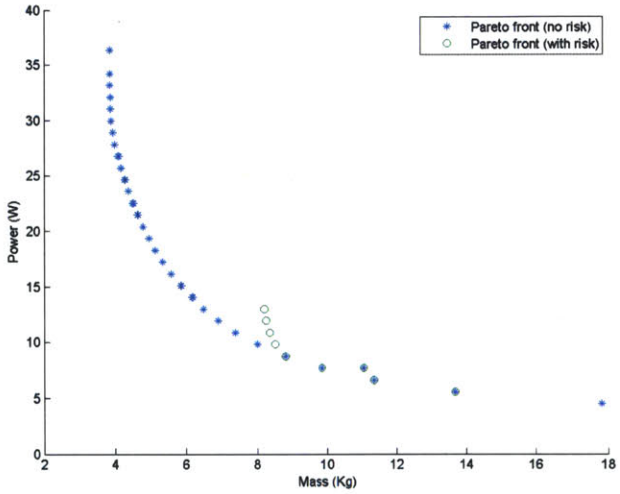

Figure 7-6: Pareto fronts comparison: the blue dots generate the Pareto front without risk analysis, while the green dots generate the Pareto front with risk analysis (5% risk threshold and Bayesian statistics). Risk analysis moves the shape of the Pareto front.

this example because the focus of this chapter is to show how statistical risk estimation can be a useful tool at the initial stages of the design when it is necessary to subselect specific options from a big set of alternatives. When the initial subselection is performed, a refined analysis on each solution can be developed.

A summary of the number of acceptable solutions for different risk thresholds is presented in Table 7.1.

The number of solutions shown in Table 7.1 validates the optimization tool proposed. The list of solutions show that the model behaves as expected. In fact, the number of acceptable solutions increases as the acceptable risk threshold increases. The three simulated statistics behave according to what has been predicted. The uniform distribution represents a lower bound in the number of acceptable solutions since uniform distribution achieves the upper bound in risk estimation under certain conditions (Appendix C). The number of acceptable solutions increases if triangular distributions and normal distributions are used, since these two distributions respect the conditions under which the upper bound is valid (Appendix C).

The real statistics exhibit different behavior due to the fact that they are not neces-

Table 7.1: Number of acceptable solutions for different thresholds of risk. The total number of architectures analyzed is 400. The number of acceptable solutions at low risk thresholds for expert statistics and data statistics is very limited.

Statistic	Allowed solutions for different thresholds of risk				
	1%	5%	10%	20%	25%
Uniform (simulated)	32	60	94	110	126
Triangular (simulated)	90	127	135	148	150
Normal (simulated)	130	140	145	150	159
Database statistic	10	20	86	130	154
Expert statistic	12	24	100	119	137
Bayesian statistic	6	20	56	150	154

sarily unimodal and symmetric, hence the upper bound conditions cannot be applied. Only a limited number of solutions are acceptable for the risk threshold below 5%. This result is due partially to the difficulty that experts have modelling extreme quantile probabilities [100], and to the limited number of samples for data statistics.

In conclusion, these results represents a validation of the optimization model as they show how risk analysis can be embedded in an optimization framework to select communication architectures as a constrained optimization problem. The results show that the model allows the designer to subselect the solutions for which risk is contained below a certain threshold. The application of risk analysis changes the Pareto front of the optimal solutions.

Regarding the effects of the statistics, the sets of acceptable solutions present some differences if data statistics, or expert statistics, or Bayesian statistics are used. However, certain architectures are acceptable for each of the three statistics. This result is important in facilitating the selection of acceptable solutions. However, the expert statistics used for the optimization problem are generalized with respect to the data collected for the test cases described in Chapter 6. Hence, a more refined analysis with specific expert elicitation is suggested to finalize the selection of the optimal solution.

7.4 Application of Optimization to PDR Design

The previous plots do not show the probability density. However, each point previously described is surrounded by a probability density in the two axes of mass and power. This probability density is important to characterize each of the solutions at PDR in order to allow the designers to identify the best solution. The bi-dimensional probability distribution, which describes the fluctuations of the architecture in the two metrics (mass and power), can be computed under the assumption of independence across mass and power. The hypothesis of independence is assumed in Chapter 6 when the risk model computes risk. The dependencies between mass and power consumption are resolved in the baseline model and after that each probability distribution is computed independently. As a consequence of the previous assumption, the joint distribution can be computed as the product of the marginal densities. Defining the mass support (x) and the power support (y), the joint distribution $f_{risk}(x, y)$ is computed as:

$$f_{risk}(x, y) = f_{mass}(x) \cdot f_{power}(y) \tag{7.23}$$

The joint distribution is computed for all of the architectures shown in the previous plots. Showing all the joint distributions together into a unique plot would appear extremely chaotic. Hence, Figure 7-8 shows the contour plots of the joint probability densities for four selected points. The points considered are examples of two acceptable solutions (specifically the ones with minimum mass and minimum power), and of two unacceptable solutions. The exact location of these points with respect to the other solutions is indicated in Figure 7-7.

The contour plot shows that, as expected, the unacceptable solution presents a high level of risk. In fact the contour lines are pretty much out of the acceptable borders (solid red lines) that indicate the caps. In the cases of acceptable solutions, the contours of the probability densities indicate that only a small portion of the probability density lies outside the boundaries. The shapes and the magnitudes of each of the four densities shown are significantly different. They are the result of convolutions of various probability densities for different components. The joint probability dis-

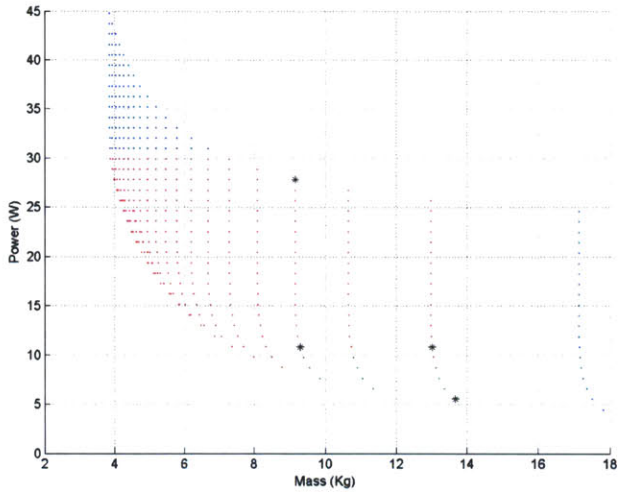

Figure 7-7: Solutions for 5% risk threshold and data statistics. The black points are the solutions shown in Figure 7-8.

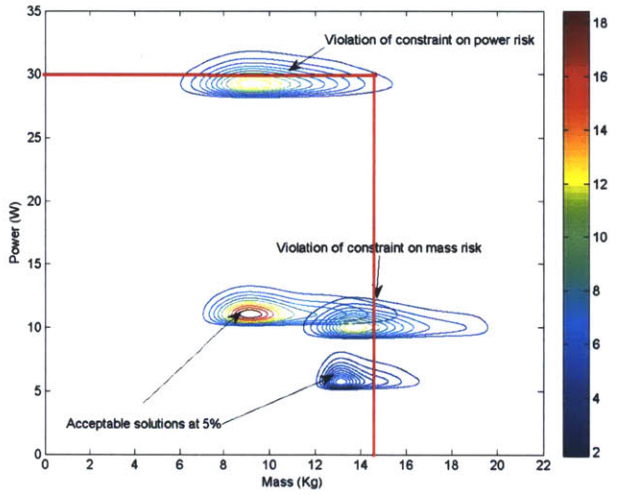

Figure 7-8: Contour view of joint probability densities for mass and power consumption fluctuations. The risk threshold is 5% and the statistics used are data statistics. The four points selected include two acceptable solutions (minimum mass and minimum power respectively), and two unacceptable solutions. The amount of density which exceeds the two boundary (red) lines determines whether a solution is acceptable or not.

tributions allow the designers to identify solutions that are acceptable and solutions that are too risky by looking at how close the cloud of uncertainties are to the caps

in mass and power. Another important tool to compare design options at PDR is the analysis of variance proportions, which is proposed in the next subsection.

7.4.1 Analysis of Variance Proportions

Analysis of variance proportions is a way to perform sensitivity analysis on the optimization result obtained. The analysis considers how the variance of each factor affects total variance. It is very similar to the ANOVA method [18], but it is more suitable for non-linear systems with non-normal distributions. The objective of this analysis is to identify the main drivers for mass and power fluctuations.
To compute the analysis, the hypothesis of independence between the distributions of the single components is assumed. Hence, the variance of fluctuations for mass and power is the sum of the single variances, as shown in the following two equations:

$$\sigma_{Mtot} = \sum_{i=1}^{n} \sigma_{Mi} \tag{7.24}$$

$$\sigma_{Ptot} = \sum_{i=1}^{n} \sigma_{Pi} \tag{7.25}$$

The variance proportions' coefficients are defined as the proportion of each component variance with respect to the total variance. They are computed as:

$$c_{Mi} = \frac{\sigma_{Mi}}{\sigma_{Mtot}} \tag{7.26}$$

$$c_{Pi} = \frac{\sigma_{Pi}}{\sigma_{Ptot}} \tag{7.27}$$

The number of mass coefficients is the same as the number of components, while the number of power coefficients is the number of active components, because for the passive ones the probability density function is zero. Hence, in the specific case of the example discussed, the analysis features four mass coefficients (S-Band antenna, S-Band transceiver, UHF antenna, and UHF transceiver), and two power coefficients (S-Band transceiver and UHF transceiver). The analysis can be performed for any of the possible dots (solutions), and for any of the possible statistics. In this section, the

analysis has been performed for two dots: the yellow dot and the light blue dot of Figure 7-3. The yellow dot is selected because it is the non-risky solution which achieves the minimum power consumption according to data statistics, to expert statistics, and to Bayesian statistics. The light blue dot is the non-risky solution which achieves the minimum mass according to data statistics. This solution is not the optimal one according to expert statistics and to Bayesian statistics as shown in Figure 7-1 and in Figure 7-5. However, it is still an acceptable (non-risky) solution according to these two statistics.

The minimum power consumption solution (yellow dot) corresponds to a set of components listed in Table 7.2. The table shows average values of mass and power consumption for the proposed design, as computed by the baseline model. However, the mass and power for each of the components fluctuate according to the statistics computed by the risk model and shown in Figure 7-8. For each component, the risk

Table 7.2: Breakdown of components mass and power consumption for the minimum power communication architecture. The influence of the different components on the total fluctuation is analyzed with variance proportions' coefficients.

Component	Average mass (Kg)	Average power consumption (W)
S-Band transceiver	1.5	2.8
S-Band antenna	3.5	0
UHF transceiver	1.65	3.3
UHF antenna	7	0
Total	13.6	6.1

model computes mass and power fluctuations. The analysis of variance compares the ratios (coefficients) made up of the variance of each component over the total variance of the distribution. The coefficients for mass fluctuations are shown in Figure 7-9. The coefficients for power fluctuations are shown in Figure 7-10. In Figure 7-9, the S-Band transceiver is the most significant contributor to the total variance, even if it is not the heaviest component (the average mass is only 1.5 Kg). Generally, both transceivers, the UHF and the S-Band, seem to contribute to the variance more than the two antennas. Another important aspect to notice is that the three statistics

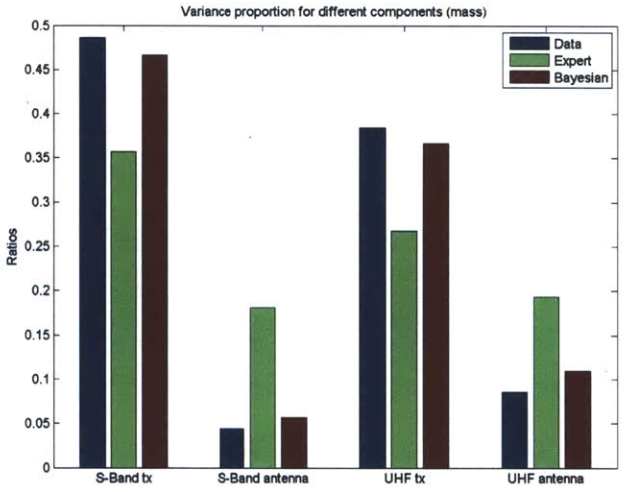

Figure 7-9: Variance coefficients (Ratios) for mass fluctuations computed using different statistics (minimum power solution). The S-Band transceiver is the most significant contributor to mass fluctuations according to the three statistics.

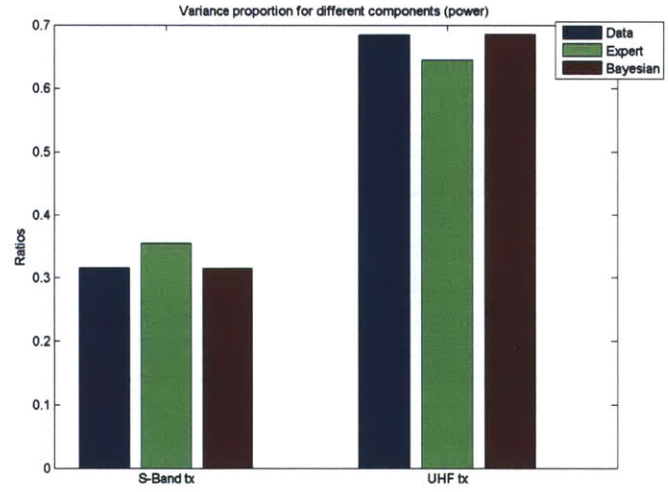

Figure 7-10: Variance coefficients (Ratios) for power fluctuations computed using different statistics (minimum power solution). The UHF transceiver is the most significant contributor to power fluctuations according to the three statistics.

agree on which component is the biggest contributor to the variance.

Figure 7-10 shows the power coefficients. The three statistics agree that the UHF transceiver is the main driver of uncertainty on the power consumption fluctuation.

The same analysis described for the minimum-power solution has been performed for the minimum-mass solution. The objective is to investigate how variance proportions change by selecting different architectures. The minimum-mass solution (light blue dot) corresponds to a set of components shown in Table 7.3. The table shows average values of mass and power consumption for the proposed design. This architecture, differently from the minimum power consumption case, presents a total mass of 9.2 Kg, at least a third less than the minimum-power solution. However, the power consumption is almost doubled. This architecture is designed with more powerful transceivers that are slightly more massive, but with a higher power consumption. Since the transceivers transmit more power, the antennas require less gain. As a consequence, the antennas' mass is reduced.

For each of the components the risk model computes mass and power fluctuations.

Table 7.3: Breakdown of mass and power consumption for components for the minimum-mass communication architecture. The influence of the different components on the total fluctuation is analyzed by using variance coefficients.

Component	Average mass (Kg)	Average power consumption (W)
S-Band transceiver	1.8	5
S-Band antenna	1.3	0
UHF transceiver	2.2	6.1
UHF antenna	3.9	0
Total	9.2	11.1

The analysis of variance proportions compares the ratio (coefficients) made up of the variance of each component over the total variance of the distribution. The coefficients for mass fluctuations are shown in Figure 7-11. The coefficients for power fluctuations are shown in Figure 7-12. In Figure 7-11, the different statistics disagree on what is the biggest driver for mass fluctuation. For data statistics the main driver is the S-Band antenna, while for Bayesian statistics the main driver is the S-Band transceiver. According to expert statistics, S-Band and UHF antennas are the drivers for the mass fluctuation. In this case, it is very difficult to identify the main contributor to variance fluctuations. In similar cases, a more refined expert

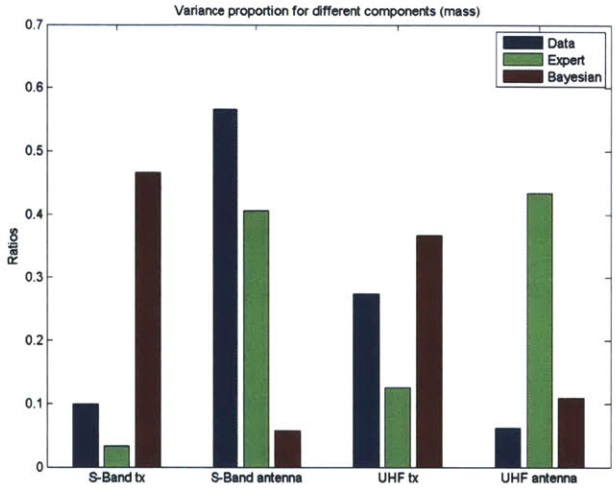

Figure 7-11: Variance coefficients (Ratios) for mass fluctuations computed using different statistics (minimum mass solution). The three statistics offer different results.

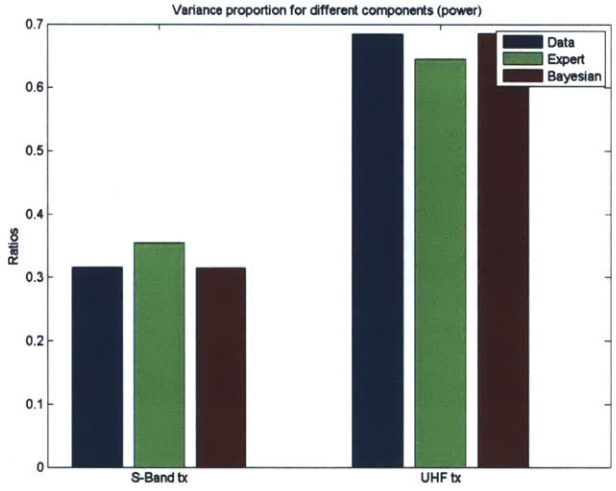

Figure 7-12: Variance coefficients (Ratios) for power fluctuations computed using different statistics (minimum mass solution). The UHF transceiver is the most significant contributor to power fluctuations according to the three statistics.

elicitation is probably recommended. Differently, regarding power consumption (Figure 7-12), the situation is almost identical to the minimum-power design point: the UHF transceiver is the biggest driver according to the three statistics.

In conclusion, analysis of variance proportions can be applied to the different design

solutions to identify the main drivers of mass and power fluctuations. The tool helps to break the design into the different components to further investigate which components are causing the greatest impact on the distributions. The analysis is useful when the three statistics agree, while when the statistics present contrasting results a deeper investigation is required.

7.5 Summary

Chapter 7 presents the mathematical framework for optimization. Each component is described by three elements (mass, power and gain). The objective function is to minimize mass and power while respecting constraints on the signal quality (EIRP), and on the risk of exceeding boundaries (caps) in mass and power consumption. The statistics used to compute the risk constraints are the same as those described in Chapters 4, 5, and 6. However, in Chapter 7 multiple solutions need to be compared in a limited amount of time. Hence, expert elicitation cannot be performed on each of the solutions. As a result, in this chapter the knowledge of experts has been generalized. Finally an example of the optimization framework is presented. The case of a satellite with two communication channels in S-Band and UHF is discussed. The optimization is performed for different risk thresholds (1%, 5%, 10%, 20%, and 25%), and for different statistics (experts, data and Bayesian). The result shows that risk analysis changes the shape of the Pareto front. Each of the different statistics produces a slightly different set of optimal solutions. However, some solutions are acceptable for all the three statistics: these solutions can be considered for a further analysis with more refined statistics. In fact, the main goal of this optimization is to provide a tool at the initial stage of the design to identify the most promising solutions, and to compare them in a risk trade-off perspective.

The next chapter discusses model extensions to additional metrics like cost, and generalizations of the model to different subsystems.

Chapter 8

Model Extensions and Generalizations

An interesting feature of the approach is the possibility of extending it to different metrics, and of generalizing it to other subsystems by modifying only a limited portion of the modules. The chapter is structured as follows: first, an extension to cost risks is discussed. This discussion includes a literature review on cost modelling, identification of causes for cost fluctuations, and adaptations of the thesis approach to the new metric. An example developed with the CASTOR mission is proposed. Finally, model generalization to different subsystems is presented through the application of this research to the META project. An overview of META and of the synergies between META and statistical risk estimation is presented. Then, an application of statistical risk estimation to finite element analysis for a mission studied with META (TerSat satellite) is discussed.

8.1 Model Extension to Cost Risks

8.1.1 Overview

This section is dedicated to the extension of the model to cost fluctuations. Cost estimates fluctuate over time in a similar fashion to fluctuations in mass and power

estimates. However, causes for cost fluctuations are slightly different from the causes of mass and power fluctuations. As a consequence, the methodology proposed is slightly different. In fact, the first step to extend the model to different metrics is to investigate the causes for metrics' fluctuations in order to properly adapt the methodology.
The following sections present: principal literature works in cost modelling and cost-risk analysis, adaptation of the methodology, and application to the CASTOR test case.

8.1.2 Literature Review in Cost Modelling

Different models have been developed to assess spacecraft costs. The approach developed by Harmon, Warfield and Rosenberg in [59] is focused on cost modelling as a point estimate. However, the authors underline the necessity of pairing the estimation with a cost risk analysis tool to model uncertainties. They identify three main sources of uncertainties in cost estimation: model uncertainty, inputs uncertainty, and correlation between model uncertainty and inputs uncertainty. They suggest modelling the fluctuations with a lognormal distribution centered in point estimate and with a variance of 36% (calculated as a fitting of empirical data).
A different methodology is presented by Lillie and Thompson in [80]. They based their cost estimations on expert opinion (without a quantification of the quality of the expert), and their analysis is developed only as point estimation. A similar approach is proposed by Rosenberg, Hihn, Roust, and Warfield [107] in the DNP (Develop New Projects) model, and by Sherif and Remer in estimating the cost for Deep Space Network (DSN) [114].
Larson and Wertz in [135] classify the causes for cost fluctuations in three main categories:

- Fluctuations due to lack of human interaction: exactly as in Chapter 1 of this thesis for mass and power consumption.
- Fluctuations due to lack of knowledge: exactly as in Chapter 1 of this thesis for

mass and power consumption.

- Fluctuations due to technical difficulties: These fluctuations are due to the fact that some components may be fabricated for the first time with associated challenges in the fabrication process. As a consequence, cost increases as more human labor is required to solve fabrication issues. Technical difficulties are generally modelled as a function of Technological Readiness Level (TRL) as in [135].

This classification of fluctuations' causes resembles the approach detailed in the previous chapters. Hence, this classification is used in this research to analyze cost. When the causes for fluctuations are identified, they can be modelled using the approach previously developed for mass and power. Similarly to what has been done in the rest of the thesis, fluctuations due to lack of human interaction are not modelled, since they can be overcome by the development of interactive facilities like NASA JPL Team X and ESA CDF (Chapter 1). Hence, in this extension to cost modelling, only fluctuations due to lack of knowledge and to technical difficulties are considered. The next section summarizes the methodologies developed to analyze cost fluctuations.

8.1.3 Methodology

The methodology is adapted from the approach described in Chapter 3, 4, 5, and 6. As previously described, the risk of exceeding a cost threshold is computed at the component level and at the system level. The total cost risk is the sum of the risks for each single component. Under the hypothesis of independence across components, the total probability density is the convolution of the respective densities:

$$C_{Risk_{tot}} = C_{Risk_1} + C_{Risk_2} + ... + C_{Risk_n} \Leftrightarrow f_{Risk_{tot}}(x) = f_{Risk_1}(x) * f_{Risk_2}(x) * ... * f_{Risk_n(x)} \tag{8.1}$$

The cost risk for each component is calculated as the composition of two sources: lack of knowledge and technical difficulties.

$$C_{Risk_1} = C_{Risk_{1\,knowledge}} + C_{Risk_{1\,technical}} \Leftrightarrow f_{Risk_1}(x) = f_{Risk_{1\,knowledge}}(x) * f_{Risk_{1\,technical}}(x) \tag{8.2}$$

In terms of statistics, both risks due to lack of knowledge and to technical difficulties are computed as a Bayesian update of data information and expert opinion. An example of the model is proposed in the next section.

8.1.4 Example: CASTOR

The CASTOR mission was designed entirely at MIT over the past three years. The advantage of using this test case is the availability of cost data for the communication system.

As described in Chapter 6, CASTOR communication system includes two transceivers, and three antennas. Cost analysis is focused on these components.

For CASTOR antenna, the initial estimate (at PDR), was about $1,000 while the final cost was $1,400. Differently, the transceiver initial cost estimate was about $800, and the final cost was $750.

In our model, PDR data are used as initial point estimation (no baseline model has been implemented), and fluctuations are calculated as a composition of lack of knowledge and of technical difficulties. To compute the statistics, a real database and real expert elicitations are not available. However, this case is just an example to show the extendibility of the methodology. Hence, database information was substituted with already existing statistics. These statistics can be found in the literature [135], and they are computed through fitting relations using existing databases. Their characteristics are as follows:

- Data statistics for fluctuations due to lack of knowledge: For both components (antenna and transceiver), statistics are modelled as normal distributions centered in the initial point estimate and with a variance of 36% (value estimated through data fitting [135]).

- Data statistics for fluctuations due to technical difficulties: They are modelled

as normal distributions centered in the point estimate and with a variance that depends on the TRL (Technological Readiness Level) as in [135]. Specifically, the variance selected for the transceiver was 10%, correspondent to a TRL of 7, which was appropriate since the transceiver had a flight heritage. For the antenna, the variance was 25% which corresponds to a low TRL reflecting the new stage of development of the antenna.

Regarding expert opinion, a formal elicitation process (as described in Chapter 5) was not performed due to the unavailability of experienced engineers. To present this example, expert opinion was artificially introduced by the author, who worked actively in designing CASTOR satellite in the past. Hence, the information reflects the uncertainty in the cost values that the team faced at PDR. Specifically, the probability distributions used to model expert opinion are as follows:

- Expert statistics for fluctuations due to lack of knowledge: At PDR the transceiver was already developed by a commercial supplier (COTS). Consequently, very little cost fluctuations were expected. Hence, the statistic was modelled as a normal distribution with 10% variance. However, the antenna design at PDR was ongoing and different possible dielectric materials where under consideration. Since dielectric materials are key elements in cost estimation for antennas, expert opinion was modelled as a normal distribution centered in point estimate with a 40% variance.

- Expert statistics for fluctuations due to technical difficulties: also in this case expert opinion was introduced by the author. On the basis of similar considerations to the ones discussed for lack of knowledge, the expert opinion for the transceiver is assumed as a normal distribution with a variance of 5%. For the antenna, the distribution is also assumed normal and the variance is 30%.

The statistics are composed using the Bayesian approach. Figures 8-1 and 8-2 show respectively the PDF (probability distribution function) and the tail function of cost fluctuations for the CASTOR transceiver. In each plot, the red curve represents data statistics, the green one expert statistics, and the blue one the Bayesian composition.

In this example, expert opinion is beneficial to the statistic of cost fluctuations since

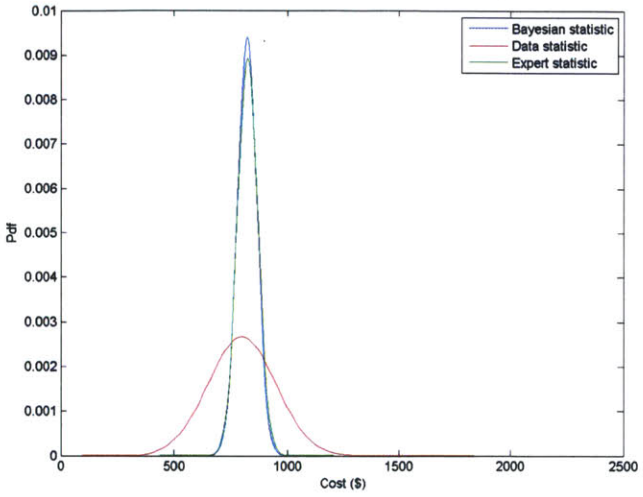

Figure 8-1: Cost fluctuations for the CASTOR transceiver (PDF). Each statistic includes the two sources of fluctuations. Expert and Bayesian statistics show a skewed distribution. The knowledge of the level of development of the component helps to reduce the cost uncertainty.

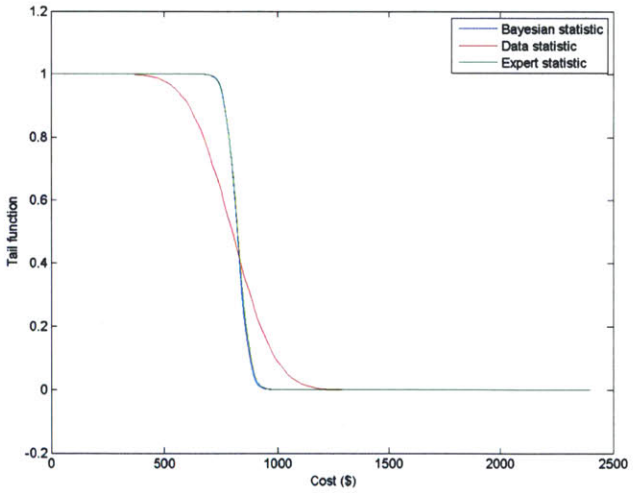

Figure 8-2: Cost fluctuations for the CASTOR transceiver (tail). Each statistic includes the two sources of fluctuations. Expert and Bayesian statistic show a skewed distribution. The knowledge of the level of development of the component helps to reduce the cost uncertainty.

the knowledge of the expert reduces the uncertainty. This phenomenon happens gen-

erally when the component is well characterized.

If, however, the component is in fabrication the three statistics provide wide variance. In fact, in the case of the antenna, the design changed multiple times over the project's evolution: different dielectric materials were considered to improve radiation properties. Figures 8-3 and 8-4 show cost fluctuations for the CASTOR antenna. The strong variance is evident in all the statistics. The following section presents

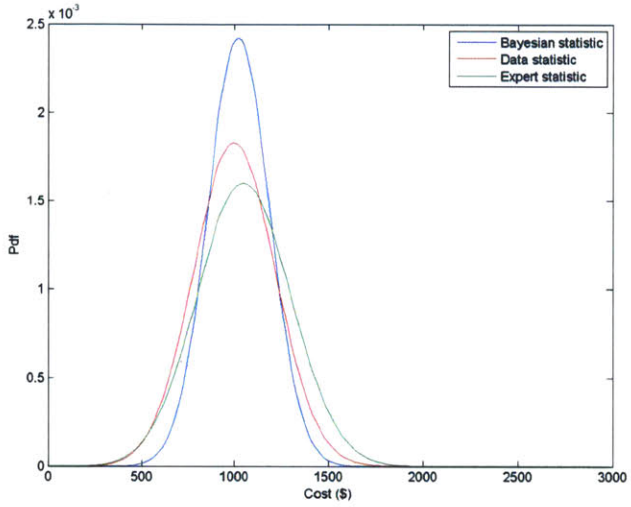

Figure 8-3: Cost fluctuations for the CASTOR antenna (PDF). Each statistic includes the two sources of fluctuations. The uncertainty in the design produces a high variance fluctuation according to the three statistics.

conclusions on the model's extensions and on the CASTOR example.

8.1.5 Conclusion and Guidelines on Extensions

An extension of the model to cost fluctuations was discussed. A literature review explored state of the art in cost modelling. An approach derived from the one developed in Chapters 3, 4, 5, and 6 is proposed. The main difference between the cost model and the model developed previously in the thesis is that fluctuations due to technical difficulties also need to be considered for cost fluctuation. For each source of fluctuations, the statistics can be generated as described in Chapters 4 and 5.

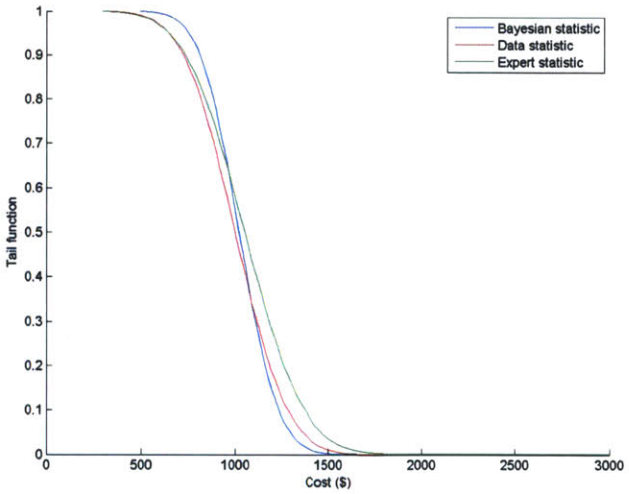

Figure 8-4: Cost fluctuations for the CASTOR antenna (tail). Each statistic includes the two sources of fluctuations. The uncertainty in the design produces a high variance fluctuation according to the three statistics.

The model is tested with the CASTOR satellite. Cost fluctuations for antenna and transceiver are modelled and composed into a unique estimate. Previously developed statistic [135] are used to replace the unavailable samples for the data statistics. Expert opinion is also artificially introduced. The model shows that expert opinion can decrease cost risk whenever the component is known and developed (CASTOR transceiver). A different case is verified when the uncertainty on the component's development is high (CASTOR antenna).

In general, the example shows the procedure to extend statistical risk estimation to different metrics. The key steps are as follows:

1. Define new metric: in this case cost.

2. Identify fluctuations' causes: in this case lack of human interaction, lack of knowledge, and technical difficulties.

3. Select the fluctuations that will be statistically modelled: in this case lack of knowledge and technical difficulties.

4. Create a database and implement approach described in Chapter 4. Alterna-

tively, as in this case, use already developed data statistics.

5. Perform expert elicitation as described in Chapter 5. Alternatively, use already available expert statistics.

6. Compute risks.

8.2 Model Generalizations

The idea behind generalization is that very few aspects of the model need to be changed in order to extend it to other spacecraft subsystems or to the whole system. Specifically, the key idea of generalization is presented in Figure 8-5. The blocks that

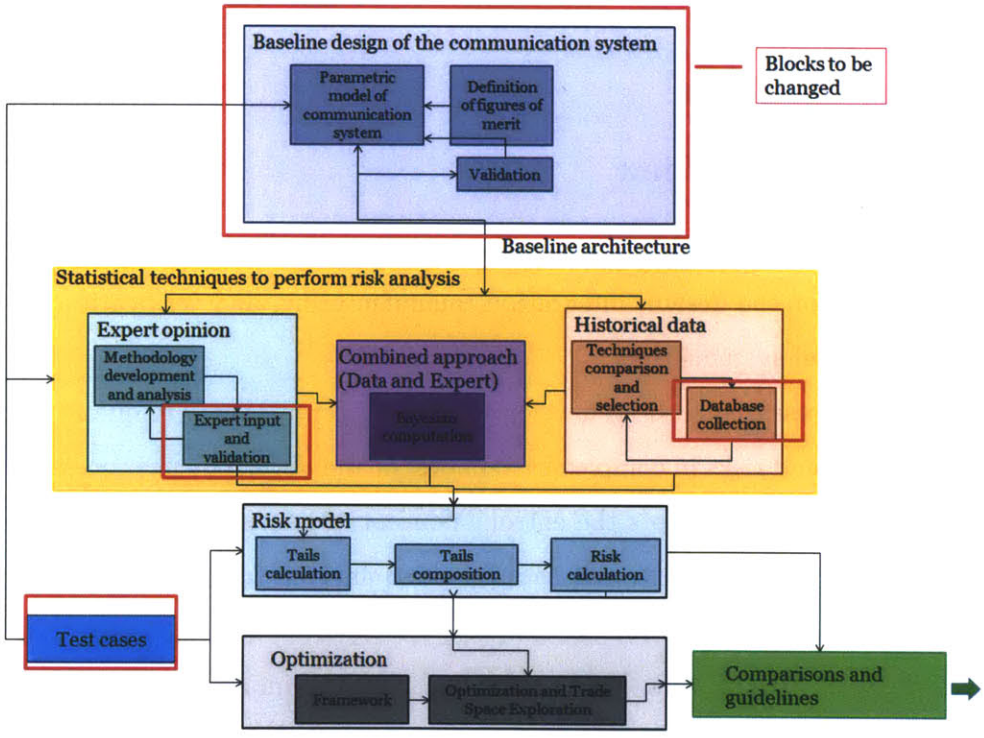

Figure 8-5: Generalization of the approach. The red blocks need to be changed as the model is applied in a different context.

need to be modified are:

- Baseline model: Baseline generates an initial estimate of the quantity of interest. Depending on the specific subsystem, different models can be used.

- Expert inputs: The expert modelling approach is the same as the one described in Chapter 5. However, Part 2 and Part 3 of the interview need to be adapted to the specific case analyzed.

- Database: A new database needs to be generated to apply the methodology of Chapter 4. If no data is available, expert opinion only is applied.

- Test cases: Test cases with data from different stages of the design need to be selected. In this way, it is possible to test the predictive nature of the model [11].

An example of model generalization is currently under study as part of the META project. The following subsections present an overview of the META project and of the application of statistical risk estimation.

8.2.1 META Overview

META is a project funded by Darpa. The goal is to reduce the level of effort and timeline needed for the design, integration, manufacturing, and verification of complex cyber-physical systems [136].
The MIT META research group is developing a fundamental theory to quantify inherent uncertainties and risks in complex system design and development processes [136]. The approach developed adapts the entropy concept of information theory to create a metric for system complexity, and uses Bayesian analysis to model propagation of uncertainty over the evolution of the design.
In the context of META, this thesis project can help by improving statistical characterizations of uncertainty for the metrics of interest. The methods described in Chapters 4 and 5 can be used at the initial stage of the design, when a statistical assessment of the uncertainties is generally lacking.
The MIT META team uses different test cases to validate the research approach. One of them is a spacecraft currently being designed at MIT Space System Laboratory, called TerSat (Trapped Energetic Radiation Satellite). The following section

describes the analysis performed by the MIT META team and how the methodology developed in this thesis has been applied to that particular context.

8.2.2 Application of Statistical Risk Estimation to META: Natural Frequency for TerSat Structure

TerSat [31] is a satellite currently developed at MIT Space System Laboratory. It is an ESPA-class satellite developed as part of a two semester CDIO (Conceive Design Implement and Operate) undergraduate project class. The satellite aims to provide a better understanding of how Very Low Frequency (VLF) waves interact with plasma in the Van Allen Radiation Belts [7]. TerSat will orbit at 550 km altitude, deploy a 5-m dipole antenna (two 2.5 m antennae), transmit less than 1 kW at 3-50 KHz, and measure the strength of its own echoes using a VLF receiver. The project is part of the UNP-7 (University Nanosatellite Program) competition, sponsored by AFOSR (Air Force Office of Sponsored Research). The satellite is now in its CDR stage, and if it wins the competition it will be launched in the next couple of years.

To win the competition, the team has to provide evidence that the satellite is accomplishing mission requirements and the safety requirements imposed by AFOSR. One of those requirements concerns the natural frequency of the satellite's structure, which has to be greater than 100 Hz, to guarantee safety in the launch of the satellite, as a secondary payload on an Air Force Rocket.

The MIT META team is helping the TerSat team with this issue by trying to model the natural frequency of the structure and by trying to measure the uncertainty around the estimate of this quantity of interest. A picture of the current satellite's structure is shown in Figure 8-6.

Statistical risk estimation can contribute to assess the uncertainty on this metric. In fact, natural frequency can be seen exactly as mass and power consumption for communication systems. Hence, the approach developed in this thesis can be applied to this problem as well. The steps followed to adapt the model to this problem are described as follows. Note that these steps can be used in a general sense to apply the

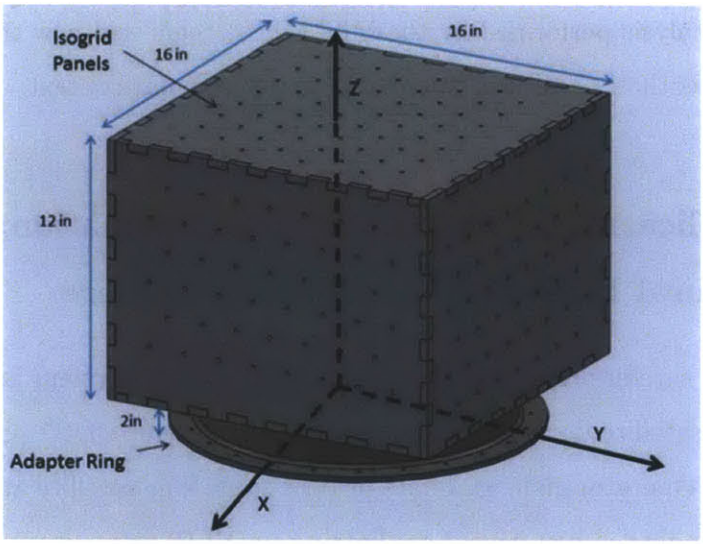

Figure 8-6: CAD model of TerSat's structure. The natural frequency has to be greater than 100 Hz.

model to a wider range of problems. The first step (Problem definition) is a pre-step, which means that it happens before applying the approach summarized in Figure 8-5. The other steps listed below corresponds to the changes required for the red blocks of Figure 8-5.

Problem Definition and Identification of Causes for Fluctuations

Problem definition includes defining the type of risk and the metric of interest. In this case, the problem is to characterize the uncertainty of the natural frequency for TerSat structure, and to model the risk that it will exceed 100 Hz. The new quantity of interest (metric) becomes the natural frequency of the structure. Regarding the identification of the causes for fluctuations, the design team is very small, hence lack of human interaction does not play a significant role in this case. The lack of knowledge due to the fact that the structure is still under development represents the most significant source of fluctuations. Specifically, the components need to be finalized and the boxes that will contain them need to be designed. The sizes of the boxes and their placement in the structure affect the natural frequency. The material of the structure itself and its design also affect the natural frequency. Hence, the

model aims to characterize fluctuations due to lack of knowledge.

Baseline Adaptation

As in the case of communication systems, baseline design provides an initial design estimate of the metric of interest. In this case, a structural model for TerSat and the correspondent Finite Element Analysis (FEA) provides the initial point estimation of the natural frequency for the structure.

Database Adaptation

To compute data statistics, an adaptation of the database is necessary. Specifically, new samples need to be collected: FEA data and data from similar spacecrafts can be used to create the database.

Expert Elicitation Adaptation

Part 1 (Probabilistic Thinking) of the interview/methodology (Chapter 5) remains the same, while Part 2 (Calibration) and Part 3 (Elicitation) need to be modified. The interview is currently under design. The questions necessary to develop Part 2 are point estimation questions that refer to similar design problems. Similar spacecraft developed at MIT and elsewhere are used as examples. For these missions the values of natural frequency are known. The expert receives information on the basic characteristics of the structure (size, mass, boundary condition, center of gravity, sensitivity of inside components), and then he/she is asked to estimate the natural frequency. The results are compared to the real values to estimate expert calibration. Then the expert estimates for the quantity of interest: the natural frequency of TerSat. The results of FEA for TerSat and the initial estimate of the natural frequency are shown to the expert. He/she is asked to model the probability distribution that represents the fluctuations around the natural frequency. Then expert opinions are aggregated according to the methodology described in Chapter 5.

Test Cases for Approach Validation

In this case test cases are not required, because vibration tests can be used to acquire the real value of the natural frequency. The approach should show that the statistics help in quantifying the risk of not exceeding the 100 Hz constraint in the natural frequency.

The study is ongoing, and the interviews will be performed in the next months. This study aims to repeat expert elicitation multiple times over the evolution of the design to measure a reduction of fluctuations as more knowledge of the structure is acquired.

The example shows how the approach can be generalized to different subsystems. The steps previously described can be used as guidelines in the generalization process. In terms of model' s application, a manager who has to decide between multiple options can use the generalized model to quantify the risks for each solutions and to obtain an estimate of the contingencies for different risk thresholds, as shown in Chapter 6 for CASTOR and HETE-2.

8.3 Summary

This chapter provides guidelines on how to extend the model to different metrics and on how to generalize it to different subsystems.

To extend the model to different metrics, the causes for fluctuations of the metrics of interest need to be analyzed, as they can be different from the ones identified for mass and power in Chapter 1. In fact, the extension to cost fluctuations reveals that it is necessary to include technical difficulties in product development as one of the causes.

Multiple steps need to be followed to generalize the methodology to different subsystems. In this chapter, the generalization is shown through an example taken from the META project: risk analysis for TerSat natural frequency. The example illustrates the procedure to generalize statistical risk estimation for different problems.

Chapter 9

Conclusions

9.1 Thesis Summary

This thesis explored the problem of fluctuations of mass and power consumption for communication system components over the evolution of the design. Fluctuations are a challenge for designers, since they can cause designs to exceed their constraints. As a result, designs may need to be drastically redesigned, and the process can be expensive in terms of time and human resources involved. Traditional approaches to the problem include contingencies which are applied at the component-, subsystem- and system-levels. However, contingencies are generally computed as fixed percentages with respect to the design values, and they do not take into account the real probability that a component, subsystem, or system will exceed a certain constraint. Hence, contingencies can lead to either overconservative designs, which generate an inefficient use of resources, or risky designs, which do not fit within the constraints and which require re-design. This thesis approaches the problem by developing a statistical methodology that computes the probability distributions of mass and power fluctuations. These distributions are used to properly size contingencies. In this way, contingencies are no longer arbitrary values, but rather they are based on real probabilities that the mass and power consumption of a component will exceed specific values. The research is focused on the communication system because it can be a key driver in mass and power for small satellites, commercial satellites, and relays for

deep space exploration. The analysis described in Chapter 1 underlines that communication systems can drive up to 28% of the total spacecraft mass, and up to 32% of the total spacecraft power consumption. Moreover, this thesis aims to be a proof of concept for the approach which can be further generalized to different subsystems and to the whole system. Chapter 1 also provides a detailed explanation of the thesis topic, together with a formal definition of the problem statement.

Chapter 2 discusses relevant literature areas for the thesis. The works are categorized into three topic areas: risk characterization, statistical techniques, and modelling and optimization for communication systems. The first literature area collects works related to risk estimation. Most of them are qualitative analyses, and the few quantitative works that are available lack the development of an integrated methodology which uses data and expert elicitation. The second literature area includes works in the field of probability density estimation and of expert elicitation. Different techniques in the field of density estimation have been developed; however, a comparison of the techniques for small sample sizes is missing. Regarding expert elicitation, different approaches have been developed and the most applied is Cooke Classical Model. However, Cooke's model requires the use of seed variables with known probability distributions which are unavailable for this study. Hence, a new methodology in the form of a three-part interview has been developed for this thesis. The third literature area explores the state of the art of modelling and optimization for communication system design. The works in this area lack a communication design tool which includes design risk analysis, and this gap is filled by this thesis.

Chapter 3 summarizes the key aspects of the approach applied in the thesis. The principal modules are introduced: baseline design, data statistics, expert statistics, risk model, optimization, extensions and generalizations. The test cases used to validate the methodology (CASTOR and HETE-2) are also briefly presented. Chapters 4, 5, 6, 7, and 8 describe the different modules.

Specifically, Chapter 4 presents the methodology developed to use historical data to construct probability density functions for mass and power consumption of different components. Initially, an analysis of different density estimation techniques is performed. This analysis aims to identify the most suitable technique to model a database composed of few samples (in most categories no more than 20). The density estimation techniques compared are: Histogram, Naïve Estimator, Kernel Density Estimator (for different kernels: normal, triangular, uniform, Epanechnikov), Nearest Neighbor Estimator, Variable Kernel Estimator, and Saddle-Point Estimator. The comparison is performed by using known benchmark distributions such as normal, exponential, beta, gamma, and lognormal. These distributions are randomly sampled and the samples are inputs for the different estimators. The estimators generate the tail function (complement of the CDF). Finally, the estimated tail function is compared with the real one and the estimation error is computed. The analysis reveals that Kernel Density Estimator with normal kernel achieves the lowest error (best performance) across all the estimators. Estimation error is greater for benchmarks that are non-normal, but in all the cases the Kernel Density Estimator performs better than the other techniques. Hence, Kernel Density Estimator with normal kernel is selected as the technique to be used to model data statistics. The methodology to compute data statistics includes the development of a database of components. The components are divided into different categories. To compute the probability density functions, type and performances of the components are used to select the proper vector of samples. Then, Kernel Density Estimator and a scaling process are applied to generate the distributions. Test case examples reveal that data statistics are accurate in predicting the final value of the component's mass and power consumption. In fact, the final value of the component is in most cases very close to the peak of the probability density. A drawback of the data statistics is that the sparsity of data generates very heavy tail functions which can cause excessive contingency allocations. This problem is solved by introducing expert elicitation.

When the number of samples is extremely limited (i.e., less than 10), or when the

characteristics of the components evolve as the result of technological evolution, it is important to add to the approach statistics based on expert opinion. Experts in the field can provide insights on the fluctuations for mass and power consumption. However, a methodology is necessary to ensure that the information collected represents experts' true beliefs. Chapter 5 describes the expert elicitation methodology developed for this research. The methodology is centered on a three-part interview which quantifies: the ability of experts to model probabilistic events (Probabilistic Thinking), the ability of experts to estimate quantities in their field of expertise (Calibration), and the elicitation of the quantities of interest for selected components (Elicitation). The results from Probabilistic Thinking and from Calibration are used to weight the opinions of different experts and to aggregate them into a unique estimate. The interview was tested with MIT students (17 undergraduate students, 17 graduate students) and with engineers (8 subjects from NASA JPL). The performance of the approach was compared with the case of experts equally weighted. For each test case, performance criteria were defined and a score from 0 to 5 was given to the estimates. Results show that the methodology developed improves the estimation with respect to the cases in which expert are evenly weighted: performance of the aggregation method proposed in this thesis is 4.3 over 5, while the performance of the aggregation with experts equally weighted is 2.75 over 5. Moreover, expert estimates are an improvement with respect to data statistics in term of tail functions, which become lighter as a result of experts' confidence. However, the peaks of experts' distributions do not always align well with the final real value. In Chapter 5 a Bayesian approach to compose data and expert information is proposed. The Bayesian combined approach improves the estimation with respect to the statistics based on data and to the statistics based on expert opinion, since it inherits the positive aspects of both estimation methods. Bayesian estimates are centered on the final design value of the component, as in the case of data statistics, and their tail functions (risk) are light, as in the case of expert statistics.

In Chapter 6 the risk model uses the previous statistics to perform risk analyses.

However, before computing risk, an initial point estimate of the system is required. The initial point estimate is basically a communication architecture which includes the list of components. For each of the components, the design value of mass, power, and gain is estimated. This initial estimate may be already available in the form of a PDR design, or it can be generated by baseline design. Baseline design is a module developed in this thesis which computes an initial estimate of a communication system design given certain parameters. The model is an integrated STK-Matlab tool, and it has been validated by comparing its result with real missions designed at NASA JPL. Once the initial design is computed, then the risk model is applied. Tail functions are computed for each component in mass and power, and they are compared with the real PDR and CDR values of mass and power for nine different test cases (components from CASTOR and HETE-2 missions). Test cases are divided into cases of overestimation (where the initial PDR estimate is too conservative, bigger than the final CDR value), and cases of underestimation (where the initial PDR estimate is smaller than the final CDR value). Results show that the Bayesian statistic is the best approach in almost every overestimation case. Its estimate is the closest to the final design value, and it also exhibits a light tail function in comparison with the CDR value. A comparison across the overestimation cases, between the mass/power that would have been allocated by using the traditional approach, and the one that would have been allocated by using the method proposed in this thesis, reveals that the use of this thesis approach would have reduced mass and power allocation, saving between 4% and 45% of the resources. Results for the cases of underestimation show that the model proposed in this thesis would have been able to identify all of the risky cases, and that the contingency applied by using the approach developed in this thesis would have been closer to the final CDR value.

Chapter 7 is dedicated to the development of an optimization framework. The framework aims to include risk estimation so as to optimize the system while keeping the estimated risk below a certain threshold. The framework developed aims to select the communication architecture which minimizes mass and power consumption.

Constraints on link quality (iso-EIRP) and on risk are introduced in the problem. The statistics used are the same previously discussed, with the exception of expert statistics that are generalized to be applied to multiple architectures. Results show that the application of risk analysis modifies the shape of the Pareto front. Also the analysis of variance allows designers to identify the components that are most influential in driving fluctuations in mass and power consumption.

Chapter 8 presents extensions and generalizations of the model. The extension to cost fluctuations is used as an example. The key step in the extension of the model is enabling the identification of causes of fluctuations in the design values. When the causes are identified, the statistics can be used to model each of the causes separately. Regarding generalization to different subsystems, the example of the TerSat structural analysis is used to illustrate the process of generalization. In the case of TerSat, risk analysis aims to assess the likelihood of not exceeding 100 Hz in the natural frequency of the structure. The methodology developed in the thesis can be applied to this problem through a redefinition of the baseline design, a change in the values of the database, and changes in the Calibration and in the Elicitation parts of the expert interview.

The next sections describe thesis contributions and suggestions for future work.

9.2 Contributions

The primary contribution of the thesis is as follows:

Creation of a statistical model to improve PDR design by assessing the risk that a certain component or a group of components in the communication system will exceed the caps in mass and power as the design matures: This thesis presents the first model to perform statistical risk assessment for spacecraft communication systems. This model is unique with respect to previous works

in the field because it is the first quantitative approach which integrates data and expert information to estimate the probability that mass and power of components will fluctuate over the evolution of the design.

Additional contributions are detailed below:

Creation of an approach, based on known benchmarks, to compare density estimation techniques, and identification of the Kernel Density Estimator as the most appropriate technique to model space mission historical data: This thesis provides, for the first time, a comparison of density estimation techniques which is independent from the data sets, since it uses benchmark distributions. Hence, the results of this comparison are of general applicability, independently of the data sets used. The comparison provides, for the first time, an assessment of performance of density estimation techniques for small sample sizes through a numerical analysis, since no analytical approach can be used as Silverman [115] underlines. The result of the analysis identifies the Kernel Density Estimator as the most appropriate technique for the thesis. An approach to use Kernel Density Estimator to model the data of this thesis is also developed.

Creation of a methodology to elicit expert opinion by measuring expert sensitivity to biases and mis-calibrations: This thesis presents a unique methodology to elicit expert opinion, while ensuring a quantitative measurement of experts' performance. The model developed in this thesis is substantially different from the state of the art in expert elicitation, especially with respect to the Cooke Classical Model. Cooke's model uses only calibration to assess experts' performance, and he measures calibration by using seed variables for which the probability density is available. In this thesis, the unavailability of seed variables, with correspondent probability densities motivated the necessity of developing a new approach. The assessment of performances of experts developed in this model, does not require the availability of probability densities for seed variables, and it is organized into two

parts: Probabilistic Thinking which checks the ability of experts to model probabilistic phenomena, and Calibration which measures the ability of experts to estimate quantities in their field of expertise. The approach proposed in this thesis is general and can be adapted to different problems, and it provides an alternative to Cooke's model for cases in which probability densities for seed variables are unavailable.

Creation of an approach to aggregate opinions from multiple experts into a unique estimate: The aggregation of opinions from different experts is performed by ranking the experts and by applying a bisection method to assign the weights. This approach is different from the state of the art in which the weights are: proportional to experts' scores (Hagan in [100]), or binary values derived from hypothesis testing (Cooke in [34]). The problem with proportional weights is that they tend to penalize the top-performing experts, while the problem with binary weights is that some experts (i.e., those weighted zero) are not considered. The bisection method offers a compromise because it takes into account contributions from every expert, while giving greater weight to the most promising ones.

Development of a methodology, derived from Bayesian analysis, to integrate expert opinion and historical data: The methodology proposed in this thesis is unique since it incorporates, for the first time, historical data and expert statistics. The Bayesian approach is used to compose the two statistics into a unique estimate, by exploiting all the sources of information used to generate the statistics.

Application of the model to test cases and improvement with respect to traditional approaches: The analysis of test cases provides the unique opportunity to show the advantages given by the model in real situations. The analysis performed reveals how the model identifies cases of overestimation and underestimation. Moreover, the test cases verify that the model provides a more accurate set of statistics, and a more efficient utilization of mass and power resources with respect to traditional approaches.

Development of a mathematical framework to model risk estimation for a communication system as a constrained optimization problem: The framework proposed incorporates, for the first time, risk analysis in the optimization framework. The tool developed represents a first attempt to pair state of the art design tools for the communication system with a methodology that complements the design by providing risk analysis. Changes in the Pareto front of optimal solutions show that risk is an important factor to consider as different solutions are compared.

Identification of guidelines on performing risk estimation for space communication systems: The model developed in this thesis provides guidelines to perform statistical risk estimation for communication systems. The guidelines are:

- Risk estimation cannot be performed without a clear understanding of risk causes. In this thesis, the primary causes of design risk are distinguished as lack of human interaction and lack of knowledge.
- Risk estimation is beneficial to communication system design when it is performed using statistics, meaning that contingencies should not be fixed percentages of the component's design values, but rather they should be the result of a statistical analysis.
- Risk estimation should be performed in the form of a statistical assessment which incorporates data and expert opinion. In this way, the methodology exploits complementary available sources of information.

Identification of guidelines to generalize and to extend the methodology to other subsystems up to the system level: The methodology presented in this thesis is general in nature and likely to be useful in performing risk estimation for different metrics and for other subsystems. Chapter 8 provides a set of guidelines for the extension and the generalization of this approach to a wider set of problems.

The next section provides suggestions for future research.

9.3 Future Work

There are a number of ways the research of this thesis could be extended. A few suggested extensions are discussed below.

Application of the model to different classes of missions: The current methodology has been tested mostly with small spacecraft, due to the limited data available. Applying the model to spacecraft equipped with high power and high mass communication components will provide insights on the behavior of the model for different classes of spacecraft. Also, applications of the model to commercial satellites would be important since they carry demanding communication equipment. This communication equipment can drive a significant portion of the total mass and power of the spacecraft.

Extension of the model to the correlated combination of probability densities: The approach shown in the thesis assumes that the risks associated with each of the components are independent from each other. This assumption is valid in some cases, but not in all circumstances. Hence, it is important to investigate and develop approaches to compute risks for the cases in which the probability distributions of the different components are correlated.

Application of Robust Optimization methodology: Robust Optimization [19] is a recently developed concept that instead of seeking to immunize the solution in some probabilistic sense to stochastic uncertainty, it constructs a solution that is optimal for any realization of the uncertainty in a given set. This approach could be beneficial, whenever dependencies across parameters are unavoidable, and it could also allow reducing the computational complexity of modeling joint bi-dimensional distributions.

Extension of the approach to different metrics: This aspect has been briefly

touched in Chapter 8. However, lack of data limited the assessment of the results for the cost extension. Hence, collection of data to validate the claim of extendibility to cost-risk analysis is important. Furthermore, other metrics can be considered such as data rate and Signal to Noise Ratio (SNR).

Generalization of the approach to other subsystems and to the entire spacecraft: This aspect has been briefly described in Chapter 8. However, the analysis of the TerSat test case is currently in process. Further examples of statistical risk analysis for different subsystems are important to validate the claim of model generalization. Finally, the generalization to the entire spacecraft would provide the designers with a model to quantify the total risk of exceeding mass and power consumption caps. This tool could be extremely important to assess risks in spacecraft design and to provide a mathematical support to the current risk analysis approaches.

Application of the approach to failure risks: The approach discussed in this thesis has been designed to perform design risk characterization. However, the approach can be used to develop a similar model to characterize failure risks. The main challenge would be identifying all the possible causes for failures. However, when the causes of failures are identified the statistical techniques developed for this thesis could be applied to failure risk estimation.

Application of the approach to different fields: The approach presented in this thesis can be extended to different fields from spacecraft design. For example, financial analyses could benefit from the use of a statistical methodology to estimate risks.

Appendix A

Additional Results in Density Estimation

As described in Chapter 4, one of the quantities used to assess the performance of density estimation techniques is the estimation error for a certain number of samples. This quantity is defined as:

$$\overline{\Delta} = \int x \cdot \Delta(x) dx \tag{A.1}$$

where Δx is the tails' divergence across the k trials between real tail ($T(x)$) and estimated tail($\hat{T}(x)$) for each point of the distribution (x):

$$\Delta(x) = \frac{1}{k} \sum_{j=1}^{k} \left| T(x) - \hat{T}_j(x) \right| \tag{A.2}$$

The following plots (Figures A-1, A-2, A-3, A-4, A-5) show the behavior of Δ for different benchmark distributions.

The estimation error increases considerably when benchmarks are not normal. For example, looking at the exponential benchmark (Figure A-2), it can be verified that to obtain an estimation error less than 4% at least 40 samples are required, while for normal benchmarks (Figure A-1) the samples required are only 20.
Exponential distribution together with lognormal distribution are the benchmarks that achieve the highest values of estimation error. This result indicates that if the

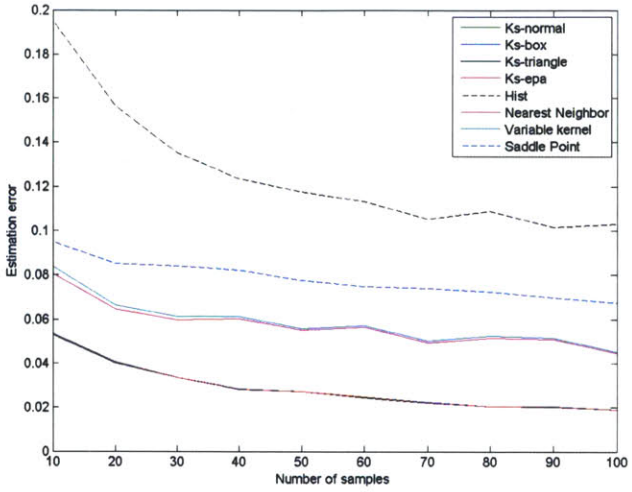

Figure A-1: Estimation error for different number of samples using normal benchmark distributions. Histogram achieves the poorest performance, and Kernel Density Estimators achieve the best performance.

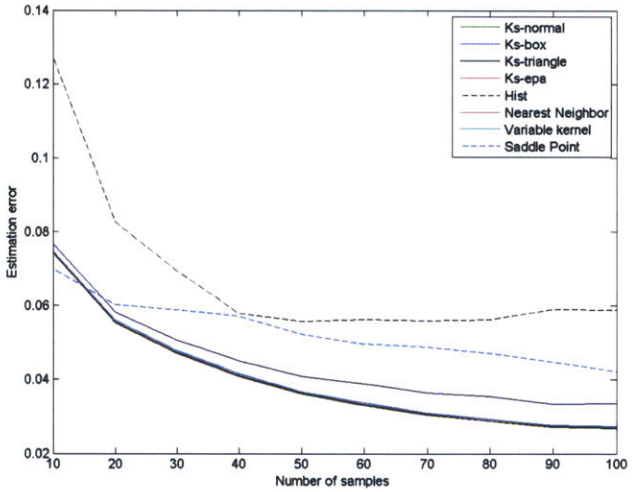

Figure A-2: Estimation error for different number of samples using exponential benchmark distributions. Histogram achieves the poorest performance, and Kernel Density Estimators achieve the best performance. The error is greater than in the case of normal benchmark.

distributions of mass and power consumption behave as exponential or as lognormal, density estimation techniques face more difficulties to predict the probability density with respect to distributions that behave like normal.

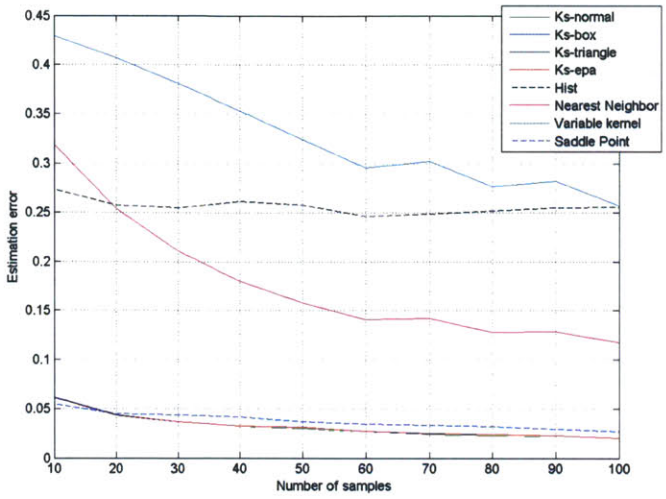

Figure A-3: Estimation error for different number of samples using beta benchmark distributions. Histogram achieves the poorest performance, and Kernel Density Estimators achieve the best performance. The error is greater than in the case of normal benchmark.

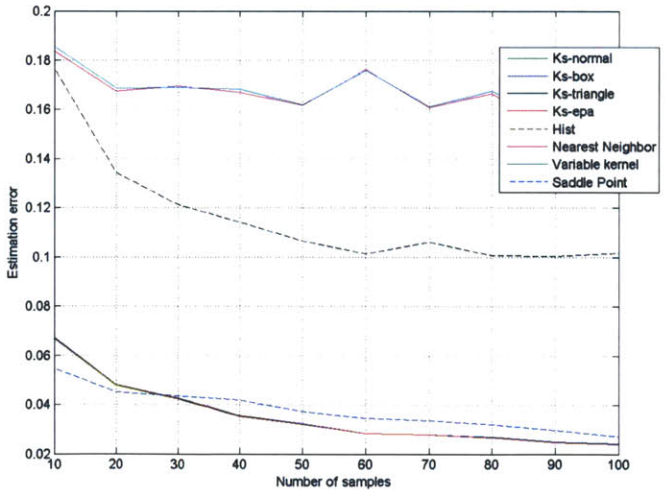

Figure A-4: Estimation error for different number of samples using gamma benchmark distributions. Histogram achieves the poorest performance, and Kernel Density Estimators achieve the best performance. The error is greater than in the case of normal benchmark.

However, Figures A-2, A-3, A-4, A-5 show that Kernel Estimators perform better than other techniques (Histogram, Saddle-Point, Variable Kernel, Nearest Neighbor)

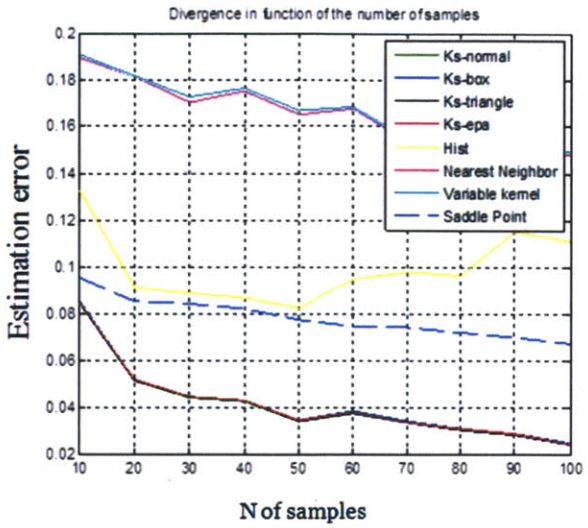

Figure A-5: Estimation error for different number of samples using lognormal benchmark distributions. Histogram achieves the poorest performance, and Kernel Density Estimators achieve the best performance. The error is greater than in the case of normal benchmark.

even if the performance decreases when benchmarks are not normal.

Appendix B

Expert Interview

The interview is divided in three parts: Probabilistic Thinking, Calibration, and Elicitation. Details for each of the parts are described in Chapter 5. In this section, all the questions are listed in their original formulation together with an explanation.

B.1 Part 1: Probabilistic Thinking

As mentioned in Chapter 5, the objective of Probabilistic Thinking is to investigate the ability of the expert to think probabilistically. The questions are as follows:

1. Given the following possible places to spend a weekend, where do you most likely think a JPL engineer will go? Give percentages to the following options (The options are exhaustive and mutually exclusive):

   (a) Hiking

   (b) Beach

   (c) Disneyland

   (d) Las Vegas

   (e) Museum of Arts

Explanation: This question is focused on coherence, and specifically on the first law of probability. The statistician wants to verify that the expert knows that the sum of probabilities for distinct events which are a partition of the sample set has to be 1 (or 100%).

2. Given that only 30% of the population is affected by disease x, and only 20% of the population is affected by disease y: what is the percentage of the population affected by both diseases?

   Explanation: This question is focused on coherence, and specifically on the second law of probability. The statistician would like to identify whether the expert knows what independent events are: the right answer is 6% (the product of 30% and 20 %).

3. Given that 50% of the Los Angeles population love snow sports (ski, snowboard), which percentage do you think loves snowboarding?

   Explanation: This question is a representativeness question. The statistician wants to check if the expert knows that a subset of an event has a probability less than or equal to the event. Hence, if 50% of the population loves snow sports, the subset that loves snowboarding is less than or equal to 50%.

4. Linda is 31 years old, single, outspoken and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations. Please tick the most likely alternative:

   (a) Linda is a bank teller.

   (b) Linda is a bank teller, and she is active in the feminist movement.

   Explanation: This question is presented in [100], and it is used to check conjunction fallacy, which is a special case of representativeness. The second answer cannot be more likely than the first one, since the second alternative is a subset of the first. However, people tend to wrongly select the second option as they are biased by Linda's description.

5. Mark is 30 years old and has a college degree in Music with a minor in Economics. From his childhood, music has been his true passion. He loves playing guitar which he started studying when he was only 4 years old. He also sings and composes his own songs. During the high school and in college, he also played basketball. Express (in percentage) your likelihood that Mark is a (The options are exhaustive and mutually exclusive):

   (a) Famous (world known) artist/song writer

   (b) Bank teller

   (c) Music teacher

   (d) Basketball coach

   Explanation: This question is used to check for representativeness and awareness of underling conditional probabilities. Being a famous artist is extremely unlikely compared to the other alternatives, even if the description matches the profile of a singer/song writer. Additionally, this question is used to check coherence (first law of probabilities), because the sum of all the percentages has to add to 100.

6. Mr. X is described as "meticulous, introverted, meek, and solemn." What is the probability that he is working in one of the following professions? (The options are exhaustive and mutually exclusive).

   (a) Salesman

   (b) Farmer

   (c) Pilot

   (d) Librarian

   (e) Physician

Explanation: This question is described in [53], and it is used to check for representativeness and awareness of the understated conditional probabilities. Salesmen and farmers are numerically more than librarians; hence, the probability of Mr X. belonging to these two categories has to be greater than the others regardless of his description. Additionally, this question is a coherence check.

7. If we choose a random English word it is more likely to:

   (a) Start with an r.

   (b) Have r as its third letter.

   Explanation: This question is presented for the first time in [53], and it is used to check judgment by availability. It is easier to recall words that start with r, so the first answer is generally the one that is selected while the correct answer is the second.

8. Suppose you have to run an experiment on twenty subjects and have obtained a significant result which confirms your theory ($z = 2.23$, $p < 0.05$, two tailed). You now have cause to run an additional group of ten subjects. What do you think the probability is that the results will be significant, by one-tailed test, separately for this group?

   Explanation: This question is used in [53], and it is used to check small sample size bias. Generally, the median of the answers to this question in previous studies is 0.85. However, the real probability computed using Bayesian analysis and a non-informative prior gives 0.48. Generally, people think that all samples behave the same, even if they are a very small set.

9. For a person in 1988, what was the probability that the Berlin wall would come down within the next five years?

   Explanation: This question is formulated in [53]. It is used to check hindsight bias. In 1988 it seemed extremely unlikely that the wall would come down. However, the fact that the expert knows what happened bias his/her answer.

10. The population of a city has 70% of engineer, 30% of lawyers. Andy is 30 year old, married with no children. Andy is a person of high ability and high motivation, quite successful, well liked by colleagues. Express the probability of Andy being:

    (a) Engineer

    (b) Lawyer

    Explanation: This question is tested for the first time in [53], and it is used to check for representativeness. The probabilities are 70% and 30%, but sometime people put 50% because the description does not recall any of the engineer or lawyer stereotypes.

11. A certain town is served by two hospitals. In the larger hospital about 45 babies are born each day, and in the smaller hospital about 15 babies are born each day. As you know, about 50 percent of all babies are boys. However, the exact percentage varies from day to day. Sometimes it may be higher than 50 percent, sometimes it may be lower. For a period of 1 year, each hospital recorded the days on which more than 60 percent of the babies born were boys. Which hospital do you think recorded more of such days?

    (a) The larger one

    (b) The smaller one

    (c) About the same (in a 5 percent with each other)

    Explanation: This question is presented in [127], and it is used to check small sample bias. The correct answer is the second, but generally people tend to select the third one.

12. Imagine an urn filled with balls, 2/3 of one color, 1/3 of another. One individual has drawn 5 balls from the urn, and found that 4 were red and 1 was white.

Another individual has drawn 20 balls and found that 12 were red and 8 were white. Which of the two should be more confident of having 2/3 red and 1/3 white?

(a) First individual

(b) Second individual

Explanation: This question is used for the first time in [127], and it is used to check for small sample bias. People tend to select the first one, but the correct answer is the second.

13. The experts will listen to a list of names: Donnie Hathaway, Madonna, Rob Eaton, Shakira, Luca Jurman, Darren Chris, Mariah Carey, Alicia Keys, Brian Mcknight, Bobby V., Whitney Houston, Elton John, Alan Robert, Barbra Streisand, Eminem, Rihanna, Matt Morris, Jon Maddox, Lady Gaga, Katy Perry.
    At the end of the list, the expert will be invited to say if the list was composed of:

    (a) More male names than female names.

    (b) More female names than male names.

    Explanation: This question is used to check small sample bias, representativeness, and availability. When female names are more famous, people are inclined to indicate the first answer even if the number of male names is greater.

14. The mean IQ of the population of eighth graders in a city is known to be 100. You have a random sample of 50 children to test. The first one tested has an IQ of 150. What do you expect for the whole sample?
    Explanation: This question is presented in [127], and it is used to check for anchoring. The correct answer is 101. People tend to believe that the knowledge of the first eighth grader tested does not change the result, and they tend to assess a mean IQ for the whole sample of 100.

15. The following question is divided in two parts.

Part a: Suppose that a doctoral student has completed a difficult and time-consuming experiment on 40 animals. He has scored and analyzed a large number of variables. His results are generally inconclusive, but one before-after comparison yields a highly significant $t = 2.70$, which is surprising and could be of major theoretical significance. Considering the importance of the result, its surprising value, and the number of analyses that your student has performed- Would you recommend that he replicate the study before publishing? If you recommend replication, how many animals would you urge him to run ?

Explanation: This question is presented in [127], and it is used to check for small sample bias. Generally people suggest replicating the experiment, and they indicate at least 20 samples. However, if mean and variance in the second set of samples are identical then the final $t = 1.88$. Hence, the student chance of obtaining the same results would be small ($p < 0.05$).

Part b (indicated as question number 16 in the thesis): Assume that the unhappy student has in fact repeated the initial study with 20 additional animals, and has obtained an insignificant result in the same direction, $t = 1.24$. What would you recommend now?

(a) He should pool the results and publish his conclusion as fact.

(b) He should report his result as a tentative finding.

(c) He should run another group of animals.

(d) He should try to find an explanation for the difference.

Explanation: There is not a right answer: the best answers are b and c. a is not a good answer because after repeating the experiment the initial conclusion are not true anymore. d is also not a good answer because the explanation for the difference is not a mystery, but it is already available and it is represented by the small sample bias.

B.2 Part 2: Calibration

The objective of this part is to identify whether experts tend to be overconfident or underconfident in their estimations. The following questions are referred to point design values, for which real values of mass and power consumptions are known.

1. A monolithic parabolic antenna in X-Band has a gain of 62 dB. What is the expected mass value?
   Explanation: The mass value is 53 Kg, and the antenna is used in the Voyager mission [124].

2. A deployable parabolic antenna in S-Band has a gain of 32 dB. What is the expected mass value?
   Explanation: The mass value is 8 Kg, and the antenna was used for the Galileo mission [120].

3. An array of patch antennas in X-Band has a gain of 39.6 dB. What is the expected mass value?
   Explanation: The mass value is 7.7 Kg, and the antenna belongs to the Dawn mission [119].

4. A monolithic array of patches in X-Band has a gain of 17 dB. What is the expected mass value?
   Explanation: The mass value is 3.5 Kg, and the antenna was used in the Phoenix mission.

5. An LGA antenna of 8 dB is used in X-Band. What is the expected mass value?
   Explanation: The mass value is 0.8 Kg, and the antenna belongs to the Mars Reconnaissance Orbiter [73].

6. A small 3 dB dipole antenna is used in the X-Band. What is the expected mass value?
   Explanation: The mass value is 0.042 Kg, and the antenna is used in the Odyssey mission [85].

7. An S-Band parabolic and monolithic antenna is used for a space mission. It has a gain of 30 dB. What is the expected mass value?
   Explanation: The mass value is 16 Kg, and the antenna is a COTS product.

8. An S-Band parabolic and monolithic antenna is used for a space mission and has a gain of 15 dB. What is the expected mass value?
   Explanation: The mass value is 1.4 Kg, and it is a COTS product.

9. A patch antenna in S-Band has a gain of 6 dB. What is the expected mass value?
   Explanation: The mass value is 0.1 Kg, and it is an MIT SSL developed antenna [1].

10. A transceiver is used in Ku-Band to reach a transmitting power of 10 W. What are the expected mass and power consumption values?
    Explanation: The mass value is 5.4 Kg, and the power consumption is 114 W (COTS product used in space applications).

11. A transceiver is used in the S-Band to reach a transmitting power of 10 W. What are the expected mass and power consumption values?
    Explanation: The mass value is 2.26 Kg, and the power consumption is 37 W (COTS model used in space applications).

12. A transceiver is used in the S-Band to reach a transmitting power of 1 W. What are the expected mass and power consumption values?
    Explanation: The mass value is 0.05 Kg, and the power consumption is 4.5 W (COTS model used in space applications).

13. A TWTA is used for a space mission in X-Band with an output transmitting power of 102 W. What are the expected mass and power consumption values?
    Explanation: The mass value is 1.9 Kg, and the power consumption is 172 W (Data from the Mars Reconnaissance Orbiter [73]).

14. A TWTA is used for a space mission in X-Band with an output transmitting power of 20 W. What are the expected mass and power consumption values?
    Explanation: The mass value is 5.4 Kg, and the power consumption is 53 W (Data from the CASSINI-HUYGENS mission [125]).

15. A switch for a space mission is developed for X-Band. What is the expected mass value?
    Explanation: The mass value is 0.375 Kg (Data from the CASSINI-HUYGENS mission [125]).

16. A diplexer for a space mission is developed for X-Band. What is the expected mass value?
    Explanation: The mass value is 0.438 Kg (Data from the Mars Exploration Rover [123]).

17. Given a mission for deep space exploration, with a complex communication system composed of an S-Band and an X-Band channel (including 2 antennas, 2 amplifiers, 2 transceivers, and a diplexer), what is the expected mass for the cables?
    Explanation: The mass value is 2.3 Kg (Data from the CASSINI-HUYGENS mission [125]).

B.3 Part 3: Elicitation

In this part the expert is invited to perform the elicitation. Two test cases are considered:

1. CASTOR
2. HETE-2

The following subsections describe the questions for the different test cases. The different questions help the experts to assess the summaries described in Chapter

5. Motivations for the formulation of these questions, and for the selection of these specific summaries are discussed in Chapter 5.

B.3.1 CASTOR

CASTOR test case is introduced as follows:

> CASTOR is a small ESPA class satellite for LEO orbit. To perform its mission, it needs to be equipped with a communication system in the S-Band with 6 dBW EIRP. The selected components are a 6 dB patch antenna and a transceiver.

After the introduction, the questions are the following:

1. The antenna is a 6 dB patch with an expected mass value of 0.5 Kg. What do you think are the lower and upper bounds for this mass value? Which shape of the probability distribution would you choose? What are the 50% and the 16% quantiles?

2. The transceiver has 1 W of output power, an expected mass value of 0.1 Kg, and an expected power consumption of 4.5 W. What do you think are the lower and upper bounds for mass and power consumption? Which shape for the probability distributions would you choose? What are the 50% and the 16% quantiles for each distribution?

B.3.2 HETE-2

HETE-2 test case is introduced as follows:

> The High Energy Transient Explorer (HETE-2) is a small scientific satellite designed to detect and to localize gamma-ray bursts. The communication system includes an S-Band channel, a VHF-Band channel and

a GPS system. The following questions ask you to express opinions on some components of this system.

The questions are the following:

1. GPS channel: the GPS system is composed of an antenna and a transceiver for a total EIRP of 13.5 dBW. The following questions refer exclusively to GPS channel.

   (a) The GPS antenna is a 13.5 dB patch with an expected mass value of 0.2 Kg. What do you think are the lower and upper bounds for this mass? Which shape of the probability distribution would you choose to describe the mass distribution? What are the 50% and the 16% quantiles of the mass probability distribution?

   (b) The GPS transceiver has 1 W of output power, an expected mass value of 1 Kg, and an expected power consumption of 10 W. What do you think are the lower and upper bounds for mass value and power consumption? Which shape for the probability distributions would you choose? What are the 50% and the 16% quantiles for each of the two distributions?

2. S-Band channel: the S-Band channel is composed of 5 patch antennas, a switch and a transceiver for a total EIRP of 8 dBW. The following questions refer to that channel.

   (a) Each of the 5 patch antennas provides a gain of 5.5 dB. The expected value for mass including the supports is 0.2 Kg for each antenna. What do you think are the lower and upper bounds for this mass? Which shape of the probability distribution would you choose to describe the mass distribution? What are the 50% and the 16% quantiles of the mass probability distribution?

(b) The S-Band transceiver has 2 W of output power, an expected mass value of 1.5 Kg, and an expected power consumption of 30 W. What do you think are the lower and upper bounds for mass and power consumption? Which shape for the probability distributions would you choose? What are the 50% and the 16% quantiles for each of the two distributions?

B.4 Consent Form

In order to perform the interview, MIT requires satisfying the requirements of the COUHES (Committee on the Use of Humans as Experimental Subjects). One of the requirements is that each subject has to sign a consent form. A sample of the form (the one used for engineers) is reported here.

CONSENT TO PARTICIPATE IN INTERVIEW: form used for Professors/Engineers

Study title: Elicitation of expert opinion: assessment of biases, mis-calibrations, and probabilities elicitation. (Part of the PhD Thesis: Statistical Risk Estimation in Communication System Design).

You have been asked to participate in a research study conducted by Alessandra Babuscia from the Department of Aeronautics and Astronautics at the Massachusetts Institute of Technology (M.I.T.). The purpose of the study is to assess how people think about probabilities: what are their biases, heuristics, mis-calibrations. This information will help to improve the acquisition of experts' opinions in statistical risk assessments problems. The results of this study will be included in Alessandra Babuscia PhD thesis. You were selected as a possible participant in this study for your expertise in communication system design and modelling. You should read the information below, and ask questions about anything you do not understand, before

deciding whether or not to participate.

1. This interview is voluntary. You have the right not to answer any question, and to stop the interview at any time or for any reason. We expect that the interview will take about 90 minutes.

2. You will not be compensated for this interview.

3. Unless you give us permission to use your name, title, and/or quote you in any publications that may result from this research, the information you tell us will be confidential.

This project will be completed by December 2011.

I understand the procedures described above. My questions have been answered to my satisfaction, and I agree to participate in this study. I have been given a copy of this form.

(Please check all that apply)

[] I give permission for the following information to be included in publications resulting from this study:

[] my name [] my title [] direct quotes from this interview

Name of Subject

Signature of Subject
Date

Signature of Investigator

Date

Please contact Alessandra Babuscia (babuscia@mit.edu) with any questions or concerns.

If you feel you have been treated unfairly, or you have questions regarding your rights as a research subject, you may contact the Chairman of the Committee on the Use of Humans as Experimental Subjects, M.I.T., Room E25-143b, 77 Massachusetts Ave, Cambridge, MA 02139.

B.5 Summary

The appendix details the interview performed with the experts. Questions have been explained and motivated. The consent form has been described. The results of this interview are discussed in Chapter 5.

Appendix C

An Upper Bound Risk Calculation for Symmetric Probability Distributions

The objective of this research is to develop a statistical model to estimate the risk of exceeding a certain cap in mass and power. It is interesting to investigate what is the gain of the statistical approach, which means how much less contingency must be applied if we are willing to accept a certain risk threshold. The assumptions are the following:

- A certain component has a design value of $\bar{x}$;
- The traditional approach applies contingency to the design value (Δ).
- The final value of the mass of the component according to the traditional approach is as follows:

$$M_{contingency} = \bar{x} + \Delta \tag{C.1}$$

To compare the traditional approach with the statistical risk estimation, it is necessary to quantify the likelihood that the component's mass will fluctuate from an average of $\bar{x}$. The parameters to estimate the likelihood (probability distribution) are as follows:

- Design value: It corresponds to the average value $\bar{x}$;

- Adverse and favorable value: Given that no additional information is available, a favorable value given by the design value itself, and an adverse value given by the average plus the contingency are assumed.

- Shape of the distribution: Since no other information is available, the uniform distribution is assumed (which is the distribution that presents the maximum variance across the symmetric and bounded distributions). Moreover, the uniform distribution is limited between the average mass $\bar{x}$ and the mass with the contingency $(\bar{x} + \Delta)$. This case is a pessimistic case since the model assumes that the component's mass can only increase and not decrease. Hence, the mass fluctuation is statistically modelled as follows:

$$f(x) = U(\bar{x}, \bar{x} + \Delta) \tag{C.2}$$

$$f(x) = \begin{cases} \frac{1}{\Delta} & \text{if } \bar{x} < x < \bar{x} + \Delta \\ 0 & \text{otherwise} \end{cases} \tag{C.3}$$

The tail function of the distribution $(T(x))$ results:

$$T(x) = \begin{cases} 1 & \text{if } x < \bar{x} \\ 1 - \frac{x - \bar{x}}{\Delta} & \text{if } \bar{x} < x < \bar{x} + \Delta \\ 0 & \text{if } x > \bar{x} + \Delta \end{cases} \tag{C.4}$$

It is possible to quantify how the contingency can be reduced for a given threshold of *acceptable* risk. To do so, a percentage of acceptable risk is defined as α. The mass allocated for a given threshold of risk is calculated as:

$$T(x) = \alpha \tag{C.5}$$

$$M_{stat} = T^{-1}(\alpha) \tag{C.6}$$

The risk α is a quantity defined between 0 and 1, but in this case it is considered between 0 and 0.5, since a risk greater than 50% does not have any utility for the

analysis (designers will never select a solution with a risk greater than 50%). The value of mass for a given threshold of risk is as follows:

$$M_{stat} = \bar{x} + \Delta \cdot (1 - \alpha) \tag{C.7}$$

Comparing this expression with Equation C.1, the statistical approach presents an advantage with respect to the traditional approach, if the designer is willing to afford a minimal risk. Specifically, the mass saving in mass allocation by using the statistical approach ($G = M_{contingency} - M_{stat}$) is:

$$G = \Delta \cdot \alpha \tag{C.8}$$

The saving G is proportional to the contingency and to the percentage of risk allowed. Figures C-1 and C-2 show a graphical comparison between contingency and statistical upper bound for $\bar{x} = 10$ and a $\Delta = 30(30\%)$.

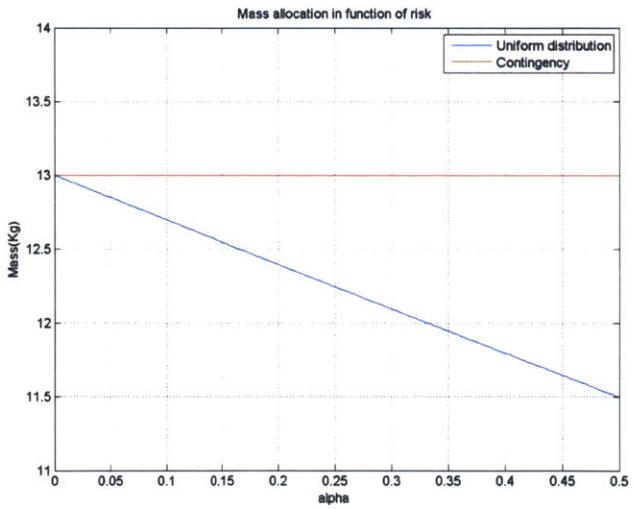

Figure C-1: Comparison of traditional approach and statistical approach in the mass allocation. A tolerable risk allows the designer to reduce the allocated contingency.

The plot compares the mass allocation using the traditional approach, and the mass allocation using a statistical approach. The statistical method allows savings in

the mass allocation, which is plotted in Figure C-2, for the same example.

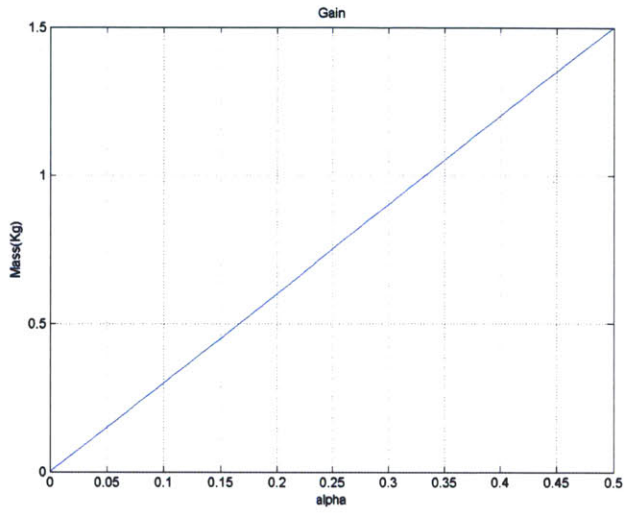

Figure C-2: Mass savings for the statistical approach. The mass allocation is reduced whenever a minimum design risk α is allowed.

The mass saving is computed and plotted assuming the uniform distribution. This is an upper bound on statistical risk estimation (for $\alpha \leq 0.5$) with respect to specific distributions. To formally prove the upper bound, Theorem 1 needs to be introduced:

> Theorem 1: If $F(x)$ has a second derivative in $[a, b]$, then a necessary and sufficient condition for it to be concave on that interval is that the second derivative $F''(x) = f'(x)$ is negative.

Another way of stating Theorem 1 is that a differentiable function $F(x)$ is concave on an interval if its derivative $F'(x) = f(x)$ is monotonically decreasing on that interval. In other words, a concave function has a decreasing slope. Using Theorem 1, Theorem 2 can be proved:

> Theorem 2: A uniform distribution ($U[\bar{x}, \bar{x} + \Delta]$) is an upper bound in statistical risk estimation for risk thresholds α less than or equal to 0.5, and with respect to any other probability distribution $f_g(x)$ that is:
>
> 1. Continuous and differentiable.

2. Unimodal and symmetric in the interval $[\bar{x}, \bar{x} + \Delta]$.

3. Bounded in $[\bar{x}, \bar{x} + \Delta]$.

Proof: Let $f_u(x)$ be a uniform PDF in the interval $[\bar{x}, \bar{x} + \Delta]$. If $f_g(x)$ is a unimodal and symmetric PDF in the interval $[\bar{x}, \bar{x} + \Delta]$, then $f_g(x)$ is peak at $\bar{x} + \frac{\Delta}{2}$, and $f_g(x_1) \geq f_g(x_2)$ for $\bar{x} + \frac{\Delta}{2} \leq x_1 \leq x_2 \leq \bar{x} + \Delta$ (i.e. $f_g'(x)$ is negative). From the symmetric property of $f_g(x)$, $F_g(x)$ and $F_u(x)$ intersect at 2 points:

$$F_g\left(\bar{x} + \frac{\Delta}{2}\right) = F_u\left(\bar{x} + \frac{\Delta}{2}\right) = 0.5 \tag{C.9}$$

$$F_g(\bar{x} + \Delta) = F_u(\bar{x} + \Delta) = 1 \tag{C.10}$$

From the unimodal property of $f_g(x)$ and Theorem 1, $F_g(x)$ is concave in the interval $\left[\bar{x} + \frac{\Delta}{2}, \bar{x} + \Delta\right]$. In other words, $F_g(x) \geq F_u(x)$ for $\frac{\Delta}{2} \leq x \leq \Delta$, and graphically $F_g(x)$ and $F_u(x)$ are related as shown in Figure C-3. Thus the contingency computed from the uniform distribution $f_u(x)$ is always greater than the contingency computed from other unimodal and symmetric distribution $f_g(x)$ defined over the same interval $[\bar{x}, \bar{x} + \Delta]$ for a given risk level α, where $0 \leq \alpha \leq 0.5$.

Figure C-3 shows the previous result by comparing the uniform bound with two distributions (a triangular and a normal truncated in $[\bar{x}, \bar{x} + \Delta]$ at the fourth sigma interval) that respect the conditions stated in the theorem. It can be noticed that the uniform distribution is an upper bound across the symmetric and bounded statistics.

Non-parametric distributions do not generally reflect the conditions stated in the theorem. Hence the results for non-parametric distributions can be different, as observed in Chapter 7 when simulated statistics and real statistics (derived from database and expert elicitation) are compared.

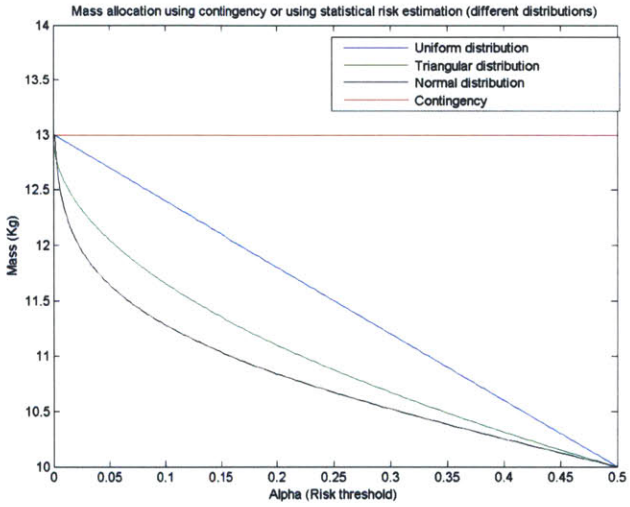

Figure C-3: Mass allocation for a given risk threshold α, and for different statistics. The uniform distribution generates an upper bound in risk estimation across the other distributions.

Appendix D

CASSINI-HUYGENS Test Case

Appendix D describes the risk analysis performed with CASSINI-HUYGENS test case.

D.1 Mission Description

CASSINI-HUYGENS is a joint NASA/ESA/ASI spacecraft mission studying the planet Saturn and its satellites since 2004. Launched in 1997, after two decades of design and development, it includes a Saturn orbiter and an atmospheric probe/lander for the moon Titan. The Titan probe, HUYGENS, entered and landed on Titan in 2005. The current end of mission plan is a 2017 Saturn impact.

D.2 Communication System

CASSINI-HUYGENS has been selected as a test case for the model since it is a very different spacecraft from CASTOR and HETE-2, carrying high energy communication equipment, and a high gain antenna. The communication system is described in [125] and it is composed of two channels: a Ka-Band and X-Band channel. The risk analysis is focused on the X-Band channel, which includes a high gain antenna, two low gain antennas, two transceivers, and two TWTA. Table D.1 summarizes mass and power

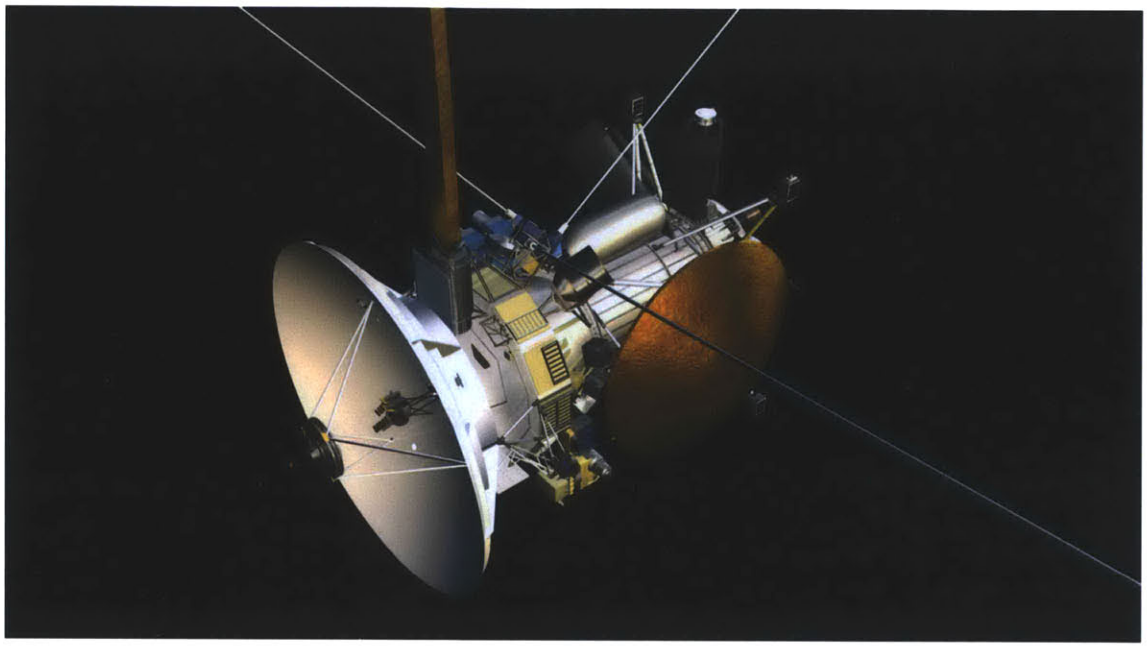

Figure D-1: CASSINI-HUYGENS spacecraft overview. Source: [125].

consumption for the system at the final stage of the design (CDR). The EIRP required for downlink transmission is 58 dB. Using this mission as a test case represents a

Table D.1: Components for CASSINI-HUYGENS communication system (X-Band channel), data from Descanso publications [125].

Component	Mass (Kg)	Power consumption (W)	Quantity
HGA	100.6	N/A	1
LGA	0.25	N/A	2
Transponder	4.5	10	2
TWTA	6	53	2
Total	115.9	126	

challenge. The motivations are: the unavailability of initial (PDR) data, and the unavailability of expert information. For these reasons, risk analysis on CASSINI-HUYGENS is performed using only data statistics and the results are excluded from the main thesis document.

To solve the issue of the unavailability of PDR data, the baseline model has been used to generate an initial PDR estimate. The baseline model is described in Chapter 6.

It creates multiple possible PDR designs able to accomplish a certain quality goal (defined as the possibility of transmitting all data to the Earth at a certain data rate). Across the multiple designs generated by baseline model, the one which seems to be closer to the CDR estimate has been selected. The PDR design estimated through baseline model is summarized in Table D.2.

Table D.2: Estimation of PDR design for CASSINI-HUYGENS communication system (X-Band channel) using data from baseline design.

Component	Mass (Kg)	Power consumption (W)	Quantity
HGA	107	N/A	1
LGA	0.3	N/A	2
Transponder	4.7	5	2
TWTA	4.8	48	2
Total	126	116	

D.3 Risk Analysis

Risk analysis is performed for each of the 4 X-Band components (HGA, LGA, transceiver, TWTA). The analysis for the HGA is performed through the following steps:

1. Component type and category: high gain antenna.

2. Performance metric: in this case the performance metric is the antenna gain (47.1 dB), used to generate the proper statistic based on historical data as described in Chapter 4.

3. Computation of probability density function and tail function.

Results are shown in Figures D-2 and D-3. The peak of the distribution is far from both PDR and the CDR values. This result can be explained by the following considerations:

- The PDR value used in the analysis is an estimate, and it is not the real value. The PDR value has been generated using baseline design. It is true that baseline

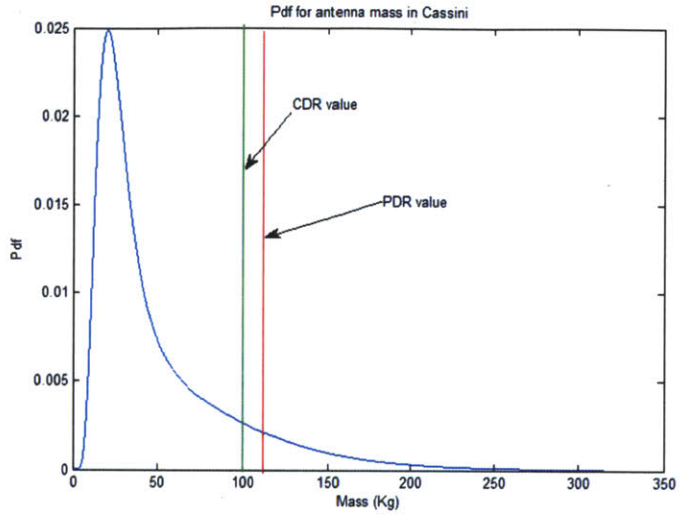

Figure D-2: Probability density function for CASSINI-HUYGENS HGA mass. The peak of the distribution is very far from both the PDR and the CDR estimates.

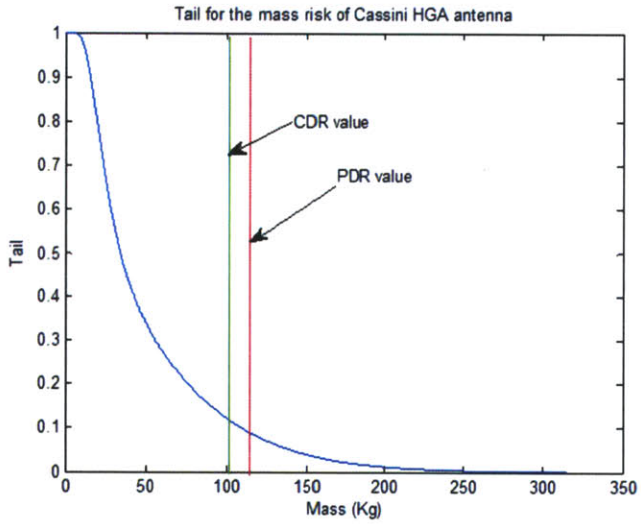

Figure D-3: Tail function for CASSINI-HUYGENS HGA mass.

design is accurate (see validation in Chapter 6). However, the author cannot be completely sure that this PDR value corresponds to the real one. Hence, this problem may be one of the reason for which the peak of the distribution and the PDR value lie far apart.

- Technological evolution of antenna design: the mass of antenna depends strongly

on the material of the reflector. In fact, CASSINI-HUYGENS HGA reflector was built using aluminum honeycomb. Currently, most reflectors for high gain antennas are built using composite honeycomb which is lighter. As a consequence, the database collects mostly new antennas which are generally lighter than the old reflector, generating the shift between the peak of the distribution and the values of PDR and CDR.

This effect is clearly observable for high gain antennas, but it can be noticed also for low gain antennas. The analysis for CASSINI-HUYGENS LGA has been performed through the same steps described for the HGA. In this case the antenna is an LGA in the X-Band with a gain of 8.4 dB. The results are shown in Figures D-4 and D-5 .

Also in this case, the statistical model tends to underestimate the mass of the

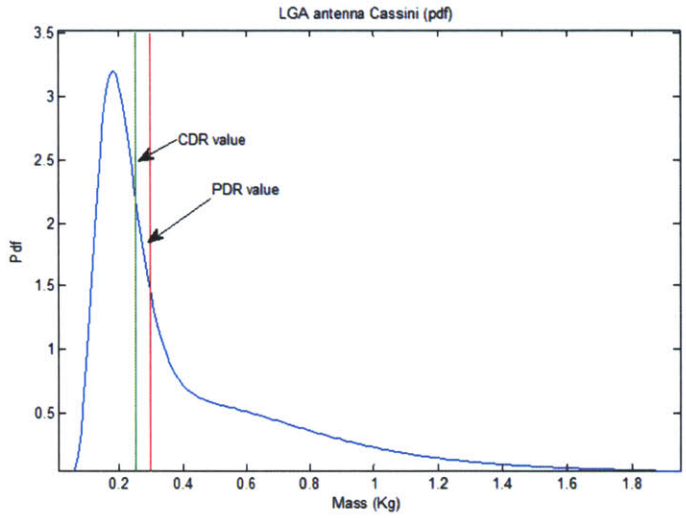

Figure D-4: Probability density function for CASSINI-HUYGENS LGA antenna mass. The model underestimates the mass of the antenna.

component. However, the phenomenon is very moderate compared to the case of the HGA. This result indicates that the risk model developed tends to be more accurate with small components, like the ones used for the other test cases (CASTOR and HETE-2).

For CASSINI-HUYGENS transceiver, the results show underestimation in the mass, but a good estimation for the power consumption. This result seems to indicate

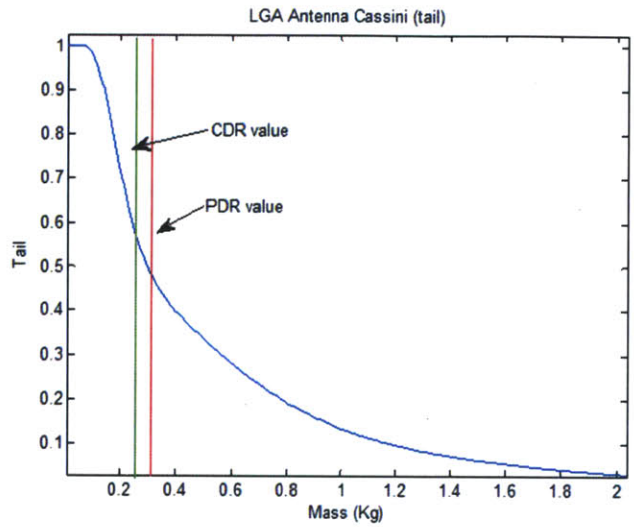

Figure D-5: Tail function for CASSINI-HUYGENS LGA antenna mass.

that the model is more accurate in estimating fluctuations in power consumption than fluctuations in mass. This difference can be observed from Figures D-6 and D-7, which show the probability density function and the tail function for mass fluctuation, and in Figures D-8 and D-9 that show probability density function and tail function for power consumption fluctuations.

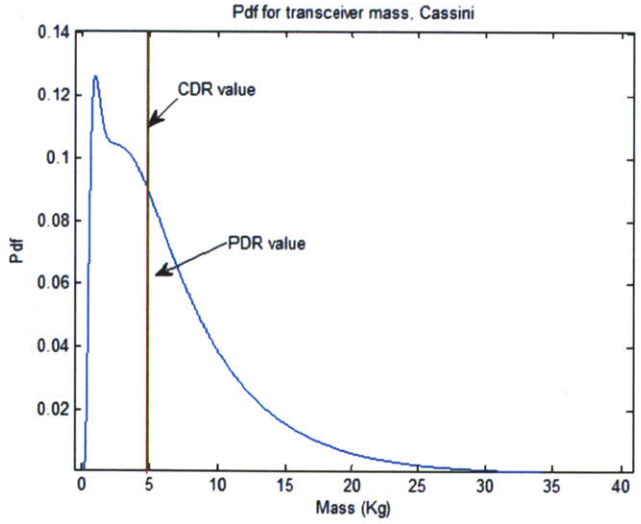

Figure D-6: Probability density function for CASSINI-HUYGENS transceiver mass. The peak of the distribution is far from PDR and CDR values.

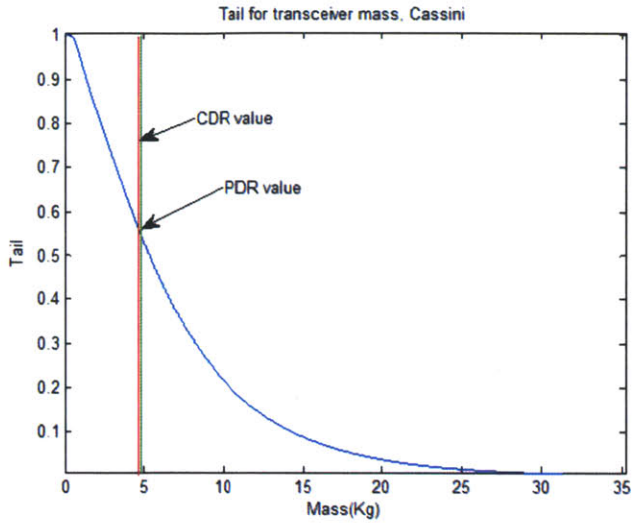

Figure D-7: Tail function for CASSINI-HUYGENS transceiver mass.

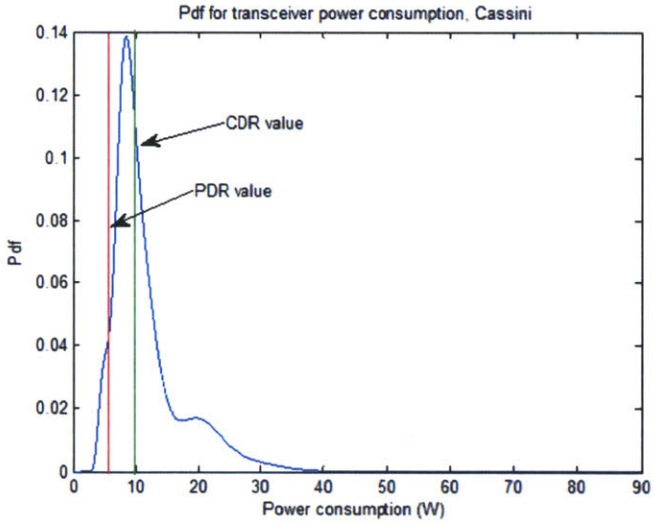

Figure D-8: Probability density function for CASSINI-HUYGENS transceiver power consumption. The peak of the distribution is centered on the final CDR estimate.

It is possible to notice that for big communication systems developed in the past the model is inaccurate in assessing the probability density function for mass fluctuation, while it is more accurate for fluctuations in power consumption. The same result can be observed with CASSINI-HUYGENS TWTA (20 dB gain). The analysis has been performed using database statistics and following the same steps described for

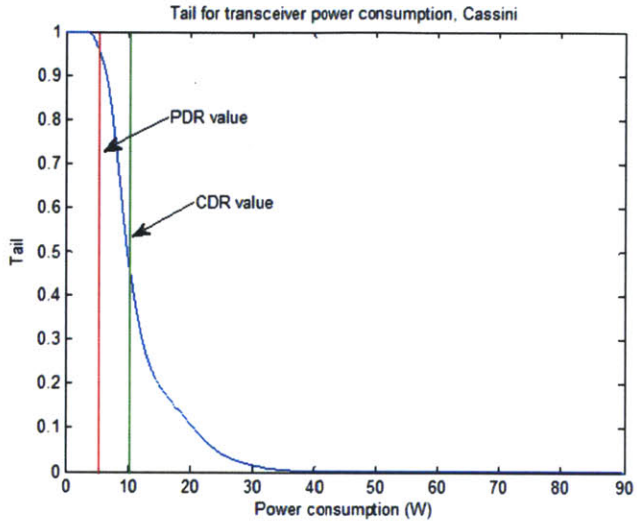

Figure D-9: Tail function for CASSINI-HUYGENS transceiver power consumption.

the other components. Results (shown in Figures D-10 and D-11 for mass fluctuation, and in Figures D-12 and D-13 for power consumption fluctuations) are similar to the transceiver case. The model estimates well the power consumption, while the mass is generally underestimated.

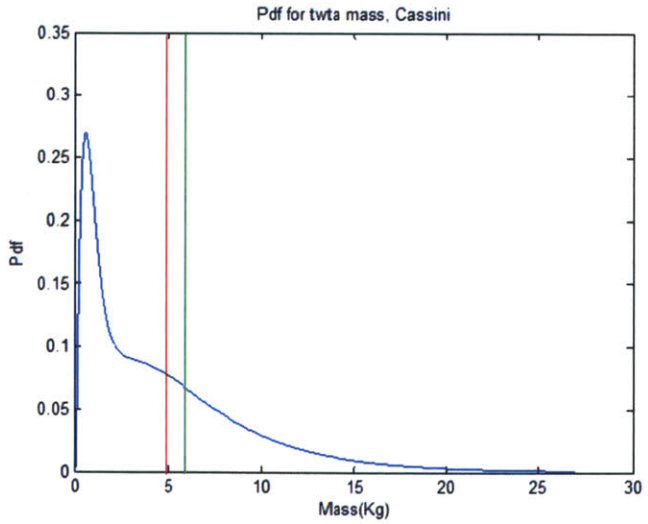

Figure D-10: Probability density function for CASSINI-HUYGENS TWTA mass. The peak of the distribution is far from PDR and CDR values.

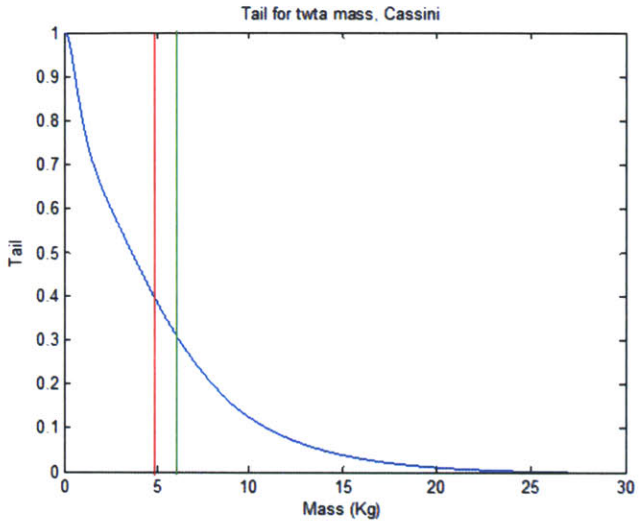

Figure D-11: Tail function for CASSINI-HUYGENS TWTA mass.

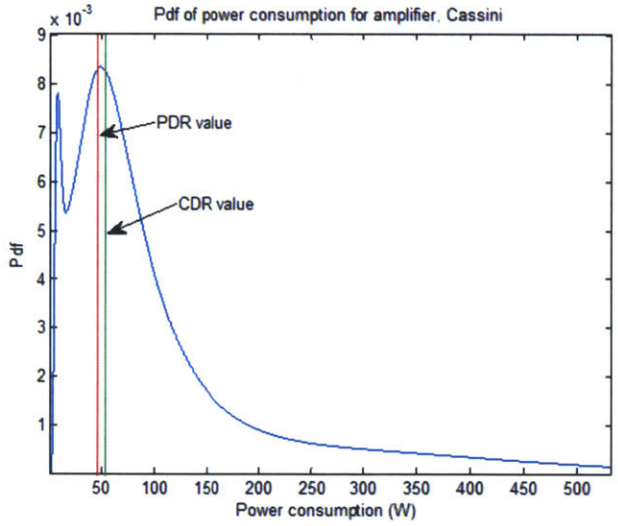

Figure D-12: Probability density function for CASSINI-HUYGENS TWTA power consumption. The distribution is multi-modal and the second peak is very close to PDR and CDR values.

D.4 Summary

A risk analysis has been performed with CASSINI-HUYGENS mission. The objective was to perform risk analysis for each component of the system, comparing the probability distributions obtained with the PDR and CDR values. In this case, with

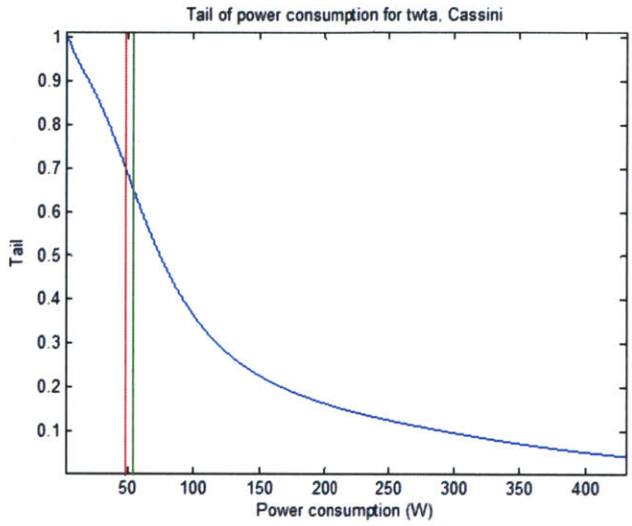

Figure D-13: Tail function for CASSINI-HUYGENS TWTA power consumption.

respect to the test cases described in Chapter 6, initial PDR point estimation values were not available, and they have been generated through baseline design.

The results of the statistical risk analysis shows a tendency of the data statistics of underestimating components' mass, while power consumption estimations are generally more accurate.

Bibliography

[1] 16.831-Class. Castor (cathode anode satellite thruster for orbital reconfiguration) design document. Technical report, Massachusetts Institute of Technology, 2010.

[2] 16.831-Class. Exoplanet satellite design document. Technical report, Massachusetts Institute of Technology, 2010.

[3] Vasil N. Alexandrov and Syed M. Hussain. Multidisciplinary design optimization: State of the art. In *Proceedings in Applied Mathematics Series*, 1997.

[4] George E. Apostolakis. The concept of probability in safety assessments of technological systems. *Science*, 250:1359–1364, 1990.

[5] Y. Asnar, V. Bryl, and P. Giorgini. Using risk analysis to evaluate design alternatives. *Agent Oriented Software Engineering VII*, 4405:140–155, 2007.

[6] Alessandra Babuscia. Statistical risk estimation for communication system design. *Technical Report, NASA Jet Propulsion Laboratory*, 2010.

[7] Alessandra Babuscia, Bruno Alvisio, Frank Schmidt, Timur Balbekov, Elizabeth Qian, Leonard Tampkins, Mark Van de Loo, and Kerry Cahoy. Mit tersat satellite: Development of a flexible, reliable and modular avionics and communication architecture for small spacecraft. *International Astronautical Conference, Naples (to be presented)*, 2012.

[8] Alessandra Babuscia and Kar-Ming Cheung. Statistical risk estimation for communication system design: a preliminary look. *The Interplanetary Network Progress Report*, 42-188:1–33, 2012.

[9] Alessandra Babuscia, Jon Elliot, Alkalai Leon, and David W. Miller. Multiobjective optimization communication methodology applied to lunar robotic exploration. *Proceedings of 61st International Astronautical Conference*, Prague, Czech Republic, September 2010.

[10] Alessandra Babuscia, Matthew McCormack, Michael Munoz, and Spencer Parra. Mit castor satellite: design, implementation and testing of the communication system. *Proceedings of 62nd International Astronautical Conference*, Cape Town , South Africa, October 2011.

[11] Osman Balci. Validation, verification, and testing techniques throughout the life cycle of a simulation study. *Annals of Operations Research*, 53:121–173, 1994.

[12] Ole E. Barndorff-Nielsen and David R. Cox. Edgeworth and saddle-point approximation with statistical applications. *Journal of Royal Statistical Society*, 41:279–312, 1979.

[13] David J. Barnhart, Tanya Vladimirova, and Martin N. Sweeting. Very small satellite design for distributed space missions. *Journal of Spacecraft and Rockets*, 44:1294–1306, 2007.

[14] Francesca Barrientos, Irem Tumer, and David G. Ullman. Design teams, complex systems and uncertainty. *ASME/IEEE*, 2010.

[15] David A. Bearden. A complexity-based risk assessment for low-cost planetary missions: when is a mission too fast and too cheap. *Acta Astronautica*, 52:371–379, 2003.

[16] James O. Berger. *Statistical Decision Theory and Bayesian Analysis.* Springer, 1985.

[17] Jose' M. Bernardo and Adrian F. M. Smith. *Bayesian Statistics.* Wiley, 2001.

[18] Dimitry P. Bertsekas and John N. Tsitsiklis. *Introduction to Probability, Second Edition.* Athena Scientific, 2002.

[19] Dimitry Bertsimas, Omid Nohadani, and Kwong Meng Teo. Nonconvex robust optimization for problems with constraints. *INFORMS Journal on Computing*, 22:44–58, 2009.

[20] Robert C. Blattberg and Stephen J. Hoch. Database models and managerial intuition: 50% model +50% manager. *Management Science*, 36:887–899, 1990.

[21] Robert S. Bokulic and Robert J. Jensen. Telecommunications system technologies for the near spacecraft. *Acta Astronautica*, 39:171–180, 1997.

[22] William Bolstad. *Introduction to Bayesian Statistics.* Wiley & Sons, 2004.

[23] Leo Breiman, William Meisel, and Edward Purcel. Variable kernel estimates of multivariate densities. *Technometrics*, 19:135–144, 1977.

[24] Charles D. Brown. *Elements of Spacecraft Design.* AIAA Books, 2002.

[25] Daylian M. Cain and Allan S. Detsky. Everyone's a little bit biased (even physician). *JAMA*, 299:2893–2895, 2008.

[26] Richard Carlson. In defense of a constitutional theory of experts. *Scholarly works*, 1:453–460, 1997.

[27] Kar-Ming Cheung. Statistical link analysis - a risk analysis prospective. *White Paper*, 2010.

[28] Kar-Ming Cheung. Probabilistic instrument data generation analysis using a variant of saddle-point approximation. In *International Planetary Probe Workshop*, Bordeaux, France, June 2007.

[29] Kar-Ming Cheung, Adans Ko, Van Dan, and David Heckman. Risk analysis for non-deterministic mission planning and sequencing. *Proceedings of IEEE Aerospace Conference*, pages 147–155, Big Sky, Montana, March 2005.

[30] Kar-Ming Cheung, Adans Ko, Van Dan, and David Heckman. Risk analysis framework for non-deterministic planning and sequencing. In *IEEE Aerospace Conference Presentation*, Big Sky, Montana, March 2005.

[31] MIT 16.83 Class. Tersat design document. Technical report, Massachusetts Institute of Technology, 2011.

[32] Robert T. Clemen and Robert L. Winkler. Combining probability distributions from experts in risk analysis. *Society for Risk Analysis*, 19:187–203, 1999.

[33] Peter Congdon. *Bayesian Statistical Modeling.* Wiley & Sons, 2001.

[34] Roger M. Cooke. *Experts in Uncertainty: Opinion and Subjective Probability in Science.* Oxford University Press, 1991.

[35] Steven L. Cornford, Martin S. Feather, Julia R. Dunphy, and Jose Salcedo. Optimizing spacecraft design optimization engine development: Progress and plans. *Proceedings of IEEE Aerospace Conference*, Big Sky, Montana, March 2003.

[36] Vittorio Cortellessa. Model-based performance risk analysis. *Proceedings of IEEE Aerospace Conference*, 31:3–20, Big Sky, Montana, March 2005.

[37] Henry E. Daniels. Saddle-point approximation in statistics. *The Annals of Mathematicals Statistics*, 25:631–650, 1954.

[38] Anirban DasGupta. *Asymptotic Theory of Statistics and Probability.* Springer, 2008.

[39] Mark E. Davis. Space based radar technology challenges. *Proceedings of IEEE Aerospace Conference*, Big Sky, Montana, March 2005.

[40] Edwin B. Dean. Parametric cost analysis: a design function. *Transactions of American Association of Cost Engineers (33rd Annual Meeting)*, 1989.

[41] Robin L. Dillon, M. Elisabeth Pate'-Cornell, and Seth D. Guikema. Programmatic risk analysis for critical engineering systems under tight resource constraints. *Journal of Operations Research*, 51:354–370, 2003.

[42] Didier Dubois and Henri Prade. Possibility theory and data fusion in poorly informed environments. *Journal of Control Engineering Practice*, 5:811–823, 1994.

[43] Edward J. Dudewicz and Satya N. Mishra. *Modern Mathematical Statistics.* Wiley & Sons, 1988.

[44] ESA. Esa cdf. *http://www.esa.int/esaMI/CDF/SEMQOF1P4HD0.html.*, 2010.

[45] Michael D. Escobar and Mike West. Bayesian density estimation and inference using mixtures. *Journal of American Statistical Association*, 90:577–588, 1995.

[46] Barry George Evans. *Satellite Communications Systems: third edition.* Institution of Engineering and Technology, 2000.

[47] Donald W. Fiske and Nick F. Pidgeon. The conjunction fallacy: the case for the existence of competing heuristic strategies. *British Journal of Psychology*, 88:1–27, 1990.

[48] Jean Pierre Florens, Michel Mouchart, and Jean-Marie Molin. *Elements of Bayesian Statistics.* Dekker, Pure and applied mathematics serie, 1990.

[49] Martin Fuchs and Arnold Neumaier. Autonomous robust design optimization with potential clouds. *International Journal of Reliability and Safety*, 3:23–34, 2009.

[50] Martin Fuchs and Arnold Neumaier. Potential based clouds in robust design optimization. *Issue on Imprecision of Statistics and Applied Mathematics*, 2009.

[51] Martin Fuchs and Arnold Neumaier. Handling uncertainty in higher dimensions with potential clouds towards robust design optimization. *Soft methods for handling variability and imprecision*, 48:376–382, 2010.

[52] Martin Fuchs, Arnold Neumaier, and Daniela Girimonte. Uncertainty modeling in autonomous robust spacecraft system design. *Proceedings in Applied Mathematics and Mechanics*, 7:4142, 2008.

[53] Paul H. Garthwaite, Joseph B. Kadane, and Anthony O'Hagan. Statistical methods for eliciting probability distributions. *Journal of the American Statistical Association*, 470:680–701, 2005.

[54] Alan E. Gelfand, Bani K. Mallick, and Dipak K. Dey. Modeling expert opinion arising as a partial probabilistic specification. *Journal of American Statistical Association*, 90:598–604, 1995.

[55] Carl E. Gilchriest. Spacecraft mass trade-offs versus radiofrequency power and antenna size at 8 ghz and 32 ghz. *TDA Progress Report*, 1987.

[56] Vincent Goulet, Michel Jacques, and Mathieu Pigeon. Modeling without data using expert opinion. *The R Journal*, 1:31–36, 2009.

[57] Gudela Grote. Uncertainty management at the core of system design. *Annuals Reviews in Controls*, 28:267–274, 2004.

[58] David H. Gustafson, Ramesh K Shukla, Andre Delbeck, and William Wilster. A comparative study of differences in subjective likelihood estimates made by individuals, interacting groups, delphi groups, and nominal groups. *Organizational Behavior and Human Performance*, 9:280–291, 1973.

[59] Corey Harmon, Keith Warfield, and Leigh Rosenberg. Jpl cost risk analysis: an approach that incorporates engineering realism. *ISPA Conference*, 2006.

[60] Rania Hassan and William Crossley. Discrete design optimization under uncertainty: a generalized population based sampling genetic algorithm. *AIAA Journal*, 45:2799–2809, 2007.

[61] Rania Hassan, Olivier L. De Weck, and Philip Springmann. Architecting a communication satellite product line. *22nd AIAA International Communications Satellite Systems Conference and Exhibit*, Monterey, California, May 2004.

[62] H. Hemmati, K. Wilson, M. Sue, D. Rascoe, F. Lansing, M. Wilhelm, L. Harcke, and C. Chen. Comparative study of optical and rf communication systems for a mars mission. *Proc. SPIE Vol. 2699, Free-Space Laser Communication Technologies VIII*, April 1996.

[63] M.I. Herman, S. Burkhart, R. Crist, J. Vacchione, R. Hughes, K. Kellogg, A. Kermode, D. Rascoe, C. Hornbuckle, W. Hoffmann, Martin Marietta Corporation, and D. Smith. Microtechnology in telecommunications for spacecraft cost and mass reduction. *IAA International Conference on Low Cost Planetary Missions*, Laurel, Maryland, April 1994.

[64] William Imbriale. *Spaceborne Antennas for Planetary Exploration.* NASA JPL Descanso Publications, 2006.

[65] Alan Julian Izenman. Recent developments in non-parametric density estimation. *Journal of the American Statistical Association*, 86:205–224, 1991.

[66] Robert A. Jacobs. Methods for combining experts' probability assessments. *Neural Computation*, 7:867–888, 1995.

[67] Harold L. Jacobson. The maximum variance of restricted unimodal distributions. *The Annals of Mathematical Statistics*, 40:1746–1752, 1969.

[68] Michel Jacques and Mathieu Pigeon. Using expert opinion in actuarial sicence. *Proceedings of 41st Actuarial Research Conference*, Montreal, Canada, 2005.

[69] Cyrus D. Jilla. *A Multiobjective, Multidisciplinary Design Optimization Methodology for the Conceptual Design of Distributed Satellite Systems.* PhD thesis, Massachusetts Institute of Technology, 2002.

[70] Cyrus D. Jilla and David W. Miller. Application of multidisciplinary design optimization techniques to distributed satellite systems. *Journal of Spacecraft and Rockets*, 37:481–490, 2000.

[71] Cyrus D. Jilla and David W. Miller. Assessing the performance of a heuristic simulated annealing algorithm for the design of distributed satellite systems. *Acta Astronautica*, 48:5–12, 2001.

[72] Cyrus D. Jilla and David W. Miller. Multi-objective, multidisciplinary design optimization methodology for distributed satellite systems. *Journal of Spacecraft and Rockets*, 41:39–49, 2004.

[73] Taylor Jim, Dennis K. Lee, and Shervin Shambayati. *Mars Reconaissance Orbiter-Telecommunications.* JPL Descanso Design and Performance Summary Series, 2006.

[74] Chris M. Jones, T. S. Marron, and S. J. Sheater. A brief survey of bandwidth selection for density estimation. *Journal of American Statistical Association*, 91:401–407, 1996.

[75] Yuen Joseph. *Deep Space Telecommunication Systems Engineering.* Plenum Press, 1983.

[76] Brian Kennedy, J. Edmund Riedel, Shyam Bhaskaran, Shailen Desai, Don Han, Tim McElrath, George Null, Mark Ryne, Steve Synnott, Mike Wang, and Robert Werner. *Deep Space 1 Telecommunications.* JPL Descanso Design and Performance Summary Series, 2001.

[77] Joel Kohler. The base rate fallacy reconsidered: Descriptive, normative, and methodological challenges. *Behavioral and Brain Science*, 19:1–17, 1996.

[78] Peter M. Lee. *Bayesian Statistics: An Introduction, Third Edition.* Wiley & Sons, 2000.

[79] Sarah Lichtenstein, Baruch Fischhoff, and Lawrence Phillips. Calibration of probabilities, the state of the art to 1980. *Technical Report, Office of Naval Research*, 1981.

[80] Charles F. Lillie and Bruce E. Thompson. Parametric cost estimation for space science mission. *SPIE Paper 7018-81*, 2008 (June).

[81] Dennis V. Lindley. *Bayesian Statistics: a Review.* Regional conference series in applied mathematics, 1972.

[82] Dennis V. Lindley and Nozer D. Singpurwalla. Reliability (and fault tree) analysis using expert opinions. *Journal of the American Statistical Association*, 81:87–90, 1986.

[83] D. O. Loftsgaarden and P. C. Quesenberry. A non-parametric estimate of a multivariate density function. *Annals of Mathematical Statistics*, 36:1049–1051, 1965.

[84] Katie Grantham Lough and Robert B. Stone. The risk in early design (red) method: Likelihood and consequence formulations. *Proceedings of International Design Engineering Technical Conference and Computers and Information in Engineering Conference*, 2006.

[85] Andre Makovsky, Andrea Barbieri, and Ramona Tung. *Odyssey Telecommunications.* JPL Descanso Design and Performance Summary Series, 2002.

[86] Andre Makovsky, Ilott Peter, and Jim Taylor. *Mars Science Laboratory Telecommunications System Design.* JPL Descanso Design and Performances Summary Series, 2009.

[87] Gerard Maral and Michel Bousquet. *Satellite Communications Systems: Systems, Techniques and Technologies (5th Edition).* Wiley & Sons, 2002.

[88] Harry F. Martz and Maurice C. Bryson. A statistical model for combining biased expert opinions. *IEEE Transactions on Reliability*, 33:227–232, 1984.

[89] Edward L. Melnick and Brian S. Everitt. *Encyclopedia of quantitative risk analysis and assessment, Volume 1.* 2008.

[90] Charles Merriam and Noah Webster. *Merriam-Webster's Collegiate Dictionary.* Library of Congress, 2011.

[91] Leila Meshkat. An holistic approach for risk management during design. *IEEE Aerospace Conference*, 6:1–5, Big Sky, Monatana, March 2007.

[92] Leila Meshkat, Steve Cornford, and Terence Moran. Risk based decision tool for space exploration mission. *AIAA Journal*, 2003.

[93] Jouini N. Mohamed and Robert T. Clemen. Copula models for aggregating expert opinions. *Journal of Operations Research*, 44:444–457, 1996.

[94] Adam Mosleh, Van M. Bier, and George Apostolakis. A critique of current practice for the use of expert opinions in probabilistic risk assessment. *Reliability Engineering and System Safety*, 20:63–85, 1988.

[95] NASA. *NASA System Engineering Handbook.* 2007.

[96] NASA-JPL. Team-x website. *http://jplteamx.jpl.nasa.gov/*, 2010.

[97] Richard De Neufville and Olivier L. DeWeck. Uncertainty management for engineering systems planning and design. *1st Engineering Systems Symposium*, 2004.

[98] William L. Oberkampf, Jon C. Helto, Cliff A. Joslyn, Steven F. Wojkiewicz, and Scott Ferson. Challenge problems: uncertainty in system response given uncertain parameters. *Reliability Engineering and System Safety*, 85:11–19, 2004.

[99] Anthony O'Hagan. Eliciting expert beliefs in substantial practical applications. *Journal of the Royal Statistics Society*, 47:21–35, 1998.

[100] Anthony O'Hagan, Caitlin Buck, Alireza Daneshkhah, Richard Eiser, and Paul Garthwaite. *Uncertain Judgements: Eliciting Experts Probabilities.* Lavoisier Libraire, 2006.

[101] Panos Y. Palambros and Douglass J. Wilde. *Principles of Optimal Design.* Cambridge University Press, 2000.

[102] Emanuel Parzen. On estimation of a probability density function and mode. *Annals of Mathematical Statistics*, 33:1065–1076, 1962.

[103] Samuel P. Pullen and Bradford W. Parkinson. System design under uncertainty: Evolutionary optimization of the gravity probe-b spacecraft. *Parallel Problem Solving From Nature*, 866:598–607, 1994.

[104] Nancy Reid. Saddle-point methods and statistical inference. *Statistical Science*, 3:213–238, 1988.

[105] Dan O. Reudnik. Maximum power transmission in mass limited radio systems. *IEEE Transactions on Communications*, 30:414–417, 1982.

[106] Madhavendra Richaria. *Satellite Communications Systems: Second Edition.* McGraw-Hill, 1999.

[107] Leight Rosenberg, John Hihn, Kim Roust, and Keith Warfield. Parametric cost modeling of space missions using the develop new projects (dnp) implementation process. *AIAA Journal*, 2000.

[108] Murray Rosenblatt. Remarks on some non-parametric estimates of a density function. *Annals of Mathematical Statistics*, 27:832–835, 1956.

[109] Harold Sackman. *Delphi critique; expert opinion, forecasting, and group process.* Lexington Books, 1974.

[110] Ron C. Schultz and Hill Stuart. The new horizons high gain antenna: Reflector design for a spin-stabilized bus at cryogenic temperatures. *IEEE Aerospace Conference Proceedings*, 2:966–974, 2004.

[111] David Scott. Multi-dimensional density estimation. *Handbook of Statistics*, 24:229–258, 2000.

[112] David Seaver, Detlof von Winterfeldt, and Ward Edwards. Eliciting subjective probability distributions on continuous variable. *Organizational Behavior and Human Performance*, 21:3790381, 1977.

[113] John Shawe-Taylor and Alexander N. Dolia. A framework for probability density estimation. *Journal of Machine Learning Research*, 2:419–444, 2002.

[114] J. Sherif, D. Remer, and R. Buchanan. Long-range planning cost model for support of future space missions by the deep space network. *NASA TDA Progress Report*, 1990.

[115] Bernard W. Silverman. *Density Estimation for Statistics and Data Analysis.* Chapman&Hall, 1998.

[116] Bernard W. Silverman and Chris M. Jones. Fix and hodges (1951): An important contribution to non-parametric discriminant analysis density estimation. *International Statistical Review*, 57:233–247, 1989.

[117] Joseph Sobieczanski-Sobieski and Raphael T. Haftka. Multidisciplinary aerospace design optimization: survey of recent developments. *Structural Optimization*, 14:1–23, 1997.

[118] Surrey. *Surrey Company: Antenna Pointing Mechanism HighGain X Band Antenna. s.l. : Surrey Company Datasheet, 2010.*

[119] Jim Taylor. *Dawn Telecommunications.* JPL Descanso Design and Performance Summary Series, 2009.

[120] Jim Taylor, Kar-Ming Cheung, and Dongae Seo. *Galileo Telecommunications.* JPL Descanso Design and Performance Summary Series, 2002.

[121] Jim Taylor, Kar-Ming Cheung, and Chao-Jen Wong. *Mars Global Surveyor Telecommunications.* JPL Descanso Design and Performance Summary Series, 2001.

[122] Jim Taylor and Hansen David. *Deep Impact Flyby and Impactor Telecommunications.* JPL Descanso Design and Performance Summary Series, 2005.

[123] Jim Taylor, Andre Makovsky, Andrea Barbieri, Ramona Tung, Polly Eastbrook, and Gail A. Thomas. *Mars Exploration Rover Telecommunications.* JPL Descanso Design and Performance Summary Serie, 2005.

[124] Jim Taylor and Ludwig Roger. *Voyager Telecommunications.* JPL Descanso Design and Performance Summary Series, 2002.

[125] Jim Taylor, Laura Sakamoto, and Chao-Jen Wong. *Cassini Orbiter / Huygens Probe Telecommunications.* JPL Descanso Design and Performance Summary Series, 2002.

[126] R. Terrile, M. Kordon, M. Postma, J. Salcedo, D. Hanks, and E. Wood. Automated design of spacecraft telecommunication subsystems using evolutionary computational techniques. *IEEE Journal*, 1:1–9, 2009.

[127] Amos Tversky and Daniel Kahneman. Belief in the law of small numbers. *Psychological Bullettin*, 76:105–110, 1971.

[128] Amos Tversky and Daniel Kahneman. Judgments under uncertainty: heuristics and biases. *Science*, 185:1124–1131, 1974.

[129] Amos Tversky and Daniel Kahneman. Extensional vs. intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, 90:293–315, 1990.

[130] Amos Tversky and Daniel Kahneman. Belief in the law of small numbers. *A handbook of data analysis in the behavioral sciencies: methodological issues*, 1:341–349, 1993.

[131] Roland Vanderspek and Joel Villasenor. Hete (high energy transient explorer). *http://space.mit.edu/HETE/*, 2007.

[132] Olivier L. De Weck and Marshall B. Jones. Isoperformance: Analysis and design of complex systems with known or desired outcomes. *Fourteenth Annual International Symposium of the International Council on Systems Engineering (INCOSE)*, 2004.

[133] Olivier L. De Weck and David W. Miller. Multivariable isoperformance methodology for precision opto-mechanical systems. *AIAA Journal*, 2002.

[134] Olivier L. De Weck, Richard F. De Neufville, and Mathieu Chaize. Staged deployment of communications satellite constellations in low earth orbit. *Journal of Aerospace Computing, Information, and Communication*, 3:119–136, 2004.

[135] James R. Wertz and Wiley J. Larson. *Space Mission Analysis and Design.* Space Technology Library, 1999.

[136] Karen Willcox, Doug Allaire, John Deyst, Chelsea He, and George Sondecker. Stochastic process decision methods for complex cyber-physical systems. *Air Force Research Laboratory Report*, AFRL-RZ-WP-TR-2011-2094, 2011.

[137] Bram Wisse, Tim Bedford, and John Quigley. Expert judgment combination using moment methods. *Reliability Engineering and System Safety*, 93:675–686, 1990.

[138] William Wright and U. Anderson. Effects of situation familiarity and financial incentives on use of the anchoring and adjustment heuristic for probability assessment. *Human Decision Processes*, 44:66–82, 1989.

[139] Du Xiaoping and Wei Chen. Efficient uncertainty analysis methods for multidisciplinary robust design. *AIAA Journal*, 40:545–555, 2002.

[140] Joseph Yuen and Laura Sakamoto. Spacecraft telecommunication systems mass estimates. *JPL-TDA Report*, 1987.

[141] Jian Zhou, Yaoqin Zhu, and Weiqing Tang. Approach of expert opinion acquisition based on cloud model and evidence theory. *Management and Service Science International Conference*, 1:1–4, 2009.