UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA ELÉCTRICA

IONOSPHERIC ANALYSIS USING INCOHERENT SCATTER RADAR AND IN-SITU MEASUREMENTS

TESIS PARA OPTAR AL GRADO DE DOCTOR EN
INGENIERÍA ELÉCTRICA

MIGUEL MARTÍNEZ LEDESMA

PROFESOR GUÍA:
MARCOS ANDRÉS DÍAZ QUEZADA

MIEMBROS DE LA COMISIÓN:
MARCOS EDUARDO ORCHARD CONCHA
JORGE FELIPE SILVA SÁNCHEZ
MARCO ANTONIO MILLA BRAVO
MARINA STEPANOVA

Este trabajo ha sido financiado por la beca CONICYT-PFCHA/Doctorado Nacional para Extranjeros/2014-63140114

SANTIAGO DE CHILE
2019

RESUMEN

RESUMEN DE LA TESIS PARA OPTAR AL GRADO DE:
DOCTOR EN INGENIERÍA ELÉCTRICA
POR: MIGUEL MARTÍNEZ LEDESMA
FECHA: 09/05/2019
PROFESOR GUÍA: DR. MARCOS ANDRÉS DÍAZ QUEZADA

IONOSPHERIC ANALYSIS USING INCOHERENT SCATTER RADAR AND IN-SITU MEASUREMENTS

Existen problemas de ambigüedad en la estimación de los parámetros del plasma ionosférico entre los 130 y los 300 km en los Radares de Dispersión Incoherente (ISR). En esos rangos, el ISR es incapaz de distinguir entre diferentes mezclas de iones moleculares (NO^+ y O_2^+) y iones atómicos de oxígeno (O^+). Los métodos comúnmente utilizados para solucionar este problema son el uso de modelos empíricos o teóricos de la Ionosfera, o añadir información conocida *a priori* de parámetros del plasma obtenidos de la Línea de Plasma del radar ISR. Por otro lado, algunos estudios han demostrado que se pueden estimar de forma no ambigua los parámetros del plasma analizando señales casi sin ruido, aunque esas características de ruido no son típicamente obtenidas en las mediciones rutinarias de los radares ISR. En este trabajo de tesis se define un entorno teórico para cuantificar el problema de la ambigüedad y determinar los niveles de fluctuación de señal máximos para estimar de forma no ambigua las señales ISR. Realizamos diversas simulaciones Monte Carlo de diferentes parámetros de plasma que nos permiten evaluar el desempeño de la estimación del algoritmo de optimización de mínimos cuadrados no lineal (NLLS) más comúnmente utilizado en ISR. Los resultados de estas simulaciones se muestran como curvas de probabilidad de *convergencia válida* y *estimación 'correcta'*. Además, realizamos simulaciones para cuantificar el error de estimación que se obtiene cuando se utilizan modelos ionosféricos para determinar las condiciones iniciales de los parámetros de plasma del algoritmo de optimización. A su vez, determinamos el efecto del conocimiento *a priori* de diferentes combinaciones de parámetros obtenidos de la Línea de Plasma, la información que contribuyen cada uno de los parámetros del plasma, y el impacto de incrementar la incertidumbre de esos parámetros conocidos *a priori*. Los resultados sugieren que el conocimiento *a priori* de la densidad de electrones y la temperatura de electrones permite estimar de forma no ambigua incluso en niveles de fluctuación de señal elevados. Mediciones *in-situ* de sensores en satélites o cohetes pueden obtener esos parámetros de plasma conocidos *a priori* para ayudar a la estimación correcta de los radares. Los resultados obtenidos en este estudio ayudan a determinar los límites de estimación de la técnica ISR y proveen nuevas herramientas que mejoran la probabilidad de estimación no ambigua de los parámetros del plasma.

ABSTRACT

ABSTRACT OF THE THESIS TO OBTAIN THE DEGREE OF:
DOCTORATE IN ELECTRIC ENGINEERING
BY: MIGUEL MARTÍNEZ LEDESMA
DATE: 09/05/2019
ADVISOR: DR. MARCOS ANDRÉS DÍAZ QUEZADA

IONOSPHERIC ANALYSIS USING INCOHERENT SCATTER RADAR AND IN-SITU MEASUREMENTS

Unambiguously estimating the plasma parameters of the ionosphere at altitudes between 130 and 300 km presents a problem for the Incoherent Scatter Radar (ISR). At these ranges, ISR is unable to distinguish between different mixtures of molecular ions (NO^+ and O_2^+) and atomic oxygen ions (O^+). Common solutions to this problem are either to employ empirical or theoretical models of the ionosphere, or to add *a priori* known plasma parameter information obtained from the Plasma Line of the ISR spectrum. Studies have demonstrated that plasma parameters can be unambiguously estimated from ISR signals in almost noiseless scenarios, not commonly feasible during routine monitoring. In this thesis work, we define a theoretical framework to quantify the ambiguity problem and determine the maximum signal fluctuation levels of the ISR signal to unambiguously estimate plasma parameters. We conduct Monte Carlo simulations for different plasma parameters to evaluate the estimation performance of the most commonly used Non-Linear Least Squares optimization algorithm. Results are shown as probability curves of *valid convergence* and '*correct' estimation*. We also use simulations to quantify the estimation error when using ionospheric models as initial conditions of the optimization algorithm. We also determine the contribution to the estimation process of different combinations of parameters known from the Plasma Line, the particular contribution of each plasma parameter, and the effect of increasing the level of uncertainty of the parameters known *a priori*. Results suggest that knowing *a priori* both electron density and electron temperature parameters allows an unambiguous estimation even at high fluctuation levels. *In-situ* measurements of sensors onboard of satellites or rockets could measure these *a priori* plasma parameters to help determine the correct estimation parameters. Results obtained in this study help determine the estimation limits of the ISR technique and provide new tools that improve the probability of unambiguous estimations of plasma parameters.

Este trabajo ha sido financiado por una beca de doctorado del Programa Formación de Capital Humano Avanzado del Gobierno de Chile CONICYT-PCHA/Doctorado Nacional para extranjeros/2014-63140114.

Powered@NLHPC: Esta investigación fue parcialmente apoyada por la infraestructura de supercómputo del NLHPC (ECM-02).

The simulation data and codes shown in this document are available at the Zenodo public repository:

Martínez-Ledesma, Miguel, & Díaz Quezada, Marcos Andrés. (2018). *Data and Code for: Determination of the Signal Fluctuation Threshold of the Temperature Ion Composition Ambiguity Problem Using Monte Carlo Simulations (Version Version 1)* [Data set]. Zenodo. https://doi.org/10.5281/zenodo.1466184.

Dedicado a Laura.

ACKNOWLEDGMENTS

I deserve special thanks to my advisor Dr. Marcos Andres Díaz Quezada. I'm very thankful for the time and support he has given me during the development of this thesis.

I have much more than compliments to the Space and Planetary Exploration Laboratory (SPEL) group of students and engineers. It has been more than four years of work together with a great team. I really hope that all of them succeed in all their dreams.

I deserve also many thanks to Dr. Marcos Eduardo Orchard Concha and Dr. Jorge Silva Sanchez, professors of the Information and Decision Systems Group (IDS) of the Electric Engineering Department of the FCFM of University of Chile (UChile). They have spent many hours in fructiferous meetings and discussions that encouraged me to continue this project. I would like to thank also Dr. Marina Stepanova of the Physics Department of University of Santiago de Chile (USACH) for her support and to allow me assist to her classes.

I would like to thank all the help and support that Dr. Francisco Roberto Jaramillo Montoya has provided to me. Many good discussions concerning optimization have been very relevant to the development of this project. He has also development the PSO algorithm that will be used in a future study, as a continuation of this thesis work.

J. Semeter (BU), P. Erickson (MIT), L. Goncharenko (MIT), and J. Swoboda (MIT) have provided me the great opportunity to learn from them and work with them. During my short internship at Boston University, they gave me a lot of information that about how real ISR radars work. I'm really grateful for all their help and support. They have really encouraged me to continue working in this field.

Por supuesto, nada de esto hubiera sido posible sin el apoyo y cariño de mi familia y amigos. Los de aquí y los de allí. En Chile y en España. Todos ellos están en mis pensamientos en todo momento, les debo el seguir peleando día a día.

Tú has sido mi colaboradora más constante, mi directora más presente, mi ayudante más dedicada, mi revisora más detallista, mi guía en todo momento, y mi mejor ejemplo de lo que quiero ser. Tú sabes mejor que nadie lo difícil que ha sido, todo el tiempo que le he dedicado, y lo mucho que te he necesitado en todo este proceso. Te lo debo a ti. Este trabajo es todo por y para ti, Laura.

Miguel Martínez Ledesma, 09th, May 2019

"All this is a dream. Still, examine it by a few experiments. Nothing is too wonderful to be true, if it be consistent with the laws of nature; and in such things as these, experiment is the best test of such consistency."

Michael Faraday

Laboratory journal entry #10.040, 19th March, 1849

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

CHAPTER 1

INTRODUCTION

1.1. MOTIVATION

The Incoherent Scatter Radar (ISR) is one of the most powerful ground-based sounding techniques to measure the ionosphere [Evans, 1969] [Beynon and Williams, 1978]. This method is a radar that transmits a powerful electromagnetic pulse the ionosphere and receives a backscatter signal due to the Thomson Scatter effect on the ionospheric electrons. Plasma parameter estimates are obtained from the analysis of the ISR backscattered signal autocorrelation function (ACF) or its Fourier Transform, the Incoherent Scatter Spectra (ISS).

At altitudes varying from 120 km to 300 km approximately, the estimation of plasma parameters from ISR signals is ambiguous because of the existence of a mixture of different ions in the ionosphere [Oliver, 1979]. ISR signals at the Ion Acoustic frequency band of the radar are dependent on the ratio of ion temperature and mass (T_i/m_i) [Oliver, 1979] [Vallinkoski, 1988]. Therefore, similar backscatter signals are obtained with different combinations of molecular ions (NO^+ and O_2^+) and atomic (O^+) ions found at these altitudes [Aponte et al., 2007]. Two possible solutions of ion composition are obtained with almost symmetric values [Lathuillere et al., 1983] [Wu et al., 2015]. Also, different temperatures are obtained related to the solutions of ion composition, hindering the correct estimation of plasma parameters. This *Temperature-Ion Composition Ambiguity* (TICA) is a relevant scientific issue that affects the determination of long-term trends of ionospheric and thermospheric variables currently under discussion [Perrone & Mikhailov, 2017] [Zhang et al., 2018] [Perrone & Mikhailov, 2018].

1.2. PROBLEM STATEMENT

The TICA problem has been typically solved in the literature by providing additional information that constrained the feasible solutions [Aponte et al., 2007]. The most common method to solve the ambiguity is by the use of theoretical or empirical models of ionospheric plasma parameters [Waldteufel, 1971] [Evans & Oliver, 1972] [Oliver, 1979] [Cabrit & Kofman, 1996] [Litvine et al., 1998]. Nevertheless, the use of ionospheric models may induce to estimation errors when these models are inaccurate, or when particular physical phenomena are not properly represented or are smoothed by these models.

An alternative method used to solve the ambiguity is to provide simultaneously estimated plasma parameters from the Plasma Line frequency band of the radar and from the total power received at the antenna [Wand, 1970] [Waldteufel, 1971] [Bjørnå & Kirkwood, 1988] [Nicolls et al., 2006] [Aponte et al., 2007]. Nevertheless, Plasma Line information is not routinely available at most ISRs because of the small signal strength of this frequency band, although occasional enhancements allow its detection under energetic events [Akbari et al., 2017].

A different method to unambiguously estimate plasma parameters is based on the determination of the two possible solutions of ion composition and the selection of the solutions which provide the smoother altitude profile [Oliver 1979] [Lathuillere et al., 1983] [Lathuillere & Pibaret, 1992]. Nevertheless, this method requires the use of almost noiseless ISR signals [Oliver, 1979]. Long integration times were required to obtain small signal fluctuations in previous studies of this technique [Lathuillere et al., 1983] [Lathuillere & Pibaret, 1992]. However, the integration of signals coming from different radar pulses requires the assumption of stationary plasma conditions during the integration period [Farley, 1969]. The stationary plasma assumption can smooth and hide the dynamic phenomenon occurring in the ionosphere.

Recently, a new optimization algorithm (i.e. the Particle Swarm Optimization algorithm, PSO) has been used to unambiguously estimate plasma parameters of ISR signals with small fluctuation characteristics [Wu et al., 2015]. These signals were obtained with Signal-to-Noise Ratio (SNR) values in the range of 15 to 25, by using the simultaneous frequency transmission [Sulzer, 1986a] and the Coded Long Pulse [Sulzer, 1986b] techniques at the Arecibo Observatory. In this case, the PSO algorithm obtained much better estimations than the standard Non-Linear Least Squares (NLLS) optimization algorithm commonly used [Wu et al., 2015].

Results from [Lathuillere et al., 1983], [Lathuillere & Pibaret, 1992], and [Wu et al., 2015] suggest the feasibility to unambiguously estimate plasma parameters from signals with very small fluctuations. Nevertheless, no previous study has reported the unambiguous estimation of plasma parameters of signals with very small fluctuations using the most commonly used NLLS fitting algorithm in the ISR literature: the Levenberg-Marquardt (L-M) optimization algorithm [Levenberg, 1944] [Marquardt, 1963]. Furthermore, no previous study has assessed the required signal fluctuation level to unambiguously estimate plasma parameters in the TICA problem.

1.3. RESEARCH QUESTIONS

Considering the problem stated in Chapter 1.2, this doctoral thesis addresses the following general research question:

"It is possible to unambiguously estimate plasma parameters from ISR signals with very small signal fluctuation using the most commonly used NLLS optimization algorithm without providing additional information?".

This general research question generates further relevant secondary research questions:

1) "What is the signal fluctuation level required to solve the ambiguity?",

2) "Does the addition of information change the signal fluctuation threshold?",

3) "What information would help solve the TICA problem at high signal fluctuations?".

These questions intend to provide a general study framework to determine the solvability of the TICA problem using the most commonly used NLLS optimization algorithm at different fluctuation levels. Furthermore, this framework can compare the effectiveness of the different methods of adding information previously studied in the literature.

1.4. HYPOTHESES

The hypotheses presented in this doctoral thesis are the following:

- It is possible to unambiguously estimate plasma parameters from ISR signals with small fluctuation levels, even in the case when no additional information is provided, using the most commonly used NLLS fitting algorithm: the Levenberg-Marquardt (L-M) optimization algorithm.

- The signal fluctuation level required to solve the ambiguity is dependent on the *a priori* information provided (i.e. type and uncertainty of plasma parameters provided from the analysis of the Plasma Line or from *in-situ* sensors).

1.5. RESEARCH OBJECTIVES

1.5.1. GENERAL OBJECTIVE

The main objective of this doctoral thesis is to provide an assessment framework to quantify the *Temperature-Ion Composition Ambiguity* (TICA) problem of the Incoherent Scatter Radar (ISR) at different scenarios using the most commonly used NLLS optimization algorithm.

1.5.2. SPECIFIC OBJECTIVES

The specific objectives of this research are summarized below.

- Implement an ISR estimation simulation framework to study the ambiguous estimation at different signal fluctuation levels using the most commonly used NLLS optimization algorithm.

- Propose a methodology to determine the validity of the estimated solutions obtained and quantify the probability of unambiguous estimation.

- Determine the signal fluctuation level required to estimate unambiguously plasma parameters in the case when no additional information is provided.

- Determine the effect of providing *a priori* plasma parameters (obtained from the analysis of the Plasma Line frequency band of the radar or from *in-situ* sensors) in the ambiguous estimation of ISR signals.

- Determine the impact of providing *a priori* plasma parameters with different levels of uncertainty in the ambiguous estimation of ISR signals.

- Determine the effectiveness of using ionospheric models to set the initial guess of plasma parameters of the optimization algorithm search.

1.6. CONTRIBUTIONS OF THIS WORK

The main contributions of this doctoral thesis are summarized below.

New methodology to analyze the ambiguous estimation of ISR signals:

- A method to quantify the estimation performance of the most commonly used NLLS optimization algorithm in ISR (i.e. the Levenberg-Marquardt optimization algorithm) at different signal fluctuation levels by running a Monte Carlo estimation analysis of multiple input plasma parameters uniformly selected from the search range.

- A new estimation method independent on the selection of the initial parameters of the optimization algorithm search by executing multiple times the estimation process with different initial parameters uniformly selected from the search range.

- A new method to determine the convergence of the estimated solutions based on the statistical distribution of the reduced Chi-square cost function (χ_r^2).

- A new method to determine which solutions are 'correct' (i.e. select the global minima) or 'incorrect' based on the estimation error of the ion composition parameter using the Expectation Maximization clustering algorithm.

- A new ambiguity quantification method based on the probability of unambiguous estimation ($P_{correct}$) calculated as the ratio between the number of 'correct' solutions and the number of convergent solutions.

Unambiguous estimation of ISR signals with very small fluctuation levels:

- Demonstration that the probability of unambiguous estimation increases when the signal fluctuation level gets reduced. Results indicate that it is possible to solve the TICA problem ($P_{correct} \geq 95.45\%$) estimating with very small fluctuations ($\delta \leq 0.045\%$), even in the case when no additional information is provided.

Impact of adding additional information of plasma parameters:

- Demonstration that the addition of different plasma parameters improves the probability of unambiguous estimation ($P_{correct}$) at different signal fluctuation levels. Results demonstrate that knowing the ion drift parameter (V_i) does not provide any information to solve the ambiguous estimation.

- Previous studies assumed that the ambiguity problem was solved by the addition of N_e and T_e/T_i plasma parameters. Results indicate that, in this case, the ambiguity is solved ($P_{correct} \geq 95.45\%$) only for signal fluctuations levels $\delta \leq 0.57\%$. Furthermore, results suggest that knowing N_e and T_e plasma parameters solves the ambiguity problem even at high signal fluctuations ($\delta \leq 7.93\%$).

Impact of adding additional information of plasma parameters with uncertainty:

- Demonstration that the increase of uncertainty of *a priori* known parameters decreases the convergence of solutions ($P_{fit\ valid}$). Results indicate that convergences are similar to simulations without uncertainty when the signal fluctuation level (δ) is larger than the uncertainty level ($\delta \geq \epsilon$).

- Demonstration that the increase of uncertainty of *a priori* known parameters decreases the probability of unambiguous estimation ($P_{correct}$). Results indicate that the probability of unambiguous estimation is similar to simulations without uncertainty when the uncertainty level is $\epsilon \leq 1\%$. Also, results indicate that the ambiguity is solved independently of the uncertainty level (ϵ) for very signal small fluctuation levels ($\delta \leq 0.05\%$).

Effect of using ionospheric models to determine the initial set of parameters of the optimization algorithm search:

- Demonstration that the increase of uncertainty on the initial guess of plasma parameters (β) decreases the probability of unambiguous estimation ($P_{correct}$). Results indicate that for very small signal fluctuations ($\delta \leq 0.05\%$) parameters were unambiguously estimated ($P_{correct} \geq 95.45\%$) independently on the uncertainty of the initial guess (β).

- Previous studies assumed that providing an accurate initial guess would obtain an unambiguous solution. Results indicate that even with an accurate initial guess ($\beta = 1\%$), unambiguous estimations ($P_{correct} \geq 95.45\%$) are obtained only for signal fluctuation levels smaller than $\delta \leq 0.54\%$. Furthermore, even with an accurate initial guess (β = 1%), it is more probable to obtain the 'incorrect' solution ($P_{correct} < 50\%$) for signal fluctuations larger than $\delta > 5\%$.

1.7. PUBLICATIONS

During the development of this doctoral thesis, the author published the following scientific contributions:

Scientific Journals (ISI):

- ***Martínez-Ledesma, M.**; Díaz, M.; (2019) "Determination of the Signal Fluctuation Threshold of the Temperature-Ion Composition Ambiguity Problem Using Monte Carlo Simulations", Journal of Geophysical Research: Space Physics. doi:10.1029/2018JA026217. Impact Factor: 2.75.*

- *Díaz, M.; Zagal, J.C.; Falcon, C.; Stepanova, M.; Valdivia, J.A.;* ***Martínez-Ledesma, M.****; Díaz, J.; Romanova, N.; Pacheco, E.; Milla, M.; Orchard, M.; Silva, J.; Mena, F.P.; and Jaramillo, F.; (2016) "New opportunities offered by cubesats for space research in Latin America: the SUCHAI project case", Advances in Space Research, doi:10.1016/j.asr.2016.06.012. Impact Factor: 1.409.*

International Conferences and Workshops:

- *M. Díaz; C. Gonzalez; P. Moya;* ***M. Martínez Ledesma****; (2018) "Preliminary results of the first year of operation of the SUCHAI-1 Cubesat: Langmuir probe and particle counter measurements" at the AGU Fall Meeting 2018, 10-14th December 2018, Washington DC, USA.*

- ***M. Martínez-Ledesma****; C. González; A. Barjarya; M. Díaz; (2017) "Ionospheric Sensors on SUCHAI: Langmuir Probe", at The Magnetosphere: New Tools, New Thinking, New Results (2017 Conference), 13-17 of November 2017, at Puerto Varas, Chile.*

- ***M. Martínez-Ledesma****; A. Barjatya; M. Díaz; (2017) "New design of a planar fixed-potential Langmuir probe for measuring ionospheric electron density" at the First Workshop of Small Satellites for Space Weather Research and Forecasting, (SSWRF 2017); 1st to 4th August 2017, Catholic University of America (CUA), Washington DC, USA.*

- *E. Michael; F. Besser; C. Pollarolo;* ***M. Martinez****; R. Fuentes; L. Pallanca; (2015) "Fiber-based Heterodyne Near-Infrared Interferometry - An Instrumentation Study Platform on the way to the proposed Infrared Planet Formation Imager" at the European Week of Astronomy and Space Science 2015, 22 – 26 June 2015, La Laguna, Tenerife, Canary Islands, Spain.*

National Conferences and Workshops:

- *"Analysis of the Temperature-Ion Composition Ambiguity (TICA) Problem of Incoherent Scatter Radars (ISR)" by* ***M. Martínez-Ledesma*** *and M. Díaz; at Física Espacial en Chile: Presente y Futuro - Workshop del Proyecto Anillo ACT1405, 6th to 8th March 2019, Universidad de Chile and Universidad de Santiago de Chile, Santiago de Chile, Chile*

- *"Langmuir Probe Design for SUCHAI" by* ***M. Martínez-Ledesma****; M. Díaz; and A. Barjartya; at Física Espacial en Chile: Presente y Futuro - Workshop del Proyecto Anillo ACT1405, 6th to 8th March 2019, Universidad de Chile and Universidad de Santiago de Chile, Santiago de Chile, Chile*

- *"Radar Course I: Fundamental Concepts" and "Radar Course II: Ionospheric Radars"; by* ***M. Martínez-Ledesma****; at First Workshop CINFAA (Centro Interuniversitario Física Alta Atmósfera)-CEEULS (Centro Estudios Espaciales Universidad La Serena), 18th-20th December 2018, Universidad de La Serena, La Serena, Chile.*

- ***M. Martínez-Ledesma****; F. Jaramillo; M. Orchard; J. Silva; M. Díaz; (2016) "Ionospheric Inversion Using Optimization Algorithms" at Workshop de Estudiantes de Astronomía de la Universidad de Chile, 9th to 10th, May 2016, Observatorio Astronómico Nacional - Cerro Calan, Santiago de Chile, Chile.*

1.8. OUTLINE OF THE THESIS

Chapter 2 shows the basic principles of the ISR method, describes the most relevant ISR observatories, explains the general procedure to obtain ISR measurements from the ionosphere, describes the methods to analyze these measurements, determines the backscatter power strength received, explains the Plasma Line and the methods used to estimate plasma parameters from this frequency band, and explains the ambiguous estimation of the temperature and ion composition and the methods commonly used to solve this ambiguity.

Chapter 3 describes the Monte Carlo simulation method implemented, and the statistical methods implemented to determine the convergence and 'correctness' of the estimation results.

Chapter 4 describes the results of using different levels of inaccuracy in the initial guess of parameters of the optimization algorithm search at different signal fluctuation values.

Chapter 5 describes the resulting probabilities obtained at different signal fluctuation values with different parameters assumed to be known *a priori* from the Plasma Line.

Chapter 6 describes the probabilities obtained by assuming each plasma parameter known *a priori* individually to determine the most relevant parameters required to solve the ambiguous estimation.

Chapter 7 describes simulations with different levels of uncertainty in the *a priori* known parameters at different signal fluctuation values with different parameters assumed to be known *a priori* from the Plasma Line.

Chapter 8 summarizes the conclusions of this thesis, new improvements suggested to be applied to the ISR technique resulting from this work, and discusses possible future works related.

Annexes are included in this work to provide a more detailed explanation of some subjects of relevance for this study:

- **Annex 1** describes the ISR spectrum model used and shows graphics of variability for different values of plasma parameters of the Ion Acoustic frequency band of this spectrum model.
- **Annex 2** describes the standard estimation technique of ISR for Range-Gate analysis.
- **Annex 3** describes the Levenberg-Marquardt optimization algorithm used and its configuration.
- Finally, **Annex 4** describes the Expectation Maximization algorithm used and its configuration.

CHAPTER 2

BACKGROUND

2.1. INCOHERENT SCATTER RADAR

One of the most powerful and effective ground-based instruments to study the ionosphere is the Incoherent Scatter Radar (ISR). This is because of its capacity to obtain the most relevant plasma parameters from the entire ionospheric profile, from altitudes between 100 to beyond 1000 km altitude. The ionospheric plasma parameters estimated are the electron density (N_e), electron temperature (T_e), ion temperature (T_i), ion composition (p), and plasma ion drift velocity (V_i). Other secondary parameters can also be derived from the previous estimates, such as the neutral atmospheric temperature or the static electric field. For a detailed review of the ISR method see [Evans, 1969] [Beynon and Williams, 1978] [Bauer, 1975] [Walker, 1979] and references therein.

The ISR is a radar technique that sends a powerful electromagnetic pulse to the ionosphere and receives the backscattering generated by the ionospheric electrons due to the Thomson Scatter effect (see Annex 1. The Incoherent Scatter Spectrum). In this effect, the electrons get accelerated by the electric field of the transmitted signal, making them re-radiate energy in the form of electromagnetic waves with the same frequency as the incident wave. Only waves that satisfy the Bragg conditions would contribute to the returning signal: $k_{Bragg} = 2k_r$, where $k_r = 2\pi f_r/c$ is the wavenumber, being f_r the frequency of the radar and c the speed of light.

Due to the higher mass of ions, the ion acceleration is much smaller than the electron acceleration of the Thomson Scattering, so the electromagnetic radiation contribution of ions is almost negligible. For this reason, the backscatter signal was initially assumed to be generated by independent movements of electrons. This initial assumption was made by William E. Gordon when proposed the idea of a radar to measure the ionosphere [Gordon, 1958]. In this assumption, due to the randomness of the electrons in the plasma, electrons would be scattering incoherently. Therefore, received signal would have a Gaussian shaped spectrum, reflecting the Doppler effect of the Maxwellian distribution of electron speeds. Kenneth L. Bowles was the first to test this radar idea and measure a Thomson Scatter from the ionosphere [Bowles, 1958] [Bowles, 1961]. Those initial tests presented a different spectrum behavior and a more powerful backscattered signal. As Bowles correctly supposed from his observations, this effect was related to collective interactions between ions and electrons that came from macroscopic field fluctuations. Nevertheless, although the initial supposition of Gordon was found invalid, this ionospheric radar method has continued with the "incoherent" naming. A more suitable and commonly used naming for this radar technique is "Thomson" Scatter Radar method.

2.2. ISR FACILITIES

ISR observational facilities are formed by a large antenna, and a radio-frequency (RF) transmitter and receiver. Due to the small backscatter echo returned from the ionosphere, very high antenna gains and low noise temperature characteristics are required at reception. Typically, a radar with a mega-watt RF transmitter ($P_{tx} \sim 1\ [MW] = 10^6[W]$) constructed with klystrons amplifiers is used to obtain backscatter signals in the order of femto-watts ($P_{rx} \sim 1\ [fW] = 10^{-15}[W]$), which is slightly higher than the achievable reception noise power.

Typical ISR radar frequencies are between 30 MHz and 1.3 GHz, which are higher than the frequencies used by other sounding methods (such as Ionosondes). The ISR technique uses transmission frequencies (f_r) higher than the plasma frequency at the maximum density peak of the ionosphere ($f_0F_2 = (2\pi)^{-1}\sqrt{N_{e,max}e^2/m_e\epsilon_0}\ [Hz]$). This characteristic ($f_r \gg f_oF_2$) implies not receiving a specular reflection and allows measure backscatters from the entire ionospheric profile. Furthermore, the radar wavelength selected ($\lambda = c/f_r$[m]) is much larger than the Debye length of the plasma ($\lambda_D = \sqrt{\epsilon_0 kT_e/e^2N_e} \cong 69\sqrt{T_e/N_e}$ [m]). This later characteristic ($\lambda \gg \lambda_D$) allows the radar signal to affect the plasma collectively and obtain a backscattered signal proportional to density fluctuations of the plasma. Otherwise ($\lambda \ll \lambda_D$), the scale length of the transmitted variations would make the particles to act almost freely, obtaining a backscattered power extremely small and below detection limits [Evans, 1969].

Characteristics of world major ISR facilities are shown in Table 2.1 and their geographic locations in Figure 2.1. Radar parameters of observatories indicated in Table 2.1 have been obtained from [Yao et al., 2014a] [McKay and McCrea, 2009] [Erickson et al., 2008] [Balsley and Gage, 1980] [Sato et al., 2014] [Ding et al., 2018] and others. Further facilities were constructed but nowadays are unavailable. Decommissioned ISR observatories are the San Santin facility (France) [Bauer et al., 1974], and the Malvern radar (UK) [Williams and Taylor, 1974]. Also, the Chatanika radar (Alaska, USA) [Leadabrand et al., 1972] was moved to the Sondrestrom facility (Greenland, Denmark) in 1982 [McCready and Heinselman, 2013].

The earliest ISR facilities are the Arecibo Observatory (AO) at Puerto Rico (USA) and the Jicamarca Radio Observatory (JRO) at Jicamarca (Peru). The AO is the world largest antenna currently in operation, with a diameter of 305 meters (1000 feets), conceived by William E. Gordon in 1958 to receive the theoretically faint incoherent backscatter from the ionospheric electrons [Cohen, 2009]. Simultaneously to the construction of the AO, Kenneth L. Bowles designed and constructed the JRO. The JRO facility is a phased array antenna of 288m x 288m almost perfectly perpendicular to Earth magnetic fields, formed by 18432 half-wave dipoles working at a 49.92 [MHz] [Bowles et al., 1962].

Recent advances on ISR design are based in the use of modular phased array antennas, Software Defined Radio (SDR) technology, and low power solid-state transmitters [Valentic et al., 2013] [Grydeland et al., 2005] [Yao et al., 2014a] [Yao et al., 2014b]. The most recent ISR is the EISCAT_3D, which is an European Holographic ISR based on multiscatic phase array radars (still under development) [Wannberg et al., 2010] [McCrea et al., 2015]. Some of the new characteristics proposed in EISCAT_3D have been already tested at the Kilpisjärvi Atmospheric Imaging Receiver Array (KAIRA) located in Finland [McKay-Bukowski et al., 2015].

Table 2.1. System Specification of the World Major Incoherent Scatter Radar Facilities

ISR system	Location	f_{radar} (MHz)	λ_{radar} (m)	P_{tx} (MW)	η (%)	T_{noise} (K)	Antenna type	Beamwidth (°)	G (dBi)
EISCAT UHF	Tromsϕ, Norway	928	0.33	1.30	12.5	90/ 110	32m Cassegrain dish	0.6	48.1
EISCAT VHF	Tromsϕ, Norway, Receivers: Kiruna, Sweden Sodankylä, Finland	224	1.34	3.00	12.5	250	Four 30×40 m steerable parabolic cylinders (Tromsø); 32 m steerable parabolic dish (receivers)	0.6x1.8	46
EISCAT ESR	Longyearbyen, Svalbard	500	0.60	1.00	25.0	65	32m Cassegrain dish	0.6	42.5
Sondrestrom	Kangerlussuaq, Greenland	1290	0.23	3.50	3.0	65	32m parabolic dish	0.6	50
RISR	Resolute Bay, Canada	430-450	0.67	2.00	10.0	120	2x 30m x 30m dipole crossed phased array	1.0	43
PFISR	Poker Flat, Alaska	430-450	0.67	2.00	10.0	120	30m x 30m dipole crossed phased array	1.0	43
Jicamarca	Jicamarca, Peru	49.92	6.00	4.0-5.0	6.0	3000	300m x 300m phased array (18432 dipoles)	1.06	
Arecibo	Arecibo, Puerto Rico	430	0.70	2.50	6.0	75	305m spherical reflector	0.17-0.25	~60
Millstone Hill	Massachussets, USA	440	0.68	2.50	6.0	120	68m zenithal and 46m steerable dish	0.6	45
Kharkiv	Kharkiv, Ukraine	158	1.90	2.6/ 3.6	6.0	100-250	100m zenithal and 25m steerable dish	1.3/5.1	
Irkutsk	Irkutsk, Siberia	154-162	1.85-1.95	3.20		400	246m x 12m sectorial horn	0.5	35
MU	Shigaraki, Japan	46.5	6.45	1.00	4.0		Circular array of 103m diameter (475 yagis)	3.6	
EAR	West Sumatra, Indonesia	47	6.38	1.00	5.0	627	100-300m across phased array (560 yagi antennas)	3.4	33
ALTAIR	Kwajalein, Marshall Islands	155/ 415	1.94/ 0.72	6.00	1.5/ 5	992	150-ft (45.72m) Steerable dish	2.8/1.09	34.7
Qujing	Qujing, China	500	0.6	2	5		29m Steerable parabolic dish	1.6	42
PANSY	Syowa Station, Antartic	47	6.38	0.52			160m diam. phased array (1045 yagi antennas)	~4	

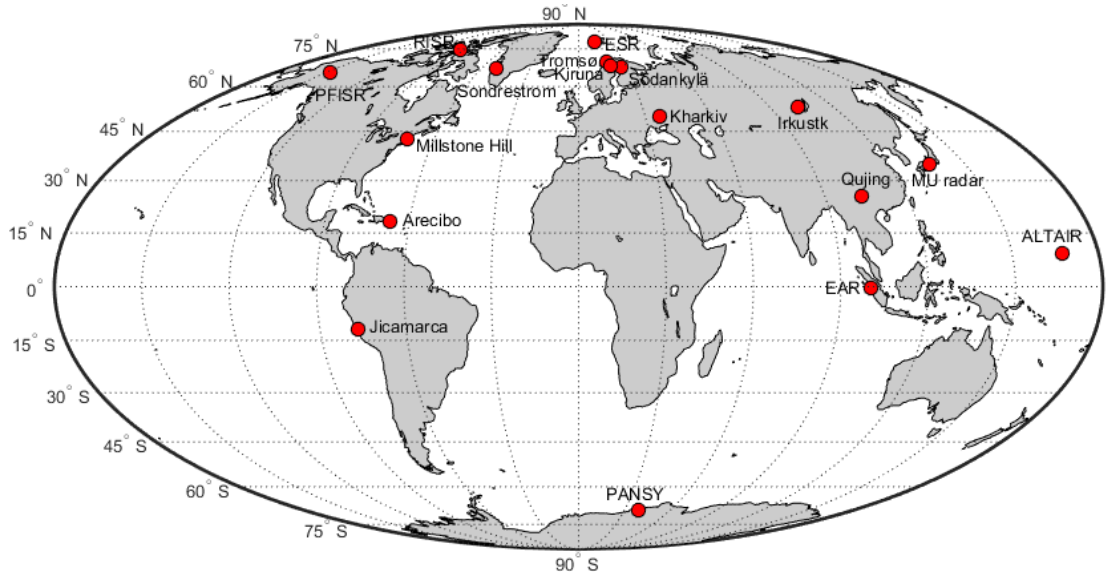

Figure 2.1. Location of World Major Incoherent Scatter Radar Facilities

2.3. ISR MEASUREMENTS

Signals received by the ISR are the backscatter of radio pulses generated by the Thomson Scatter effect on ionospheric electrons (a more detailed explanation of the signals received can be seen in Annex 1. The Incoherent Scatter Spectrum). The ionospheric backscatter ($z(t)$) is a random process that provides information of the ions and electrons of the ionosphere. There are a large number of non-stationary electrons contained in the scattering volume that contribute with a backscatter, generating a Gaussian stochastic process according to the Central Limit Theorem [Lehtinen and Huuskonen, 1996]. These backscattered contributions are governed by electromagnetic forces in the plasma that cause some regularity in the plasma fluctuations, generating the Incoherent Scatter Spectra (ISS). At a given range and during a finite time, this process is assumed Wide Sense Stationary (WSS) in the sense that its first and second order moments (i.e. mean and covariance) are constant [Nikoukar et al., 2012].

Embedded with the random backscatter signal there is a contribution of noise ($n(t)$) that comes from background electromagnetic radiations (i.e. sky noise) and from the internal receiver system noise (i.e. thermal noise) [Lehtinen and Huuskonen, 1996]. The bandwidth of this noise is much broader than the receiver bandwidth, and consequently it is commonly assumed a white noise filtered by the receiver impulse response. It is traditionally considered Gaussian, stationary, and independent from the radar signal [Nikoukar et al., 2012]. This noise is complex (i.e. with real and imaginary contributions) and it is modeled as proper complex normal random variable [Vierinen, 2012]. The filtered and discretized measurement errors are assumed identically distributed proper complex Gaussian normal random variables and independent.

Since both random processes are assumed to have Gaussian distributions with zero mean, these processes can be described by the Auto-Correlation function (ACF) of the lagged products of the signal with noise ($z(t) = v(t) + n(t)$) [Lehtinen and Huuskonen, 1996]:

$$ACF(\tau) = \langle z(t)z^*(t')\rangle = \langle \big(v(t) + n(t)\big)\big(v(t') + n(t')\big)^* \rangle \quad (1)$$

where τ is the lag time ($\tau = t - t'$), and the operator $\langle\,.\,\rangle$ represents the expected value.

Assuming both random processes independent, the ACF of uncorrelated stationary processes is the sum of two independent ACFs [Lehtinen and Huuskonen, 1996]:

$$ACF(\tau) = \langle v(t)v^*(t')\rangle + \langle n(t)n^*(t')\rangle \quad (2)$$

It is not possible to obtain an estimate of the ACF using only a single pulse transmission [Nikoukar et al., 2008]. Due to the random nature of the plasma backscatter, to compute an accurate statistical estimate of the real plasma ACF it is required to average signals obtained from many pulse transmissions [Farley, 1969]. This process is known as "integration". The "integration period" of the experiment is defined as the time span over which the N-pulse average is performed [Erickson, 1998], and the number of transmitted trains of pulses N is defined as the "integration length" [Kudeki, 2010].

Assuming a stationary process (i.e. plasma parameters varying slowly in the scattering radar volume), the statistical accuracy of the estimate will improve as much as the integration lenght is increased [Farley, 1969]. Nevertheless, for long integration periods this assumption can smooth and hide the dynamic phenomenon occurring in the ionospheric plasma. The increase of accuracy of the ACF estimate with the integration length is illustrated in Figure 2.2, where an ISS obtained using Particle In-Cell (PIC) simulations is reconstructed with different number of integrations. The results shown in Figure 2.2 were simulated without sky nor thermal noises.

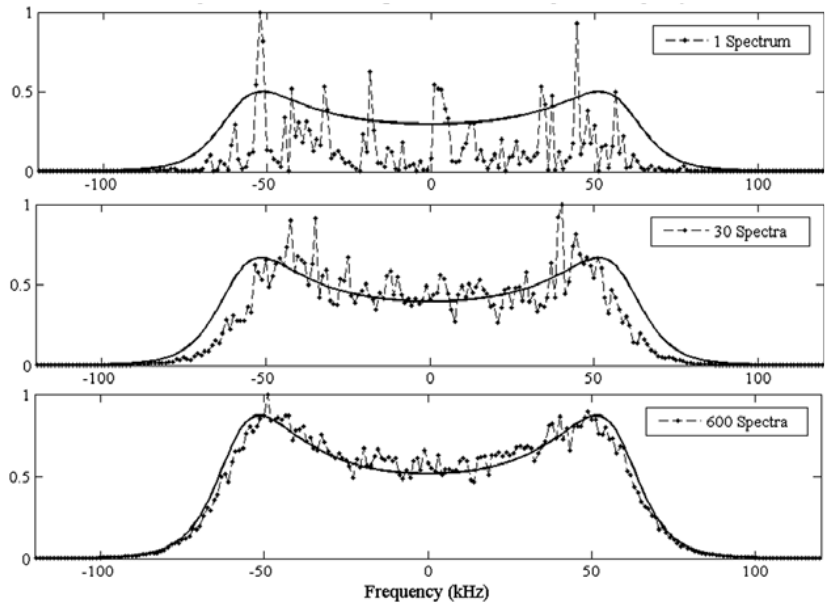

Figure 2.2. Incoherent Scatter Spectrum (ISS) for different number of independent spectra integrated, from [Diaz et al., 2008].

The ISR signal fluctuation (δ) is a measure of the variability of the estimated signal. It is equivalent to the Root Mean Square (RMS) error of the ACF estimate (i.e. standard deviation in unbiased estimators). This RMS error can be obtained as [Farley, 1969] [Sulzer, 1986a]:

$$\delta = RMS = \frac{(1 + SNR^{-1})}{\sqrt{N}} \tag{3}$$

The signal fluctuation (δ) is inversely proportional to the square root of the integration length ($\sqrt{N}$) and the Signal-to-Noise Ratio (SNR) of the backscattered signal received at the antenna [Farley, 1969] [Mathews et al., 1982]. The SNR value at the antenna can be directly calculated as indicated in Chapter 2.5. Received Power.

An increase of the number of integrations would reduce the estimate RMS representation error. Alternatively, increasing the SNR to values higher than one ($SNR > 1$) will not generate a significant reduction of the estimate errors [Farley, 1969] [Sulzer, 1986a]. In such high SNR cases, the signal power can be distributed into multiple frequency transmissions to increase the number of simultaneous samples obtained, increasing the integration length (N) [Sulzer, 1986a].

2.4. ISR ANALYSIS METHODS

There exist two different methods to analyze the ISR signal and estimate plasma parameters (i.e. inversion process): the "Range-Gate analysis" and the "Full Profile analysis",

The "Range-Gate analysis" is the standard technique to estimate the plasma parameters from the ISR signal backscatter. It is also called "Heigh-by-Height analysis" or "Gated analysis" because each height is processed separately [Hysell et al, 2008] [Nikoukar, 2010]. The range spacing of this method is commonly defined by the pulse length [Nikoukar et al., 2008]. This method assumes that the ionospheric plasma has a slow spatial variation over each range-gate [Lehtinen et al., 1997]. Summation rules are applied to compensate range smearing effects at each range-gate when using Long Pulse transmission codes [Holt et al., 1992] [Swoboda et al., 2017]. Finally, plasma parameters are estimated from the measured signal using a Non-Linear Least-Squares (NLLS) optimization algorithm, as indicated in Annex 2. Standard ISR Estimation Technique. This individual range fitting is done in the lag domain, comparing ACF values, or in the frequency domain, comparing ISS values (i.e. the Fourier transform of the ACF). This technique suffers from the assumption of constant plasma parameters in the study volume during measurements and a coarse resolution of the estimated parameters [Nikoukar et al., 2008].

Alternatively, the "Full Profile analysis" technique estimates simultaneously the complete ionospheric profile [Hysell et al, 2008]. This analysis is performed using the Inverse Problem theory over the entire column vector of lag profile estimates. In this case, the inversion theory considers the estimate sufficiently close to the real solution to assume a linearized ISR model. In this technique the plasma parameters are modelled using B-spline [Holt et al., 1992] or Lagrangian [Lehtinen et al., 1996] interpolations with user-defined range resolution of the parameters. Then, the ACF or ISS estimates are fitted to the theoretical ISR model using two-dimensional ambiguity functions [Virtanen et al., 2008]. The fitting is performed typically using the Levenberg-Marquardt (L-M) or Gauss-Newton (GN) methods.

The "Full Profile" analysis improves the accuracy of the estimates because uses the information completely and all corrections are included in the calculus (including filter corrections, plasma gradient corrections, sidelobe deconvolutions of short phase codes, ambiguity deconvolutions of ambiguous codes, spatial post-integration to specified resolutions, etc.) [Lehtinen, 1986]. This technique is optimal in the sense that uses all the available information to compute the plasma parameters, including the ambiguity function and the full error covariance matrix [Nikoukar et al., 2008].

Nevertheless, the "Full Profile" method is more sensitive to noise and interferences, and it is highly dependent on the initial guess of plasma variables to converge to the correct solution [Hysell et al., 2009]. The accuracy of this method is dependent on the resolution of the range grid of the parameters, and it has a very significant computational cost for a fine spatial resolution [Nikoukar et al., 2008]. Furthermore, solutions may not be unique or even become unstable, requiring the application of regularization techniques [Aster et al., 2012] that apply theoretical models and assumptions about the behavior of the ionospheric plasma that may be inappropriate in certain cases.

2.5. RECEIVED POWER

The backscatter power received (P_{rx}) from an ionospheric volume located at a distance r of the antenna can be calculated using the Soft Radar Equation [Evans, 1969] given by:

$$P_{rx} = \frac{P_{tx}\mathrm{T}Lc\lambda_r^2 N_e \sigma}{2(4\pi)^3 r^2} \int_{\phi}\int_{\theta} G_t^2(\theta)\sin(\theta)\, d\theta\, d\phi \tag{4}$$

where P_{tx} is the transmitted signal power, T is the transmitted pulse duration, L is a Loss Factor that takes into account internal ohmnic losses, λ_r is the radar wavelength ($\lambda_r = c/f_r$), G_t is the radar antenna gain, and σ is known as the Radar Cross-Section (RCS).

The RCS (σ) is defined as the cross-sectional area of a perfectly conducting sphere that produces the same amount of energy obtained by the measured object. In the ionospheric plasma, the volumetric RCS is dependent on the electron density variations ($\langle|n(k,\omega)|^2\rangle$). The RCS ($\sigma$) of a thermal unmagnetised plasma can be written as [Evans, 1969] [Yao et al., 2014b] [Lu et al., 2016]:

$$\sigma = \frac{\sigma_e}{(1+\alpha^2)(1+T_e/T_i+\alpha^2)} \tag{5}$$

where σ_e is the cross-section of each electron ($\sigma_e = 4\pi(r_e\sin(\theta))^2 \cong 10^{-28}(\sin^2(\theta))$, being θ the polarization angle of the scattered signal), and $\alpha = 4\pi\lambda_D/\lambda_r$ (being $\lambda_D = \sqrt{\epsilon_0 kT_e/e^2N_e}$ the plasma Debye length) [Evans, 1969].

For parabolic antennas in monostatic radars, the total power received is [Evans, 1969]:

$$P_{rx} = \left[0.76\frac{LcA_e\sigma_e}{16\pi}\right]\frac{P_{tx}\mathrm{T}}{\mathrm{r}^2}\frac{N_e}{(1+\alpha^2)(1+T_e/T_i+\alpha^2)} \tag{6}$$

where A_e is the antenna effective aperture that corresponds to the physical collecting area of an ideal antenna (computed as $A_e = G_t\lambda_r^2/4\pi$).

Alternatively, the received power noise level P_n is defined by the equivalent noise temperature T_n and the reception bandwidth B of the radar as [Lu et al., 2016]:

$$P_n = k_B T_n B \tag{7}$$

Consequently, the Signal-to-Noise Ratio (SNR) of a parabolic antenna radar can be computed as [Taran, 1988] [Yao et al., 2014b] [Lu et al., 2016]:

$$SNR = \frac{P_{rx}}{P_n} = \left[0.76\frac{LcA_e\sigma_e}{32\pi}\frac{1}{k_B T_n B}\right]\frac{2P_{tx}\mathrm{T}}{r^2}\frac{N_e}{(1+\alpha^2)(1+T_e/T_i+\alpha^2)} \tag{8}$$

The section in brackets of Equation (8) is known as the radar system constant (K_{sys}), and it is typically obtained during operation by a calibration of radar measurements [Lu et al., 2016].

2.6. PLASMA LINE

ISR backscatter signals are located at two different frequency bands (see Annex 1. The Incoherent Scatter Spectrum):

a) the Ion Acoustic band, with a kilohertz bandwidth and the strongest signal power, this frequency band is commonly used to estimate plasma parameters; and
b) the Plasma Line band [Akbari et al., 2017], which is a frequency peak located near the plasma frequency (f_p) that can be used to obtain an estimate of the electron density.

Both spectral bands provide complementary and simultaneous information of ionospheric parameters. The first detection of the ISR Plasma Line was done by [Perkins et al., 1965] at the Arecibo Observatory. Few years later, [Wand, 1970] was the first to compare the values of electron-to-ion temperature ratio (T_e/T_i) observed using the Plasma Line and the Ion-Acoustic Line simultaneously. Results of this later study demonstrated a very good agreement of the estimates obtained at both spectral lines. The study of [Wand, 1970] suggested the use of the information obtained from the Plasma Line to solve ambiguous estimation of ion composition in the Ion-Acoustic analysis.

The equation to obtain the Plasma Line frequency (f_{PL}) in presence of magnetic fields is [Yngvesson and Perkins, 1968]:

$$f_{PL}^2 = f_p^2 + \frac{12 k_B T_e}{\lambda_r^2 m_e} + f_c^2 \sin^2 \Theta \tag{9}$$

where λ_r is the wavelength of the radar, T_e is the electron temperature, f_p is the plasma frequency ($f_p = (2\pi)^{-1}\sqrt{N_e e^2/m_e \epsilon_0} \approx 8.97\sqrt{N_e(cm^{-3})}\ [kHz]$), f_c is the electron gyro frequency ($f_c = eB/2\pi m_e$), and Θ is the angle between the radar wavevector and the magnetic field.

Figure 2.3 shows an ISR spectrum highlighting the Ion Acoustic, Plasma, and Gyro lines (see [Janches and Nichols, 2007] for more information about Gyro Line).

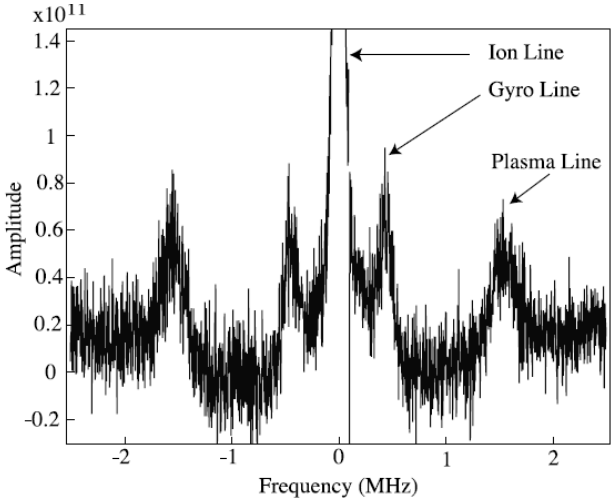

Figure 2.3. Full ISS showing the Ion, Plasma, and Gyro Lines. Measured at Arecibo at 275km, from [Aponte et al., 2007].

For a radar transmission frequency much smaller than 1 [GHz] (e.g. at the Arecibo Radio Observatory with $f_r = 430\ MHz$), it is possible to extract a highly accurate estimate of electron density from the Plasma Line resonance frequency, even when electron temperatures are not accurately known [Aponte et al., 2007]. The combination of the information of the electron density and the total power received at the antenna (P_{rx}) can provide highly precise estimates of the T_e/T_i temperature ratio. At these frequencies the value of RCS in Equation (5) can be approximated to $\sigma \approx \sigma_e/(1 + T_e/T_i)$. In this case, the T_e/T_i temperature ratio value can be inferred from the total received power (P_{rx}) of Equation (6) and the N_e parameter estimated from the Plasma Line as [Wand, 1970] [Waldteufel, 1971] [Aponte et al., 2007]:

$$\frac{T_e}{T_i} = \frac{K_{sys} G_t(r)}{r^2} \frac{N_e}{P_{rx}(r)} - 1 \tag{10}$$

where K_{sys} is the radar system constant that includes all fixed parameters of the radar experiment (as indicated in Chapter 2.5) and G_t is the antenna gain. In large antennas, the antenna gain depends on range ($G_t(r)$) at the altitudes of interest, and a precise antenna calibration is required to correctly estimate plasma parameters using this formulation [Aponte et al., 2007].

Alternatively, when the radar transmission frequency is sufficiently elevated (e.g. at EISCAT UHF radar with $f_r = 933\ MHz$), the Plasma Line resonance frequency depends also on the T_e parameter. In this case, it is possible to estimate unambiguously the ion composition by fitting together the Ion Acoustic Band and the resonance frequency [Bjørnå and Kirkwood, 1988].

A different method to determine the electron temperature from the Plasma Line has been implemented by measuring the asymmetries between the up- and down-shifted Plasma Line frequencies [Nicolls et al., 2006]. This shift depends on T_e, currents, photoelectrons, and other minor effects. In quiet ionospheric conditions, the difference between the Plasma Line frequencies (Δf_{PL}) can be approximated as [Nicolls et al., 2006]:

$$\Delta f_{PL} = f_{PL-} - f_{PL+} \approx \frac{f_r}{c}\left(4V_e - \frac{12 k_B T_e}{c m_e}\right) \tag{11}$$

where V_e is the line-of-sight electron drift velocity in [m/s].

Using Equation (11), and assuming equal electron and ion drift velocities ($V_e \approx V_i$), high-resolution measurements of electron temperature were obtained at Arecibo [Nicolls et al., 2006]. The values calculated using the frequency asymmetry agree within 5-10% of the temperatures obtained from the Ion Acoustic band, verifying the high accuracy of the method.

The study of [Bjørnå, 1989] has been able to unambiguously estimate the ion-neutral collision frequency of the plasma (ν) at lower ionospheric E regions by providing information of parameters estimated from the Plasma Line. Similar to the case of the ion composition and temperature ambiguity problem, there exist an estimation ambiguity related to the collision frequency and the plasma temperature in the ISR method.

2.7. TEMPERATURE-ION COMPOSITION AMBIGUITY

Different species of ions exist in the ionosphere at different altitudes. Figure 2.4 shows a simulation of the model IRI [Bilitza et al., 2017] of the most relevant ions located at altitudes between 100 km to 2000 km above the Arecibo Observatory. This figure shows the typical occupancy of the different ions depending on the altitude. Atomic oxygen ions (O^+) are dominant over ionospheric altitudes between 200 km to 1000 km, approximately. More weighty molecular ions (NO^+ and O_2^+) are predominant at the bottom section of the ionosphere, as demonstrated by early rocket and satellite measurements [Hoffman et al., 1969] [Evans & Oliver, 1972]. Finally, more lightly atomic ions (H^+ and He^+) are found at the upper ionospheric section. At high altitudes the ionosphere is mainly governed by atomic hydrogen ions (H^+).

There exist two transitions of major ion species at the altitudes measurable by the ISR: *a)* a transition at altitudes from ~130 km to 250-350 km changes between molecular ions (NO^+ and O_2^+) to atomic oxygen ions (O^+) [Wand, 1970] [Litvine et al., 1998]; and *b)* at altitudes higher than ~500 km atomic oxygen ions (O^+) are replaced by atomic hydrogen (H^+) and helium (He^+) ions. Typically, at this second transition the helium ion (He^+) percentage is lower than 15% and, consequently, almost negligible [González and Sulzer, 1996]. Nevertheless, the helium species is very sensitive to solar activity, reaching peach values of percentage of 50% depending on season and solar cycle level [Gonzalez et al., 2004].

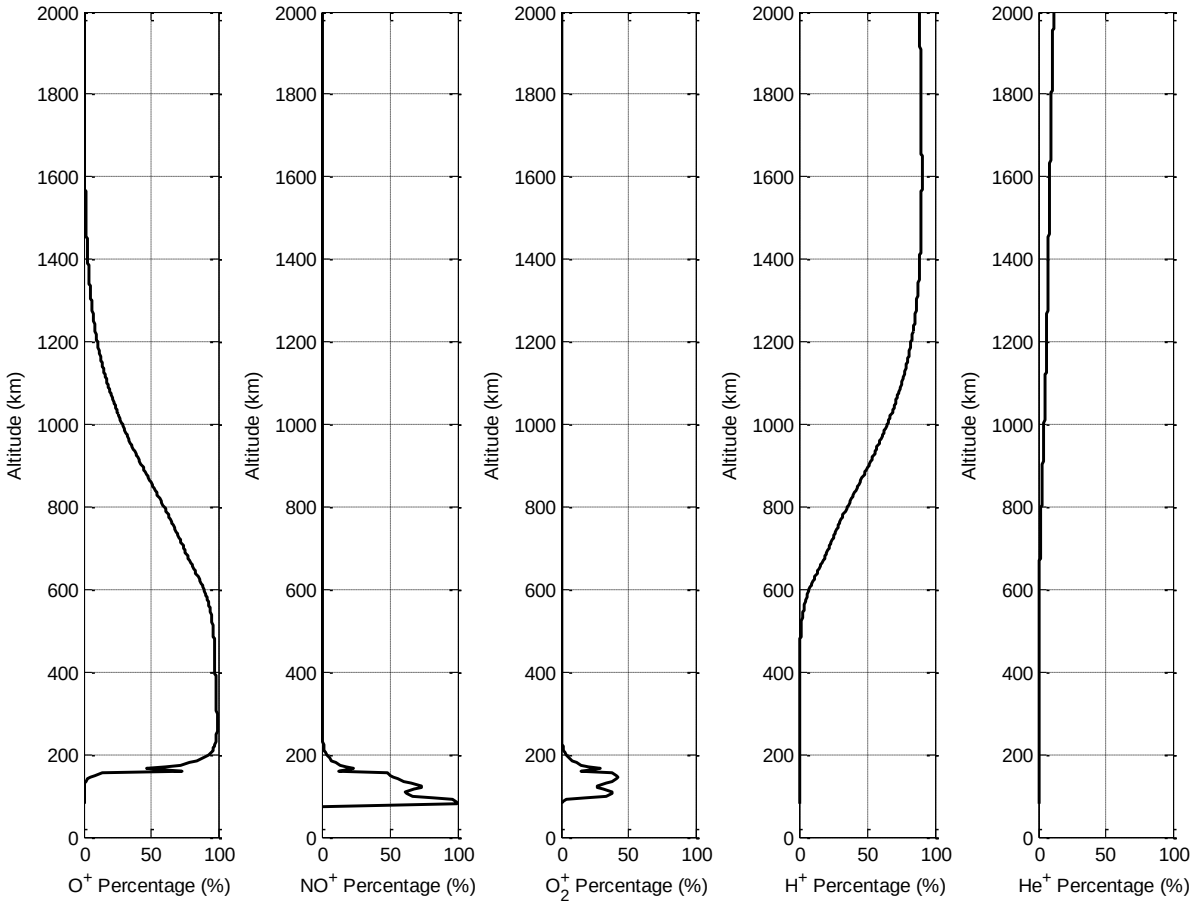

Figure 2.4. Most relevant ion composition percentages depending on altitude over Arecibo (model IRI of 8th March, 2005)

Estimating plasma parameters from the ISR signal is not always directly solvable and several problems arise from the analysis of the signal backscatter [Nikoukar et al., 2008]. Of relevance is the *Temperature-Ion Composition Ambiguity* (TICA) problem, which is generated by the mixture of different ions populating the ionosphere at lower altitudes [Oliver, 1979]. This ambiguity is related to the dependency of the theoretical ISR spectrum on the thermal speed of different ion species (Equation (30) of Annex 1. The Incoherent Scatter Spectrum), which is, in turn, proportional to $\sqrt{T_i/m_i}$ [Evans and Oliver, 1972] [Oliver, 1979] [Vallinkoski, 1988].

As the masses of molecular ions NO^+ and O_2^+ are almost identical, their final effects are indistinguishable by the ISR [Oliver, 1979]. It is not possible to know the exact mixture of these ions. Molecular ions are assumed to behave as a single molecular specie (M^+) with mass 30.5 amu for a mixture of 25% O_2^+ and 75% NO^+ [Lathuillere et al., 1983] [Cabrit & Kofman, 1996].

At the lower ionic transition of the ionosphere, the estimation of the relative abundance of atomic oxygen (O^+) and molecular (M^+) ions is ambiguous. This ion determination ambiguity is because those species have a mass ratio between them of approximately $m_{M^+}/m_{O^+} \approx 2$ [Oliver, 1979]. Therefore, combinations of temperatures and abundances of molecular (M^+) and atomic (O^+) ion species can obtain almost identical Ion Acoustic spectra, hindering the correct estimation of signals [Oliver, 1979] [Aponte et al., 2007]. Two different possible solutions of estimates of plasma parameters are obtained ambiguously [Oliver, 1979] [Lathuillere et al., 1983]. To determine unambiguously the ionospheric plasma parameters, additional information is then required [Aponte et al., 2007].

Figure 2.5 shows two examples of the TICA problem, representing similar Incoherent Scatter Spectrums obtained with two different combinations of parameters.

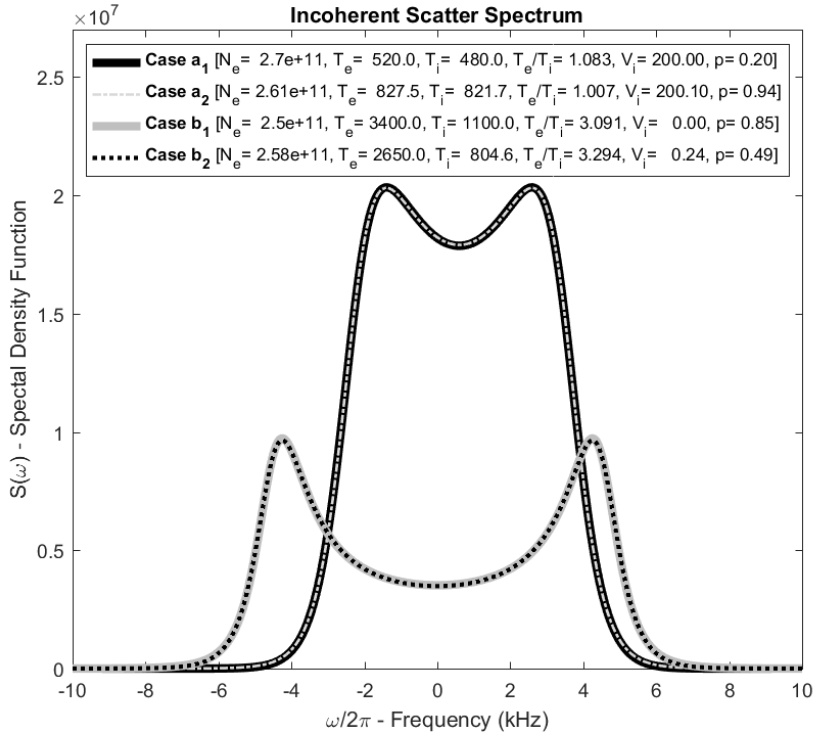

Figure 2.5. Incoherent Scatter Spectrum for 4 different parameters showing the ion composition ambiguity.

The ion composition (p) is defined as the molecular ion fraction, computed as the relative abundance of molecular ions ($n(M^+)$) with respect to the total ion concentration (assuming charge neutrality, $N_e = N_i$) [Lathuillere et al., 1983] [Aponte et al., 2007] [Wu et al., 2015]:

$$p = \frac{n(M^+)}{N_e} \tag{12}$$

Two different values of ion composition are obtained as possible solutions located approximately symmetric with respect to 0.5 [Lathuillere et al., 1983] [Wu et al., 2015].

This ambiguity problem has been widely studied in the literature and different methods have been proposed to solve it. Early methods were based on theoretical models of the behavior of ionospheric plasma parameters that assumed equal ion and neutral temperature profiles or considered a smooth altitude variation of parameters [Waldteufel, 1971] [Evans & Oliver, 1972]. Models first used at high latitudes assumed constant electron temperatures to cope with the ion temperature and composition changes generated by Joule heating [Kelly & Wickwar, 1981]. A widely used parametric model of ion composition and temperature was developed based on rocket and satellite measurements to solve this ambiguity [Oliver, 1979]. Models of ion composition have been also created to unambiguously apply the Full Profile inversion method [Cabrit & Kofman, 1996] [Litvine et al., 1998]. Different models have also been developed to estimate parameters of plasma disturbances generated by strong electric fields in the auroral region [Blelly et al., 2010] [Zettergren et al., 2011].

An alternative to use theoretical or empirical models is to use complementary information extracted from the Plasma Line frequency band and from the total power received at the antenna [Wand, 1970] [Bjørnå & Kirkwood, 1988] [Nicolls et al., 2006] (see Chapter 2.6. Plasma Line). The addition of *a priori* information constrains the feasible solutions, allowing an unambiguous ion composition estimate [Waldteufel, 1971] [Bjørnå & Kirkwood, 1988] [Aponte et al., 2007].

Another method for unambiguous estimation of ISR parameters was proposed in [Oliver, 1979] by the analysis of signals with very small fluctuations. Two possible solutions of ion composition can be obtained from the estimation of almost noiseless signals. The correct ion composition parameter can be determined by selecting the solutions that provide the smoother ion composition profile at different altitudes. Initial tests of this method made by [Oliver, 1979] obtained only partial success mainly because of a lack of signals almost without variability. The studies of [Lathuillere et al., 1983] and [Lathuillere & Pibaret, 1992] successfully demonstrated the efficacy of this method with experimental data. These studies reduced the fluctuation of signals by post-integrating multiple radar pulses with integration times of 5 minutes.

Recently, the study of [Wu et al., 2015] used a new optimization algorithm to unambiguously estimate plasma parameters of ISR signals with very small fluctuations. In their work, signals with very high SNR values, in the range of 15 to 25, were obtained at the Arecibo Radio Observatory using simultaneous frequency transmission [Sulzer, 1986a] and Coded Long Pulse [Sulzer, 1986b] techniques. This study used the Particle Swarm Optimization (PSO) algorithm [Kennedy & Eberhart, 1995] to analyze ISS data with the addition of *a priori* information from the Plasma Line, obtaining much better estimations than the standard NLLS algorithm commonly used in ISRs.

CHAPTER 3

METHODOLOGY

3.1. SIMULATION METHODS

3.1.1 IMPLEMENTATION OF THE INCOHERENT SCATTER RADAR MODEL

To perform the estimation process, an ISR spectrum model was implemented based on the formulation of [Kudeki & Milla, 2011] (shown in Annex 1. The Incoherent Scatter Spectrum). To simplify the required calculations, the magnetic field aspect angle was not included in the used model, which limits the validity to magnetic aspect angles greater than 5° [Longley et al., 2018]. Ion-neutral collisions were not included in the model because they are only relevant at the lower E region of the ionosphere [Bjørnå, 1989]. Therefore, this model represents the backscatter signal received at altitudes above 120km by radars not pointing perpendicular to the magnetic fields. Plasma parameters were electron density (N_e), electron temperature (T_e), ion temperature (T_i), ion drift radial velocity (V_i), and ion composition (p).

To improve computational speed and accuracy of simulations, the Gordeyeve integral (Equation (29)) was computed using the Imaginary Error Function (ERFI) implemented in the Faddeyeva Dawson function library of [Johnson, 2012]. A radar frequency of 450 MHz was considered in the simulation because many observatories use similar frequencies (i.e. Arecibo, Millstone Hill, AMISR, and ESR).

To avoid additional ACF conversion times at fitting comparisons, the estimation analysis was done by directly comparing the difference between simulated and theoretical Incoherent Scatter Spectra (ISS) at each range, similarly as in [Wu et al., 2015]. Theoretical spectrums were obtained at the Ion Acoustic frequency band, having frequencies between $\pm$10 kHz.

3.1.2. NOISE ADDITION SCHEME

To consider realistic radar conditions, simulated ISR signals were calculated as the theoretical backscatter signal plus a random noise (as indicated in Chapter 2.3. ISR Measurements). According to the Central Limit Theorem, the integration of multiple backscatter signals provides it Gaussian characteristics [Vallinkoski, 1988]. Therefore, the added noise was an Additive Gaussian White Noise (AWGN) [Lehtinen & Huuskonen, 1996].

A Gaussian random noise with zero mean was added directly to the ISR spectrum signals. The standard deviation (σ) of this noise was determined by the signal fluctuation percentage ($\delta(\%)$) assumed to be given by the RMS error of the ACF estimator (shown in Equation (3)) [Farley, 1969]. This standard deviation is calculated relative to the maximum absolute value of the spectrum amplitude ($\sigma = \delta(\%)/100 \cdot \max(|f(\boldsymbol{x})|)$), similarly as in [Aponte et al., 2007].

Noise contributions were assumed to be independent from the plasma parameters. This assumption requires the reception of signals with $SNR \ll 1$ [Huuskonen & Lehtinen, 1996]. Achievable signal fluctuation percentages ($\delta(\%)$) under this weak SNR condition are shown in Figure 3.1 for different integration lengths and periods. The signal fluctuation values shown in this figure were calculated using Equation (3) and an Inter-Pulse Period (IPP) of 1 millisecond.

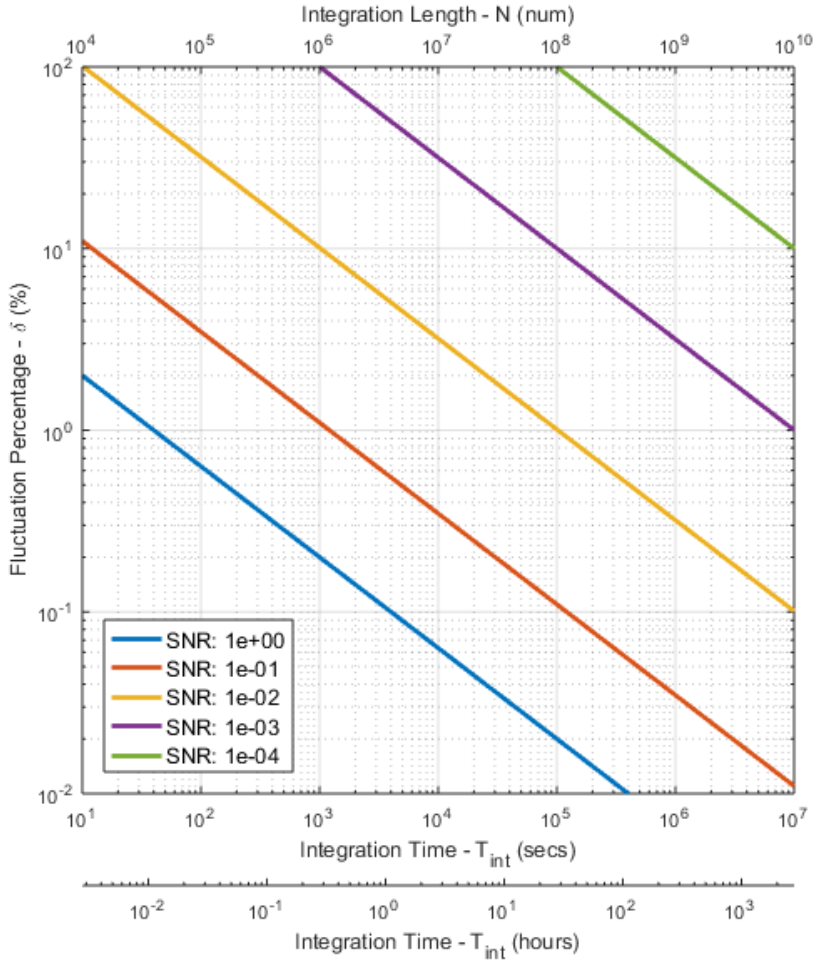

Figure 3.1. Theoretical fluctuation percentage ($\delta(\%)$) in logarithmic scale obtained at different integration times (in seconds and hours) for different signal SNR values. Upper horizontal axis indicates the required integration length.

Recent ISR simulators [Swoboda et al., 2017] implement the effect of the random backscatter contribution of electrons and the range smearing (i.e. the Range-Lag ambiguity function [Lehtinen & Huuskonen, 1996]). To reduce simulation complexity and computing time requirements, these effects were not considered in the present work. Therefore, in our study weak SNR and no range ambiguity are assumed. These same characteristics have been assumed in recent studies of the TICA problem [Aponte et al., 2007] [Wu et al., 2015] and are commonly obtained in Multi-Pulse or Alternating Code experiments [Vallinkoski, 1988]. Therefore, simulated spectra were considered to be the Fourier transform of the final ACF estimate calculated by the radar processing chain, obtained by subtracting range smearing effects (i.e. using summation rules [Holt et al., 1992]) and noise correlation estimates, following the standard Range-Gate analysis criterion explained in [Swoboda et al., 2017].

3.1.3. PLASMA PARAMETERS ESTIMATION METHOD

In this thesis work, we consider the estimation of plasma parameters using the Range-Gate analysis method (see Chapter 2.4. ISR Analysis Methods). Although the Full Profile method provides an optimum solution because it uses measurements of the entire ionospheric column, this method requires the use of theoretical models and assumptions about plasma parameters that may be inaccurate in certain cases. Furthermore, the Range-Gate analysis has been previously used in recent studies of the ISR estimation ambiguity [Aponte et al., 2007] [Wu et al., 2015].

The estimation method implemented in this work is the standard ISR estimation process indicated in Annex 2. Standard ISR Estimation Technique. This method assumes theoretical signals with Gaussian noise added and no correlation between measurement errors (i.e. weak SNR conditions). The Maximum Likelihood Estimator (MLE), which in this case is equivalent to the Least Squares Estimator (LSE), corresponds to the minimization of the Chi-Square cost function χ^2 (shown in Equation (38) of Annex 2. Standard ISR Estimation Technique).

Unbiased estimates of signals with AWGN characteristics and with known variances (σ_i^2) obtain a Chi-Squared (χ^2) cost function value approximately equal to the number of Degrees of Freedom (DoF) of the estimation problem [Taylor, 1997] [Bevington & Robinson, 2003]. To normalize the resulting cost function value, the Reduced Chi-Square (χ_r^2) cost function criteria has been used in this study [Taylor, 1997] [Bevington & Robinson, 2003]:

$$\chi_r^2 = \frac{\chi^2}{DoF} \tag{13}$$

The DoF is typically assumed to be $DoF = M - P$, where M is the input signal vector length, and P is the number of parameters of the estimation problem. This assumption is commonly used in standard error analysis studies [Taylor, 1997] [Bevington & Robinson, 2003], although it is known that this is an approximation only valid for linear models with linearly independent basis functions [Andrae et al., 2010].

Due to the randomness of signals to estimate, the χ_r^2 values have a Probability Density Function (PDF) that corresponds to a Chi-Squared (f_{χ^2}) statistical distribution, dependent on the DoF value [Bevington & Robinson, 2003] [Andrae et al., 2010]:

$$f_{\chi^2}(x, DoF) = \frac{x^{\left(\frac{DoF-2}{2}\right)} e^{-\left(\frac{x}{2}\right)}}{2^{\left(\frac{DoF}{2}\right)} \Gamma\left(\frac{DoF}{2}\right)} \tag{14}$$

where $\Gamma(z)$ is the Gamma function ($\Gamma(z) = \int_0^\infty t^{(z-1)} e^{-t} dt$).

The Gaussian distribution that best approximates the Chi-Squared statistical distribution (f_{χ^2}) has a mean $\mu_r = 1$ and a standard deviation $\sigma_r = \sqrt{2/\mathrm{DoF}}$ [Andrae et al., 2010]. Figure 3.2 shows the Chi-Squared statistical distribution and the Gaussian distribution that best approximates for different values of DoF.

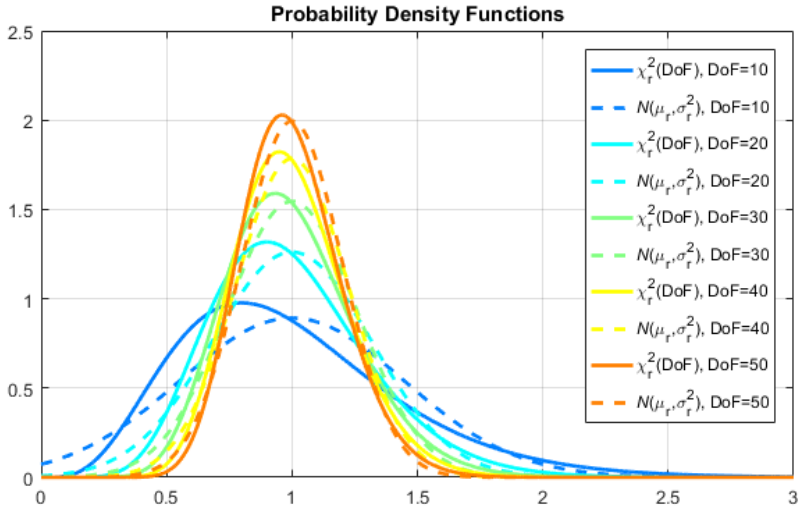

Figure 3.2. Probability density functions of the Chi-Square distribution and Gaussian approximation for different DoF values.

Figure 3.3 shows an example of the χ_r^2 cost function for different projections of electron temperature (T_e), ion temperature (T_i), and ion composition (p) parameters, assuming that electron density (N_e) and ion velocity (V_i) are parameters known *a priori*. These images were generated selecting the minimum χ_r^2 value of the non-represented parameter, allowing a 2D representation of a 3D structure. These simulations were done for different fluctuation levels to represent the impact of the increase of signal variability in the estimation performance. In scenarios with almost no fluctuation ($\delta = 0.01\%$), Figure 3.3 (Top) shows one of the two χ_r^2 cost function minimums with a much smaller value, indicating that it is the global minimum (i.e. the 'correct' solution). In this case, it is possible to determine the 'correct' set of parameters by selecting the solution with the lowest χ_r^2 value. Alternatively, Figure 3.3 (Medium and Bottom) demonstrate that it is more difficult to discriminate the 'correct' solution by selecting the lowest χ_r^2 value in more noisy scenarios ($\delta = 1\%$ and 10%, respectively).

The existence of two minimums in Figure 3.3 verifies the presence of the TICA problem. These two minimums are located at almost opposite ion composition values, symmetric with respect to 0.5, as initially indicated by [Lathuillere et al., 1983] and more recently verified by [Wu et al., 2015]. As the 'incorrect' ion composition solution ($p_{incorrect}$) is an almost symmetric solution of the 'correct' solution ($p_{correct}$), those solutions are associated as $p_{incorrect} \approx 1 - p_{correct}$. Consequently, the 'incorrect' solution is related to the abundance of atomic oxygen ions ($p_{incorrect} \approx n(O^+)/N_e$) (see Equation (12)). These two minimums of the cost function are obtained because the ISR spectrum has a dependency on the ratio T_i/m_i. Therefore, as molecular ion mass (m_{M^+}) is almost double that of atomic ions (m_{O^+}), almost identical results would be obtained with molecular ion temperatures which also double those of atomic temperatures ($T_{M^+} \approx 2T_{O^+}$), as indicated by [Oliver 1979].

Incorrectly calibrating the radar constant (K_{sys}) or wrongly estimating the equivalent noise temperature (T_n) would obtain an incorrectly estimated *SNR* value, as indicated in Chapter 2.5. The use of an invalid SNR estimate would generate wrong estimates of signal fluctuation (δ), as seen in Equation (3). Therefore, these misconfigurations would generate the use of incorrect measurement error variances (σ_i^2) in the χ_r^2 cost function criteria, affecting the correct estimation of plasma parameters. The effect of such misconfigurations should be analyzed in a future study.

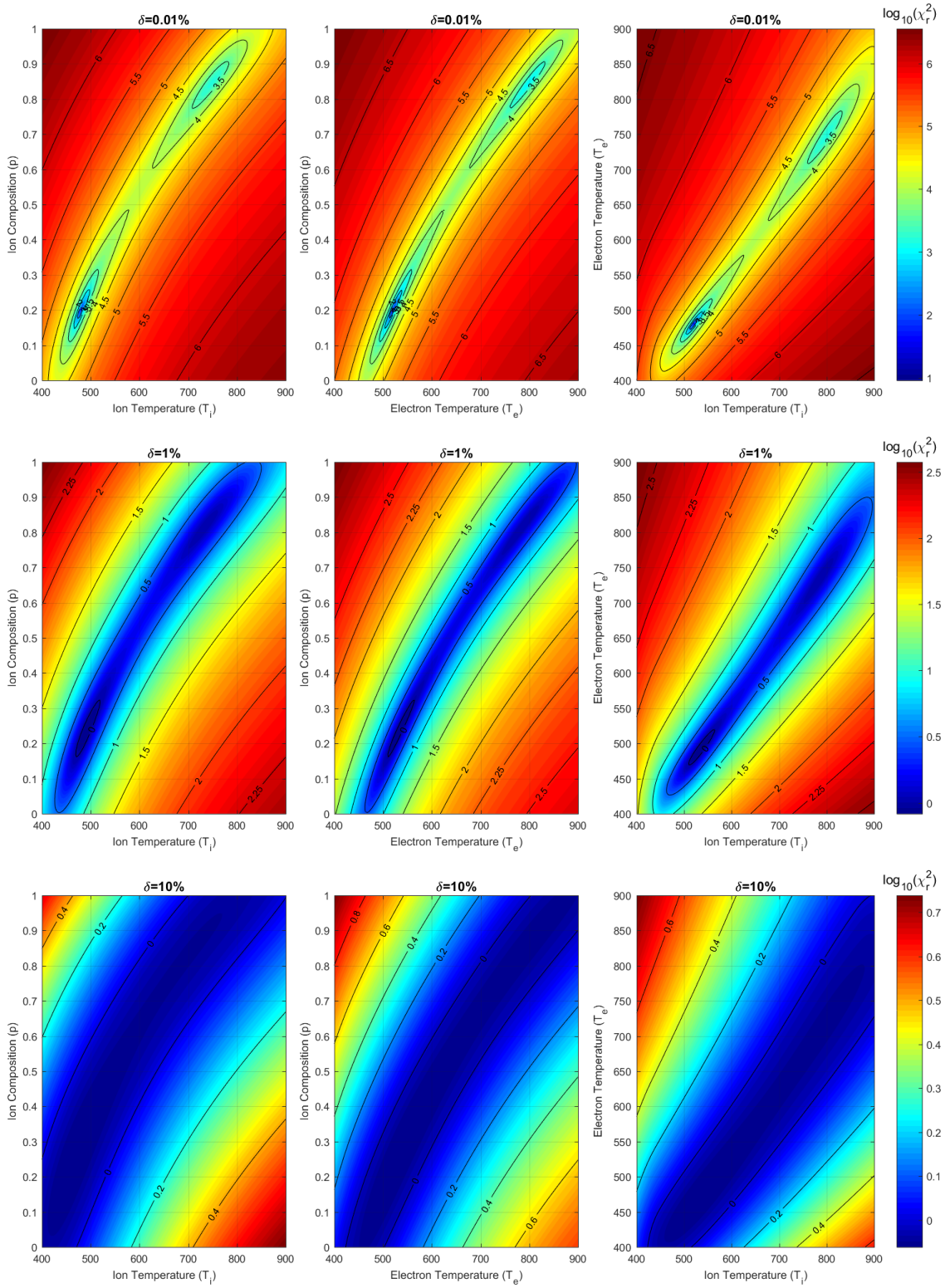

Figure 3.3. Graphic representation of the χ_r^2 cost function minimum value (in logarithmic scale) for different combinations of plasma parameters at different fluctuation levels (Top $\delta = 0.01\%$, Middle $\delta = 1\%$, and Bottom $\delta = 10\%$). Data has been obtained from the simulation of plasma parameters of spectrum Case a_1 of Figure 2.5 (T_e=520 ºK, T_i=480 ºK, and p=0.2), assuming N_e known a priori, and consequently not represented.

3.1.4. NON-LINEAR LEAST SQUARES OPTIMIZATION ALGORITHM USED

The minimization of the χ_r^2 cost function is performed by the most commonly used NLLS optimization algorithm of ISR analyses, both for Range-Gate [Erickson, 1998] [Swoboda et al., 2017] and for Full Profile [Hysell et al., 2008] [Nikoukar et al., 2008] analyses: the Levenberg-Marquardt (L-M) algorithm [Levenberg, 1944] [Marquardt, 1963]. This algorithm is explained in Annex 3. The Levenberg-Marquardt Optimization Algorithm.

The L-M optimization algorithm suffers from the local minimum determination problem. When several minimums of the cost function exist, iterative descent search algorithms may provide suboptimal results when selecting the first minimum found during the search [Wu et al., 2015]. The determination of the solution is subject to the initial search parameters of the algorithm, the parameter search range, and the algorithm configuration [Gavin, 2017]. To reduce the effect of this drawback, other optimization algorithms have been studied in the ISR literature such as the Trust-Region algorithm [Milla et al., 2013] or the Particle Swarm Optimization (PSO) algorithm [Wu et al., 2015]. These algorithms improve some characteristics of the estimation but also increase the computation time requirements.

Initial estimation tests were done with the Simulated Annealing (SA) optimization algorithm. The SA algorithm is an iterative probabilistic technique [Kirkpatrick et al., 1983] based on the Metropolis algorithm. The SA algorithm implementation was coded based on the library of [Vandekerckhove, 2008]. No previous study in the literature has analyzed the SA algorithm to estimate IS parameters. Nevertheless, the study of [Erickson, 1998] indicated that this algorithm should be computationally expensive to evaluate and consequently impractical to use in ISR estimation. The estimation tests were done considering the ion composition parameter *a priori* known and no random noise was added to the theoretical ISR signal. These simulation characteristics were considered to verify the capability of this algorithm to estimate the most basic ISR signals. Results of these estimation tests obtained computing times approximately 2000 times larger than the computing times required by the L-M algorithm in identical conditions. Furthermore, the convergence of the SA algorithm was much lower than the L-M algorithm in identical conditions, with convergence percentages of 87.48% and 99.5%, respectively. These results indicate that the SA algorithm cannot be used as a routinely algorithm for estimating ISR parameters, as previously indicated by [Erickson, 1998].

Despite PSO provided much better performance than the standard NLLS algorithm in [Wu et al., 2015], in their study they considered only signals with a very small signal fluctuation. Therefore, the results obtained by [Wu et al., 2015] may be related to the use of a particularly small signal fluctuation level. To verify the effectiveness of using PSO it is required to compare it against a standard Non-Linear Least Squares (NLLS) algorithm at different signal fluctuations. The determination of the improvement obtained by PSO requires the previous knowledge of the performance at different signal fluctuations of the L-M algorithm, which is the main focus of this thesis. The PSO algorithm performance is currently being analyzed and will be presented in a future study.

3.1.5. ADDITION OF PLASMA LINE INFORMATION

To unambiguously estimate plasma parameters, several methods that provide information measured from the Plasma Line and from the total power received by the radar have been implemented in the literature (as indicated in Chapter 2.7. Temperature-Ion Composition Ambiguity). The Plasma Line band provides complementary and simultaneous information of some ionospheric plasma parameters [Wand, 1970].

In our simulations, we quantified the effect of providing information from the Plasma line. To account for realistic cases, different combinations of plasma parameters were assumed known *a priori* in the estimation process by fixing these values during the optimization algorithm search. Similar fittings were made by [Wu et al., 2015] for different combinations of *i*-unknown parameters (*i*=1, 2, and 3). The combinations of parameters assumed known in our study were:

a) no *a priori* information;
b) electron density (N_e);
c) electron-to-ion temperature ratio (T_e/T_i) and electron density (N_e); and
d) electron temperature (T_e) and electron density (N_e).

The *a priori* knowledge of N_e and T_e/T_i parameters was assumed to be extracted from the Plasma Line frequency and from the total power received from the antenna, respectively, following previous studies [Waldteufel, 1971] [Aponte et al., 2007] [Wu et al., 2015]. Alternatively, the *a priori* information of T_e was assumed to be extracted from the Plasma Line resonance frequency [Bjørnå & Kirkwood, 1988] or from its asymmetry [Nicolls et al., 2006].

All *a priori* known plasma parameters were initially considered without uncertainty, assuming a perfect determination of plasma parameters, as in [Wu et al., 2015]. Although this assumption is unrealistic, the estimation using deterministic *a priori* information provides the *best case* estimate that can be obtained. The analysis of information with uncertainty in the *a priori* known parameters was studied separately. Simulations were done as indicated in Chapter 3.1.9 and results of those simulations are shown in Chapter 7. Determination of the Impact of the Uncertainty of *a-priori* Known Parameters.

3.1.6. MONTE CARLO SIMULATIONS OF PLASMA PARAMETERS

To quantify the effect of the ambiguous estimation of plasma parameters from ISR signals in realistic conditions, Monte Carlo simulations of the estimation of plasma parameters were performed. These simulations were done either for 1000 or 2000 (N_{MC}) different true input parameters ($\boldsymbol{x}_{true}$) uniformly selected between the following parameter ranges:

- Electron density (N_e): $10^9 \leq N_e \leq 10^{12}\ m^{-3}$;
- Electron temperature (T_e): $300 \leq T_e \leq 5000\ ºK$;
- Ion temperature (T_i): $300 \leq T_i \leq 3000\ ºK$;
- Electron-to-ion temperature ratio (T_e/T_i): $0.1 \leq T_e/T_i \leq 5$;
- Ion drift velocity (V_i): $-250 \leq V_i \leq 250$ m/s; and
- Ion composition (p): $0 \leq p \leq 1$.

The ranges of N_e, T_e, and T_i parameters selected resemble the maximum and minimum values of measurements obtained by SROSS-C2 satellite and models above the ionospheric F2 layer peak at low-latitude regions with different solar conditions [Sharma et al., 2016]. The range of electron-to-ion temperature ratios (T_e/T_i) and ion drift velocities (V_i) used were larger than the typical measurements obtained at standard mid-latitude ionospheric conditions shown in [Wand, 1970] [Scherliess et al., 2001].

In this work, the ion drift velocity parameter (V_i) was assumed to be known *a priori*. As it generates a shift of the spectrum center due to the Doppler effect, it does not affect the determination of other parameters and it can be estimated independently [Wu et al., 2015]. Results of simulations shown in Chapter 6 verify the ion drift estimation independence.

To study the effect of increasing the variability of ISR signals in the estimation process, Monte Carlo simulations were done at different signal fluctuation levels. First, different ISS signals were created using the uniformly selected true input plasma parameters ($\boldsymbol{x}_{true}$) and the theoretical ISR spectrum model explained in Annex 1. The Incoherent Scatter Spectrum. The noise contribution scheme indicated in Chapter 3.1.2 was applied to these ISS signals, adding a Gaussian random vector with zero mean and a standard deviation defined by a selected signal fluctuation level ($\delta(\%)$). The same set of uniformly selected true input plasma parameters ($\boldsymbol{x}_{true}$) was used in all simulations, allowing a direct comparison of results for different fluctuation levels. Finally, the plasma parameters were estimated using the L-M optimization algorithm, minimizing the Reduced Chi-Square cost function (χ_r^2), as indicated in Chapter 3.1.3. To provide sufficient statistical representation, 500 (N_{rep}) repetitions of the estimation process were done for each set of input parameters were done with a different random noise added.

3.1.7. UNIFORMLY SELECTED INITIAL PARAMETERS

The solution of the optimization algorithm depends on the configured initial parameters of the search [Lathuillere et al., 1983]. Very different results are obtained assuming different initial parameters, as shown in previous studies [Wu et al., 2015]. In practical ISR observatories, the most common method is to set the initial parameters as those calculated by an ionospheric model, such as the International Reference Ionosphere (IRI) [Bilitza et al., 2017] or the Mass Spectrometer Incoherent Scatter Radar (MSIS) models [Picone et al., 2002]. However, because of the TICA problem, this method would generate 'incorrect' estimates when theoretical model considerations were significantly different from actual values.

To ensure solutions independent of the initial parameters selected, different initial parameters were used at each repetition of the estimation process. Therefore, each estimation of plasma parameters is repeated with a different noise added and a different initial guess of parameters. The initial parameters were uniformly selected from the full search range of parameters (indicated in Chapter 3.1.6). Identical uniformly selected initial parameters were used in all simulations to allow a direct comparison of results of the different simulations done.

Figure 3.4 shows the graphic representation of the complete Monte Carlo analysis method, including the uniform selection of both input parameters ($\boldsymbol{x}_{true}$) and initial parameters ($\boldsymbol{x}_{initial}$). The number of different Monte Carlo input parameters (N_{MC}) was 2000 or 1000, depending on the simulation, and the number of estimation repetitions of each input parameter (N_{rep}) was 500.

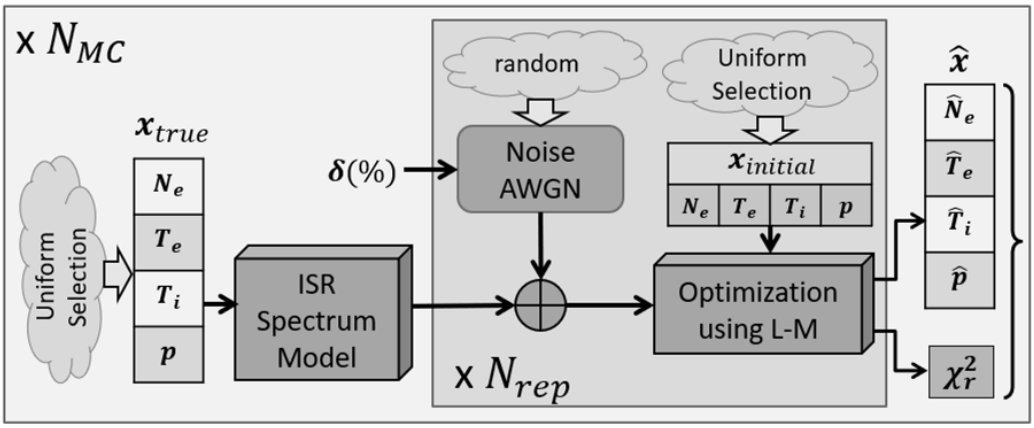

Figure 3.4. Graphic representation of the Monte Carlo simulation at each signal fluctuation value ($\delta(\%)$). Each plasma parameter estimated ($\hat{\boldsymbol{x}}$) and the cost function values obtained ($\boldsymbol{\chi_r^2}$) are matrices of $N_{MC}xN_{rep}$ elements.

3.1.8. SIMULATION OF INACCURACY OF INITIAL PARAMETERS

To determine the effects derived from the use of an inaccurate initial guess of plasma parameters obtained from an ionospheric model, Monte Carlo simulations were done with different configurations of initial parameters at different signal fluctuation levels. Results of these simulations are shown in Chapter 4. Determination of the Impact of the Inaccuracy on the Initial Parameters.

The initial parameters ($\boldsymbol{x}_{initial}$) of simulations shown in Chapter 4 were selected randomly using a uniform distribution centered on the true input plasma parameters ($\boldsymbol{x}_{true}$). The selection range was configured relative to the full search range of parameters as:

$$\boldsymbol{x}_{initial} = \boldsymbol{x}_{true} + \frac{\beta}{100} \cdot (\boldsymbol{x}_{search\,max} - \boldsymbol{x}_{search\,min}) \cdot (rand(n) - 0.5) \tag{15}$$

where β is the initial parameters range percentage ($\beta(\%) \in [0, 100]$), and $rand(n)$ is a random function generator with a uniform distribution between 0 and 1.

Consequently, the initial search parameters were located randomly in the parameter space defined as:

$$\boldsymbol{x}_{initial} \in [\boldsymbol{x}_{true} - \Delta/2\,, \boldsymbol{x}_{true} + \Delta/2] \tag{16}$$

where $\Delta = \beta/100 \cdot (\boldsymbol{x}_{search\,max} - \boldsymbol{x}_{search\,min})$ is the size of the initial search range. This parameter (Δ) resembles the uncertainty of the solution of an ionospheric model used as an initial guess of the optimization algorithm search.

3.1.9. SIMULATION OF UNCERTAINTY OF *A-PRIORI* KNOWN PARAMETERS

To provide a more realistic radar framework, different levels of uncertainty in the plasma parameters known *a priori* from the Plasma Line have been analyzed. Results from these simulations are shows in Chapter 7 Determination of the Impact of the Uncertainty of *a-priori* Known Parameters.The study of [Vallinkoski and Lehtinen, 1990] analyzed the effect of the level of uncertainty of the a priori given parameters, demonstrating the dependence of the estimation error on the *a priori* accuracy.

Previous studies obtained experimental Plasma Line parameter uncertainties for N_e between 1% and 3% at different altitudes [Djuth et al., 1994] [Nicolls et al., 2006], for T_e/T_i of approximately 0.5% [Aponte et al., 2007], and for T_e of 5% to 6% using the frequency asymmetry method [Nicolls et al., 2006].

The *a priori* uncertainty error (ϵ) has been simulated as a random variation of the *a priori* known parameters ($\boldsymbol{x}_{a\ priori}$). Simulations were done assuming a fixed value of *a priori* known parameters selected randomly from the range defined by the uncertainty error (ϵ) and centered at the true input parameter value ($\boldsymbol{x}_{true}$). The *a priori* known parameters were obtained with a uniform selection in the range of parameters defined by:

$$\boldsymbol{x}_{true}\left(1-\frac{|\epsilon|}{100}\right) \leq \boldsymbol{x}_{a\ priori} \leq \boldsymbol{x}_{true}\left(1+\frac{|\epsilon|}{100}\right) \tag{17}$$

where ϵ is the *a priori* parameter uncertainty percentage and $\boldsymbol{x}_{true}$ is the *a priori* known true input parameter.

Therefore, the *a priori* known values would be selected uniformly in the range defined by $\boldsymbol{x}_{true} \pm \Delta\boldsymbol{x}_{true}$, where the term $\Delta\boldsymbol{x}_{true} = \boldsymbol{x}_{true} \cdot \epsilon/100$ corresponds to the absolute uncertainty and the term ϵ to the percentage of relative uncertainty.

In these simulations each of the N_{rep} (500) estimation repetitions of the true input parameters was done with a different random noise added, with initial parameters selected uniformly from the entire search space of parameters, and with *a priori* known parameters uniformly selected from an uncertainty range defined by ϵ.

3.2. STATISTICAL ANALYSIS

3.2.1. DETERMINATION OF CONVERGENCE

Results from Monte Carlo simulations were analyzed statistically to determine the impact of the ambiguity in different signal fluctuation scenarios. In some cases, the optimization algorithm did not converge to a valid solution. To identify which solutions converged to a minimum of the χ_r^2 cost function, the *goodness of the fit* was calculated based on the Chi-Squared (f_{χ^2}) statistical distribution [Taylor, 1997] [Bevington & Robinson, 2003] shown in Equation (14).

Figure 3.5 shows results obtained from the Monte Carlo simulation of 2000 plasma parameters with signal fluctuation values (Top) $\delta = 0.01\%$, (Middle) $\delta = 0.1\%$, and (Bottom) $\delta = 1\%$. These simulations were done estimating 3 unknown plasma parameters (T_e, T_i, and p), assuming known *a priori* the electron density (N_e) parameter. In this figure, the histogram of the χ_r^2 cost function (blue bars) obtains very similar values to the Chi-Squared (f_{χ^2}) statistical distribution (continuous black line) calculated with $DoF = 47$ (with $M = 50$ and $P = 3$). In this figure it is also shown (red line) the Gaussian distribution that best approximates the Chi-Squared distribution (with $\mu_r = 1$ and $\sigma_r^2 = 2/DoF$, as indicated in Chapter 3.1.3) normalized to the histogram maximum value [Andrae et al., 2010]. The good agreement between the theoretical Chi-Squared statistical distribution and the histogram suggests the use of this distribution to determine the probability of obtaining a range of χ_r^2 cost function values.

Estimated parameters ($\hat{\boldsymbol{x}}$) with χ_r^2 smaller than a maximum cost function value ($\chi_r^2 \leq \chi_{r,max}^2$) were considered to have a valid convergence fit. The $\chi_{r,max}^2$ value has been selected to have a probability of $P(x > \chi_{r,max}^2 \,;\, x \sim f_{\chi^2}) = 0.00317\%$, which corresponds to the probability of the 4σ criterion of a normally distributed function (i.e. $P(x > \mu + 4\sigma \,;\, x \sim \mathcal{N}(\mu, \sigma^2)) = 0.00317\%$). The selected probability is small enough to ensure that a fit with $\chi_r^2 > \chi_{r,max}^2$ is obtained by an invalid convergence. Figure 3.5 shows the maximum value to consider a valid convergence ($\chi_{r,max}^2$) with a vertical black dotted line. In this figure, most solutions are found below the $\chi_{r,max}^2$ convergence limit. In our study, maximum and minimum values of $\chi_{r,max}^2$ are 2.073 and 2.0317, for a spectrum vector length of 50 (M) and 5 and 2 plasma parameters (P), respectively.

Figure 3.6 shows the bi-dimensional histogram of the absolute estimation error of parameters ($|\boldsymbol{x}_{true} - \hat{\boldsymbol{x}}|$) and the corresponding cost function χ_r^2 values obtained at different signal fluctuation values. Values shown in this figure are obtained from the simulations shown in Figure 3.5. The maximum cost function value ($\chi_{r,max}^2$) shown in Figure 3.6 is represented as a horizontal black dotted line. This figure shows large absolute estimation errors when $\chi_r^2 > \chi_{r,max}^2$ in all signal fluctuation cases studied. This result verifies the assumption that the optimization algorithm did not converge to a valid solution when the cost function value is larger than $\chi_{r,max}^2$.

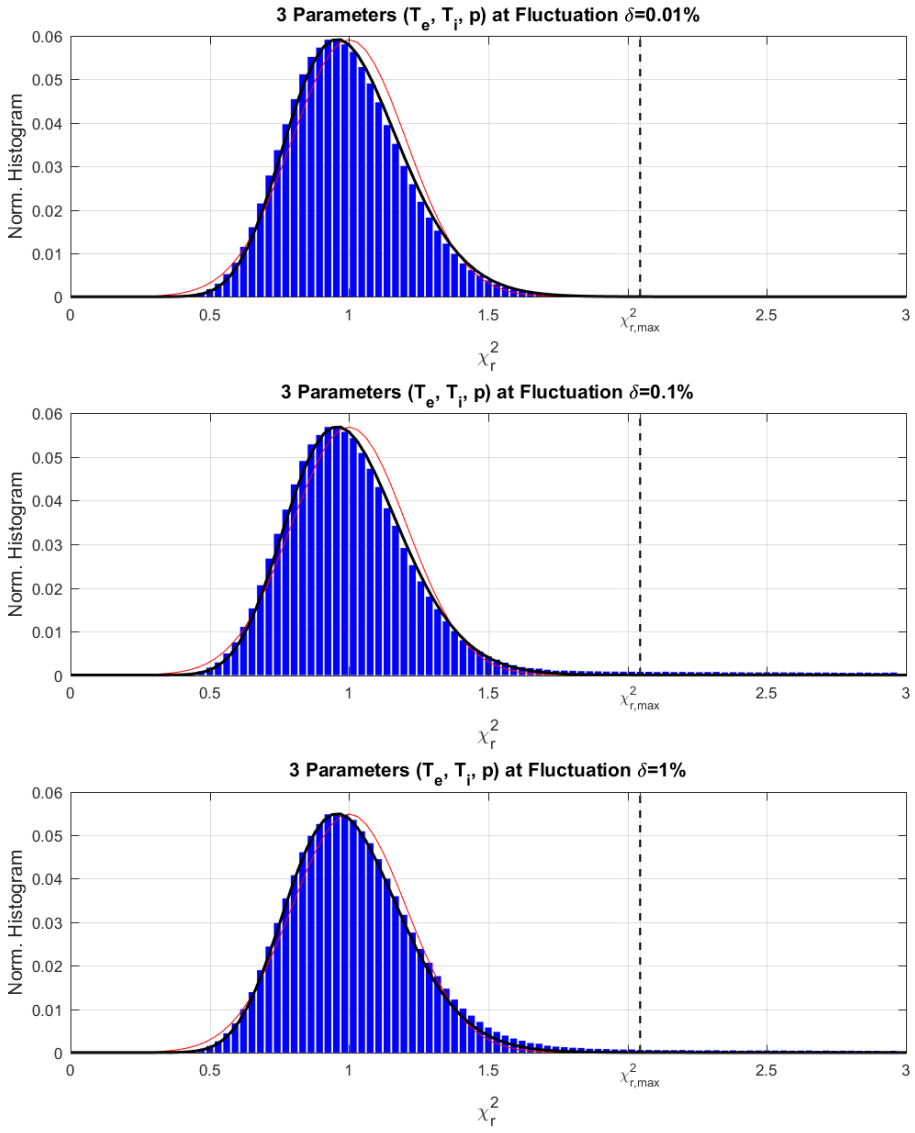

Figure 3.5. Histograms of χ_r^2 (blue bars) at different signal fluctuations. The black line of the histogram represents the PDF of the Chi-Square distribution (f_{χ^2}). This graphic is limited to values of χ_r^2 in the range [0, 3] for a detailed representation of the histogram.

Simulated cases with very small signal fluctuation shown in Figure 3.6 (left column with $\delta = 0.01\%$) obtained small absolute estimation errors when $\chi_r^2 \leq \chi_{r,max}^2$. This result indicates that only 'correct' solutions are found below $\chi_{r,max}^2$, verifying the initial hypothesis that it is possible to estimate unambiguously at very small signal fluctuations. Alternatively, when the signal fluctuation value increases in Figure 3.6 (at middle and right columns with $\delta = 0.1\%$ and $\delta = 1\%$, respectively), solutions with large absolute estimation errors are obtained below $\chi_{r,max}^2$, indicating the existence of an ambiguous estimation.

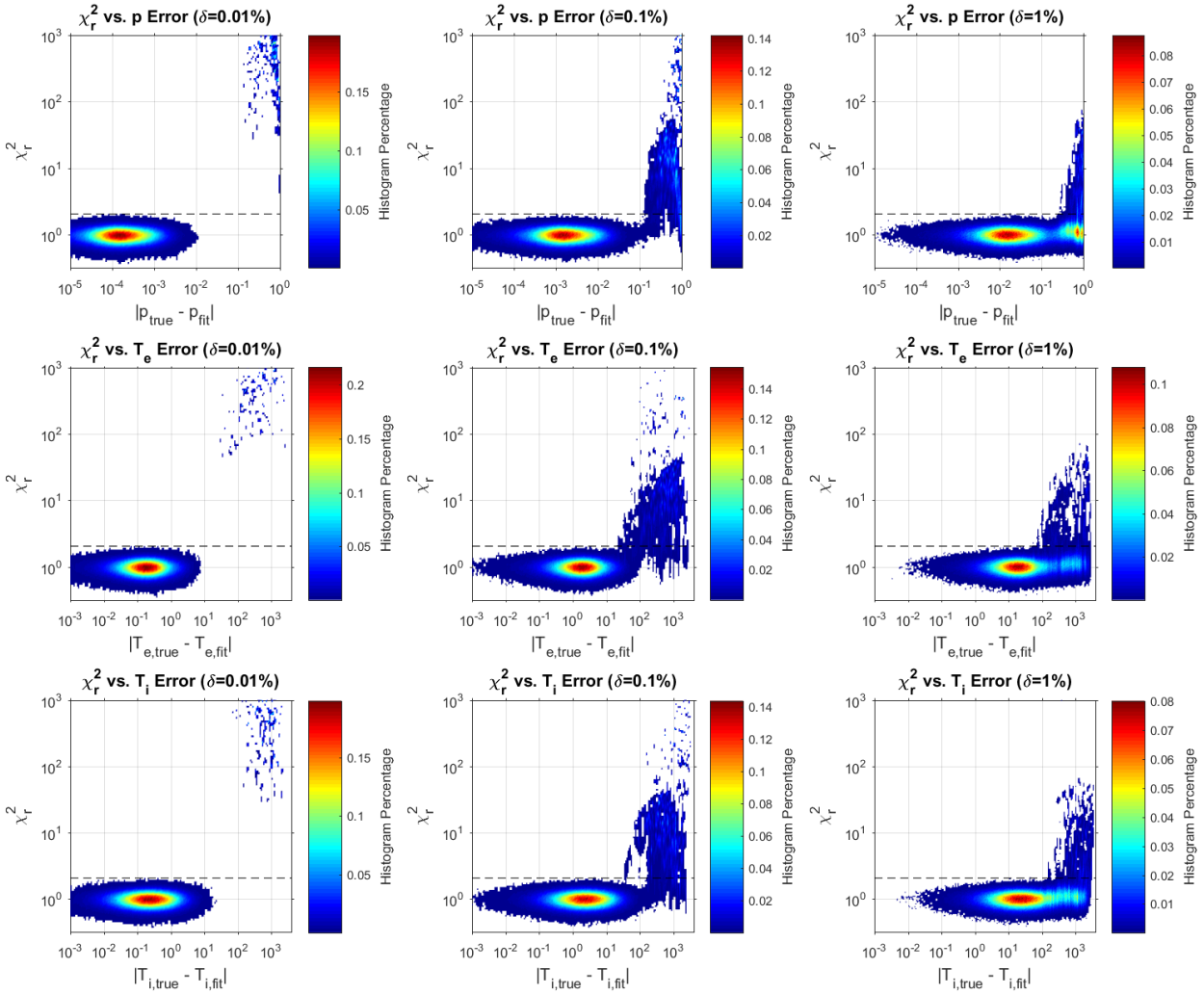

Figure 3.6. Bi-dimensional histograms of χ_r^2 and absolute errors of (Top) ion composition (p), (Middle) electron temperature (T_e), and (Bottom) ion temperature (T_i) parameters at signal fluctuation values of δ = 0.01%, δ = 0.1% and δ =1% at left, middle, and right columns, respectively. The horizontal black dotted line corresponds to the maximum value to consider a valid convergence ($\chi^2_{r,max}$).

3.2.2. DETERMINATION OF CORRECTNESS

Due to the ambiguity problem, in some cases a local minimum of the cost function (χ_r^2) was selected instead of the global minimum. A solution is selected as 'correct' or 'incorrect' depending on the distance between the estimated ($\hat{\boldsymbol{x}}$) and the true input parameters of the simulation ($\boldsymbol{x}_{true}$). Near parameters were assumed 'correct' and distant values 'incorrect'. Note that the 'correct' statement is not an indication of exact estimation without uncertainty (or deterministic), but of global minimum selection. As the estimation process is a stochastic process due to the randomness of the added noise, the estimated plasma parameters also have stochastic distributions. Therefore, the distance discrimination criterion depends on the characteristics of the probability distributions of the estimated results. The process to obtain the statistical parameters of both 'correct' and 'incorrect' distributions of parameters is equivalent to the selection of different data groups in a clustering analysis where some of the statistical information is hidden or unknown [Jain et al., 1999].

3.2.3. CLUSTERING ALGORITHM

In this work, the Expectation Maximization (EM) algorithm was used [Depmster et al., 1977] for clustering 'correct' and 'incorrect' estimated solutions. The EM algorithm is a general method commonly used for partitional clustering [Jain et al., 1999]. The EM algorithm and its configuration are explained in Annex 4. The Expectation Maximization Algorithm.

A total of 500 repetitions of the estimation process were done for each input parameter of the Monte Carlo simulation. Therefore, it was assumed that both 'correct' and 'incorrect' solutions had Gaussian probability distributions, according to the Central Limit Theorem. Consequently, the statistical distribution to be determined by the EM algorithm was a bimodal Gaussian Mixture Model (GMM) given by:

$$f_{GMM}(x \mid \alpha, \mu_0, \sigma_0^2, \mu_1, \sigma_1^2) = \alpha \cdot \mathcal{N}(x \mid \mu_0, \sigma_0^2) + (1-\alpha) \cdot \mathcal{N}(x \mid \mu_1, \sigma_1^2) \qquad (18)$$

where α is the weight of the mixture distribution ($\alpha \in [0, 1]$), and $\mathcal{N}(x \mid \mu_i, \sigma_i^2)$ is the Probability Density Function (PDF) of the Gaussian or Normal distribution, being μ_i and σ_i^2 the mean and variance of the distribution, respectively. Therefore, $\mathcal{N}(x \mid \mu_0, \sigma_0^2)$ and $\mathcal{N}(x \mid \mu_1, \sigma_1^2)$ are the PDFs of 'correct' ($\hat{\boldsymbol{x}}_{correct}$) and 'incorrect' ($\hat{\boldsymbol{x}}_{incorrect}$) results, respectively.

In this study, the EM algorithm was used to analyze the statistical distribution of the ion composition estimation error ($\mathcal{E}_p = p_{true} - \hat{p}$, where p_{true} is the true input ion composition and $\hat{p}$ is the estimated value). This parameter was selected as a discriminator of correctness because of the known existence of two different solutions of ion composition [Oliver, 1979] [Lathuillere et al., 1983] [Wu et al., 2015] and the limited range of this parameter (i.e. $\mathcal{E}_p \in [-1, 1]$). Parameters obtained with a convergent solution ($\chi_r^2 \leq \chi_{r,max}^2$) were clustered as 'correct' or 'incorrect' based on the distance of each ion composition estimation error ($\mathcal{E}_p$) to the GMM PDFs. Figure 3.7 shows a histogram of ion composition estimation errors ($\mathcal{E}_p$) and the corresponding GMM PDFs obtained by the EM algorithm. The estimate errors of this figure were obtained with the plasma parameters indicated as Case a_1 in Figure 2.5.

The 'correct' distribution mean is approximately equal to zero ($\mu_0 \approx 0$), because results obtained with a small ion composition error are 'correct' solutions ($\hat{p}_{correct} \approx p_{true}$). To improve the accuracy of the estimation in noisy scenarios, the mean of the 'correct' distribution was fixed to zero in the EM algorithm ($\mu_0 = 0$). Otherwise, results obtained without this fixing obtained highly variable results and an increase of the probability of 'correct' estimation at high signal fluctuations ($\delta > 10\%$).

Figure 3.8 shows the (red) 'correct' and (blue) 'incorrect' solutions of ion composition obtained from simulations at different signal fluctuations (δ = 0.01%, 0.1%, 1%, and 10%), assuming electron density parameter (N_e) known *a priori*. This figure shows that 'incorrect' solutions are approximately symmetric with respect to 0.5 to the 'correct' solutions (i.e. $\hat{p}_{incorrect} \approx 1 - p_{true}$), as indicated by [Lathuillere et al., 1983] and [Wu et al., 2015]. Therefore, the 'incorrect' distribution mean value could be approximated to $\mu_1 \approx 2p_{true} - 1$. Nevertheless, due to the high variability of the 'incorrect' solutions obtained in noisy scenarios ($\delta \geq 1\%$), this approximation can be used accurately only for small fluctuation cases.

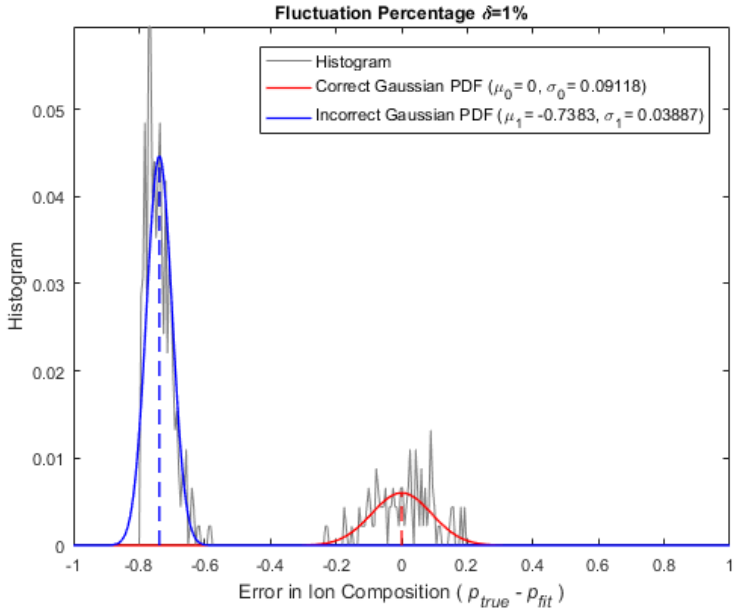

Figure 3.7. (Grey line) Histogram of error of ion composition estimations ($\varepsilon(p) = p_{true} - \hat{p}$) obtained from a Monte Carlo simulation with signal fluctuation $\delta = 1\%$. Input plasma parameters of this simulation were those of spectrum Case a_1 of Figure 2.5, assuming electron density parameter known a priori. The red and blue curves show respectively the 'correct' and 'incorrect' distributions of the Gaussian Mixture Model (GMM) calculated using the EM algorithm.

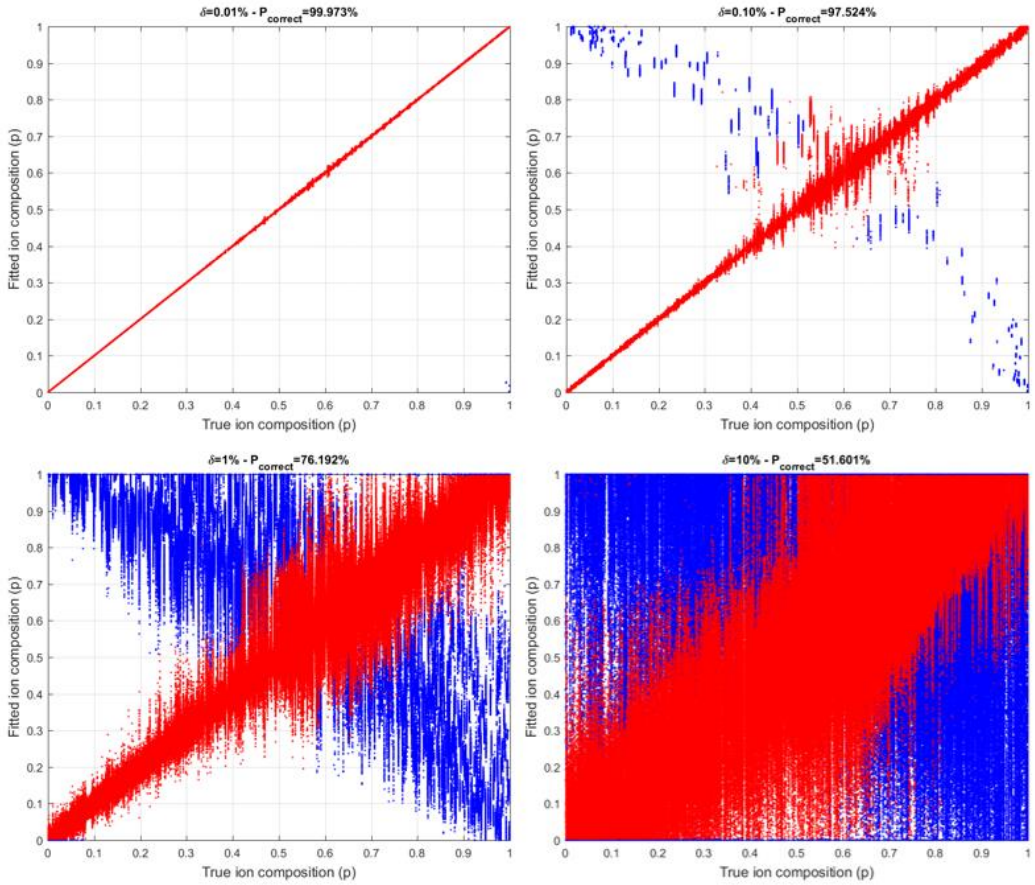

Figure 3.8. (Red) 'Correct' and (Blue) 'Incorrect' results of ion composition obtained by the EM algorithm at signal fluctuation values of (Top Left) δ = 0.01%, (Top Right) δ = 0.1%, (Bottom Left) δ =1%, and (Bottom Right) δ =10%.

3.2.4. PROBABILITIES OF CONVERGENCE AND CORRECTNESS

Different types of probabilities were calculated to quantify the estimation ambiguity at different signal fluctuation values. These probabilities provide information about the maximum signal fluctuation thresholds (δ_{th}) that can be assumed to obtain unambiguous estimates.

The probability of valid convergence of the fit ($P_{fit\ valid}$), defined as the probability of finding a minimum of the χ_r^2cost function by the L-M optimization algorithm, was calculated as:

$$P_{fit\ valid} = N_{fit\ valid}/N_{total} \tag{19}$$

where $N_{fit\ valid}$ is the number of valid convergence results from the simulation ($\chi_r^2(\hat{\boldsymbol{x}}) \leq \chi_{r,max}^2$), and N_{total} is the total number of parameters simulated. Note that $N_{total} = N_{MC} \cdot N_{rep}$, where N_{MC} is the number of different input parameters of the Monte Carlo simulation (i.e. N_{MC}=1000 or 2000 depending on the simulation), and N_{rep} is the number of repetitions of each input parameter of the simulation (i.e. N_{rep}=500).

Alternatively, the probability of 'correct' estimation ($P_{correct}$), which represents the probability to successfully select the global minimum of the cost function having already converged to a valid solution, was calculated as:

$$P_{correct} = N_{correct}/N_{fit\ valid} \tag{20}$$

where $N_{correct}$ is the number of 'correct' results of a simulation (i.e. $\hat{\boldsymbol{x}} \approx \boldsymbol{x}_{true}$) calculated by the EM clustering algorithm.

The probability of valid convergence and 'correct' estimation ($P_{fit\ valid\ \&\ correct}$), defined as the total number of parameters calculated without an ambiguous solution, was calculated as the product of previous probabilities:

$$P_{fit\ valid\ \&\ correct} = P_{fit\ valid} \cdot P_{correct} = N_{correct}/N_{total} \tag{21}$$

Note that $P_{fit\ valid\ \&\ correct}$ represents the probability of solving a set of parameters and estimating it without ambiguity, and consequently indicates the total probability of unambiguous estimation when the convergence of the solution ($\chi_r^2(\hat{\boldsymbol{x}}) \leq \chi_{r,max}^2$) is not verified.

CHAPTER 4

DETERMINATION OF THE IMPACT OF THE INACCURACY ON THE INITIAL PARAMETERS

Different Monte Carlo simulations have been done to determine the effect of using an ionospheric model to determine the initial parameters of the optimization algorithm search with different levels of accuracy. The initial parameters of these simulations were configured as indicated in Chapter 3.1.8. Simulation of Inaccuracy of Initial Parameters.

These simulations were done for 1000 different true input values ($\boldsymbol{x}_{true}$) estimating electron density (N_e), electron temperature (T_e), ion temperature (T_i), and ion composition (p) parameters. This configuration of plasma parameters corresponds to the typical estimation process for ISR radars, when no *a priori* information is provided. Results from these simulations demonstrate the necessity to study the ambiguous estimation using an uniform selection of initial parameters (as indicated in Chapter 3.1.7. Uniformly Selected Initial Parameters).

Figure 4.1 shows the results of Monte Carlo simulations with initial parameter range percentages of β=1%, 5%, 10%, 20%, 30%, 40%, 50%, and 100%, demonstrating the effects of increasing the uncertainty on the initial parameters.

4.1. ACCURATE INITIAL GUESS

The most remarkable result shown in Figure 4.1 was obtained in the case of an almost perfect guess of the initial parameters ($\beta = 1\%$). Even if the initial parameters were extremely accurate, unambiguous estimates (i.e. $P_{correct} \approx 100\%$) were obtained only with signal fluctuations smaller than $\delta < 0.1\%$. This result implies the existence of a threshold on the fluctuation level that can be tolerated to completely solve the TICA problem, even when using a precise ionospheric model as initial guess.

To obtain a $P_{correct} \geq 95.45\%$ (i.e. a 2σ probabilistic criterion), the signal fluctuation threshold was $\delta_{th(\beta=1\%)} = 0.54\%$ in the case of $\beta = 1\%$. Alternatively, $P_{correct}$ decayed to less than 50% for signal fluctuations $\delta > 5\%$, indicating that in noisy scenarios it was more probable to select the 'incorrect' solution. On the other hand, the probability of valid convergence ($P_{fit\ valid}$) was almost 100% in all signal fluctuation scenarios, indicating that the L-M optimization algorithm was always able to converge to a minimum of the cost function.

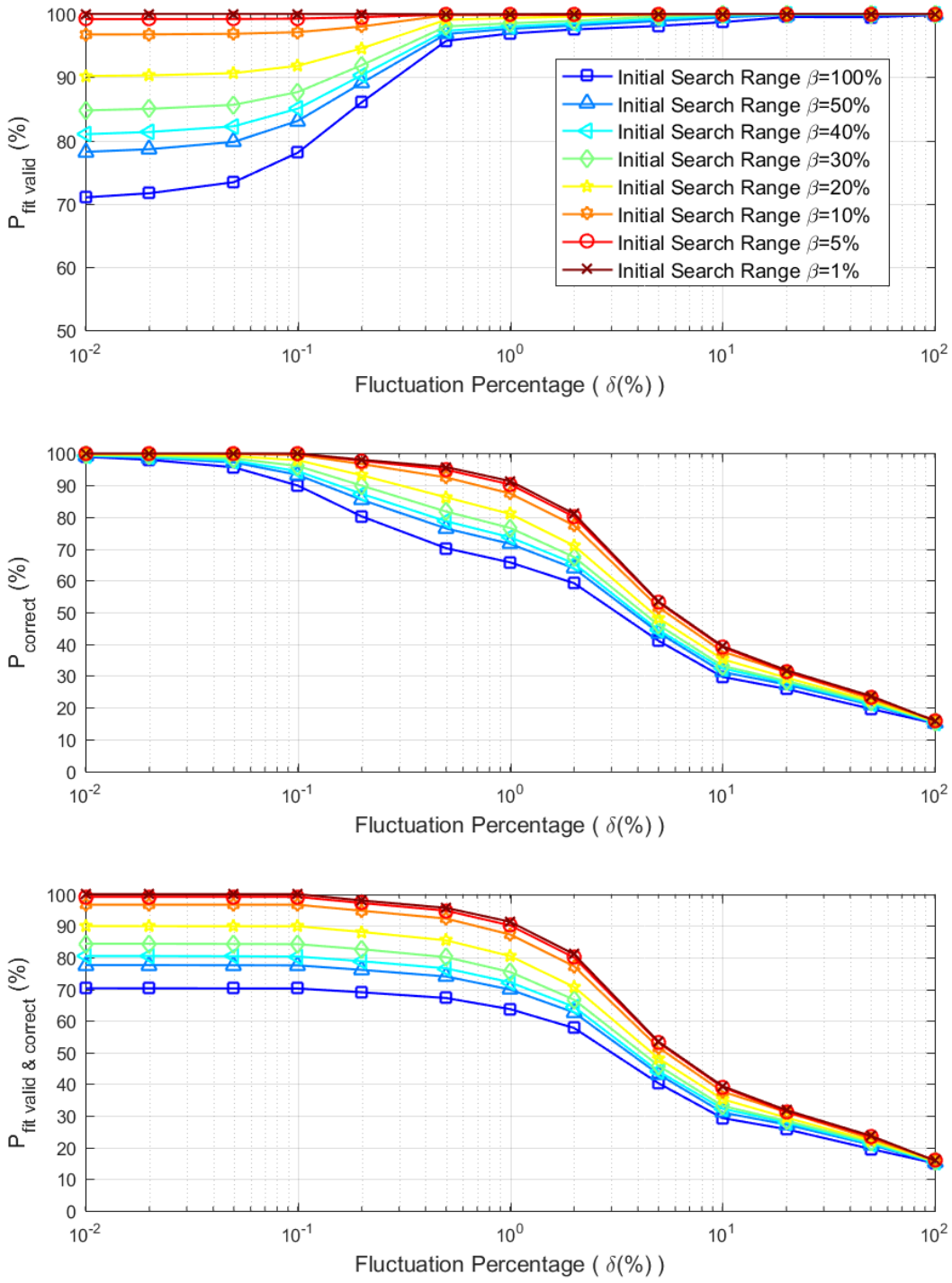

Figure 4.1. Probability of convergence ($P_{fit\ valid}$) and probability of 'correct' estimation ($P_{correct}$) at different signal fluctuation percentages ($\delta(\%)$) obtained by simulations of the estimation of four plasma parameters (N_e, T_e, T_i, and p parameters) without parameters given a priori, for different ranges of initial search parameters ($\beta(\%)$).

4.2 INCREASING THE UNCERTAINTY OF INITIAL PARAMETERS

The increase of the uncertainty of the initial parameters ($\beta > 1$) produces a reduction of both probabilities $P_{fit\ valid}$ and $P_{correct}$ (Figure 4.1). Nevertheless, the reduction of $P_{fit\ valid}$ was localized in the low fluctuation regime ($\delta \leq 0.5\%$), while the reduction of $P_{correct}$ was obtained at high fluctuations ($0.1\% \leq \delta \leq 10\%$). These different ranges indicate that two different estimation issues occurred at different ranges of signal fluctuation.

The inversely proportionality of $P_{fit\ valid}$ to β occurs because increasing this uncertainty is equivalent to increase the distance between the initial and the true input parameters, making it more difficult to find the global minimum. This effect is shown in Figure 4.2, where the number of computing iterations is proportional to the uncertainty of the initial parameters (β). Furthermore, this figure shows that more iterations are required to estimate plasma parameter at higher signal fluctuation values.

Nevertheless, the decrease of $P_{fit\ valid}$ was found only at the low fluctuation regime in Figure 4.1. A decrease of signal fluctuation (δ) is equivalent to a reduction of the estimated signal variance (σ^2) in Equation (38), resulting in higher χ_r^2 cost function values (Equation (13)). It is assumed that the increase of χ_r^2 obtained non-convergent 'incorrect' estimates ($\chi_{r,incorrect}^2 > \chi_{r,max}^2$) at the low fluctuation regime. This effect is shown in the example of Figure 3.3 (Top), where the χ_r^2 cost value of the local minimum was larger than $\chi_{r,max}^2$ for small signal fluctuations ($\delta = 0.01\%$). This effect would reduce the number of convergent solutions and increase the number of 'correct' estimates when the signal fluctuation level decreases, as shown in Figure 4.1.

At highly noisy scenarios ($\delta > 5\%$), $P_{correct}$ probabilities were smaller than 50% and had almost identical values with different uncertainties. This result indicates that 'incorrect' solutions were more selected at high fluctuation levels, independently of the uncertainty of initial parameters (β). Alternatively, at very small signal fluctuations ($\delta \leq 0.05\%$) $P_{correct} \approx 100\%$ independently of the initial guess uncertainty (β). This latter result indicates that all solutions were 'correctly' estimated even when initial parameters were far from true input parameters. Therefore, this result verifies the assumption that, independently of the initial guess accuracy, there is a threshold of signal fluctuation to completely solve the TICA problem ($\delta_{th(\forall\beta)} \approx 0.05\%$ to obtain a $P_{correct} \geq 95.45\%$).

In summary, the use of ionospheric models to determine initial plasma parameters leads to obtaining erroneous plasma parameters unless signals have almost no fluctuation. Even with a very good model prediction (i.e. $\beta = 1\%$), there is always ambiguity in the estimation of noisy signals. In these noisy scenarios, 'incorrect' solutions are more likely to be selected.

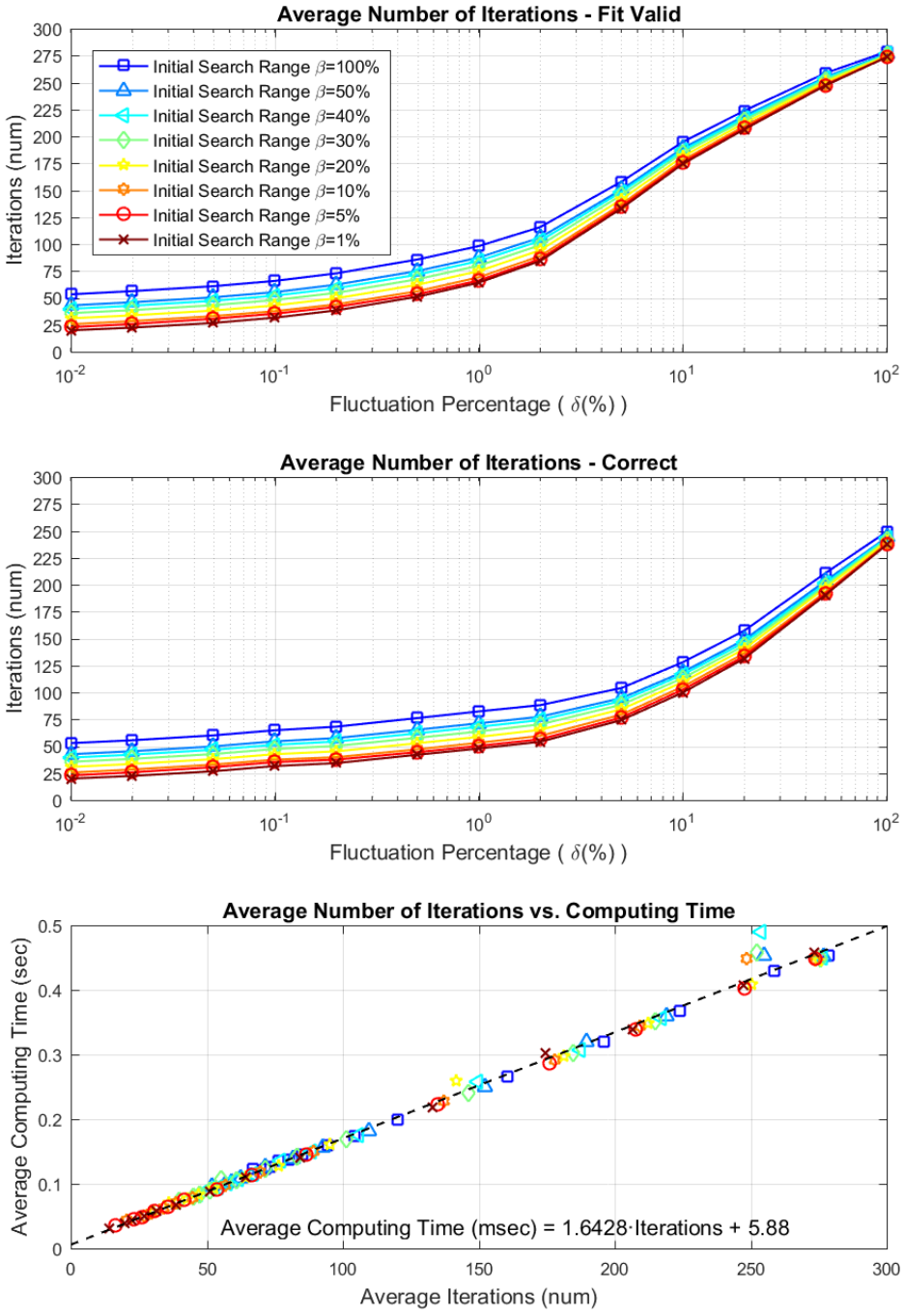

Figure 4.2. (Top) The average number of iterations required to obtain a valid convergence, (Middle) average number of iterations required to obtain a 'correct' estimation, and (Bottom) linear regression of average computing times of each fitting. Simulations were done estimating four plasma parameters (N_e, T_e, T_i, and p) without information given a priori, for different ranges of initial search parameter selected uniformly around the true input value (β(%)). Average computing times were obtained at the NLHPC supercomputer by parallelizing fitting simulations.

4.3 PROPOSED ESTIMATION TECHNIQUE

Consequently, to obtain results independent of the configuration of initial parameters, the use of a Monte Carlo selection scheme of initial parameters is suggested. This technique consists of executing the optimization algorithm several times with different initial parameters uniformly selected from the global search space of parameters, as indicated in Chapter 3.1.7. Uniformly Selected Initial Parameters. The estimated solution would correspond to the set of parameters most frequently obtained. The following sections show the probability results that are obtained using this estimation technique.

CHAPTER 5

ANALYSIS OF THE TEMPERATURE-ION COMPOSITION AMBIGUITY PROBLEM

Monte Carlo simulations were done to analyze the TICA problem adding different information from the Plasma Line without uncertainty. These simulations were done for 2000 different true input plasma parameters at different signal fluctuation levels. In these simulations, the combinations of parameters studied were:

a) 4 unknown parameters (N_e, T_i, T_e, and p) without information given *a priori*;
b) 3 unknown parameters (T_i, T_e, and p) given *a priori* N_e;
c) 2 unknown parameters (T_i and p) given *a priori* N_e and T_e/T_i; and
d) 2 unknown parameters (T_i and p) given *a priori* N_e and T_e.

5.1. AMBIGUOUS ION COMPOSITION ESTIMATE

Figure 5.1 shows estimated ion composition parameters obtained from these simulations, where the different colors represent values estimated at different signal fluctuations (δ). In the study case *d* of Figure 5.1 (Bottom Rigth), the addition of *a priori* knowledge of N_e and T_e information obtained 'correct' solutions (i.e. $\hat{p}_{correct} \approx p_{true}$) at all signal fluctuation scenarios. This result implies that this combination of *a priori* known parameters solved the ambiguity problem, as proposed by [Bjørnå & Kirkwood, 1988] and [Nicolls et al., 2006]. In the other study cases (*a*, *b*, and *c*), unambiguous estimates were only possible at scenarios with almost no fluctuation ($\delta = 0.01\%$, red colored). At middle signal fluctuations ($0.05\% \leq \delta \leq 1\%$, colors cyan to orange) 'incorrect' ion composition values were also obtained (i.e. $\hat{p}_{incorrect} \approx 1 - p_{true}$). In high fluctuation cases ($\delta \geq 5\%$, blue colored), estimated results were obtained spread throughout the entire ion composition parameter range. This latter result indicates that in noisy scenarios estimates had a high uncertainty and were also ambiguous.

The 'incorrect' ion composition values ($\hat{p}_{incorrect}$) were dependent on the *a priori* information provided. Simulations without *a priori* information obtained values of 'incorrect' ion compositions with a parabolic curve dependent on the true input parameter (p_{true}). Alternatively, results with the *a priori* knowledge of N_e or N_e and T_e/T_i parameters were described by $\hat{p}_{incorrect} = 1 - p_{true}$. These latter results agree with previous studies that obtained 'correct' and 'incorrect' solutions approximately equidistant to $p = 0.5$ [Lathuillere et al., 1983] [Wu et al., 2015]. The variability of the ion composition value increased in the vicinity of the intersection between 'correct' and 'incorrect' results ($p \approx 0.5$), indicating a higher estimation uncertainty near the intersection point. This increase of uncertainty near the intersection point was previously indicated in [Aponte et al., 2007].

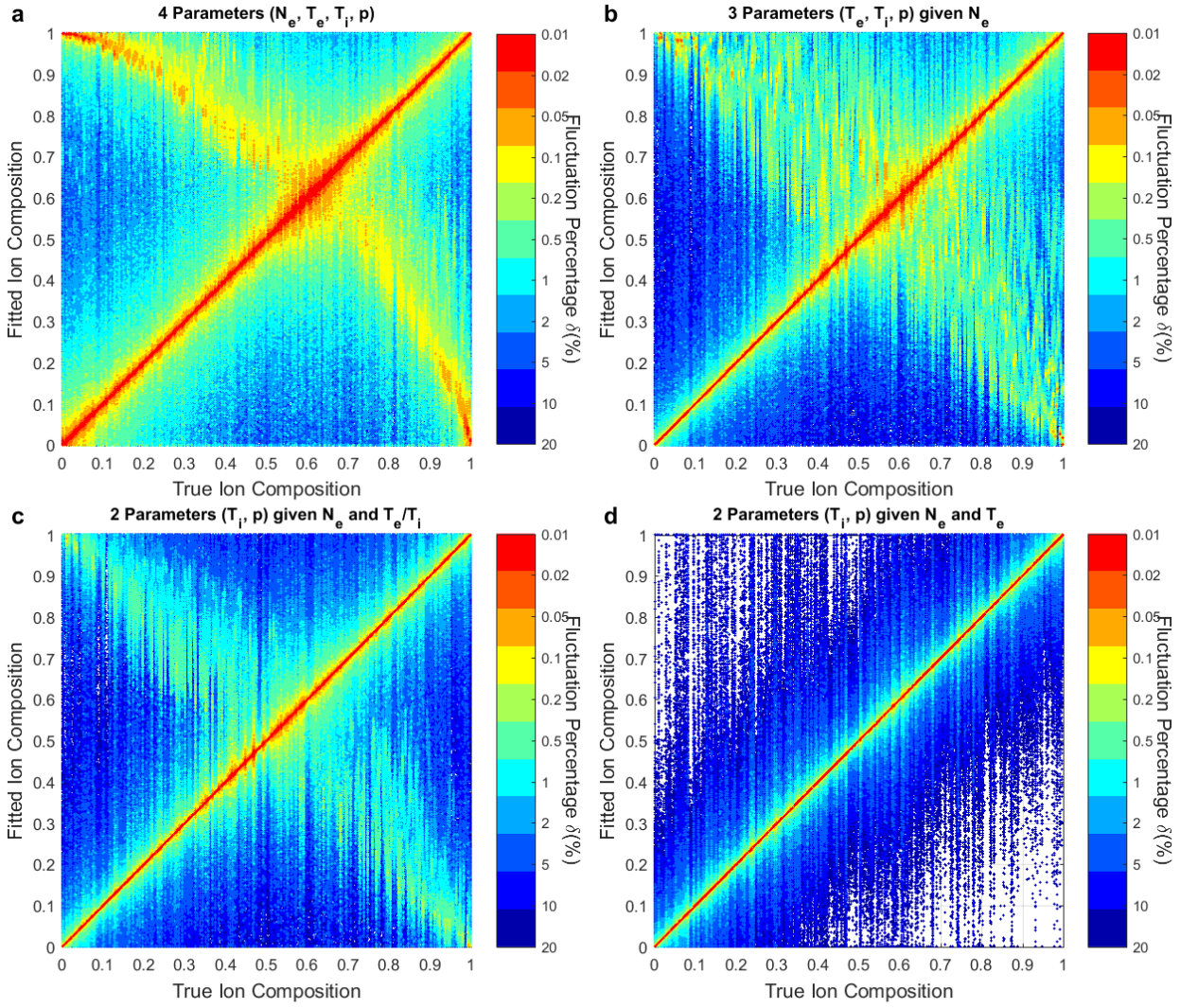

Figure 5.1. Scatter plot of estimated and input values of ion composition, obtained from the analysis of different combinations of parameters known a priori from the Plasma Line: (a) without a priori information, (b) given N_e, (c) given N_e and T_e/T_i, and (d) given N_e and T_e. Each color represents results obtained by simulations with a particular signal fluctuation percentage ($\delta(\%)$).

5.2. UNCERTAINTY OF THE ION COMPOSITION ESTIMATE

Figure 5.2 shows the standard deviation of both 'correct' and 'incorrect' distributions of the GMM PDFs calculated by the EM algorithm ($\sigma_{correct}$ and $\sigma_{incorrect}$, respectively) at different signal fluctuation levels. These values resemble the average uncertainty of the 'correct' and 'incorrect' estimated ion compositions. At the low and middle signal fluctuation regimes ($\delta < 5\%$), both standard deviations increased linearly with the signal fluctuation percentage. Estimated linear regressions of standard deviation (σ_{est}) calculated for study cases *a*, *b*, *c*, and *d* were $\delta/10$, $\delta/30$, $\delta/50$, and $\delta/150$, respectively (Figure 5.2). These different estimated increases indicate that the uncertainty of the ion composition estimate was dependent on the type of *a priori* information provided.

For all study cases, both 'correct' and 'incorrect' standard deviations have a maximum value of approximately $\sigma_{sat} = 0.18$. This saturation value is the maximum standard deviation value that was estimated by the EM algorithm in the ion composition error range ($-1 \leq \epsilon_p \leq 1$). These saturation values were reached at signal fluctuation levels approximately equal to δ_{cross}, where δ_{cross} is the signal fluctuation level at which $P_{correct} = 50\%$ in Figure 5.3. Consequently, δ_{cross} represents the maximum signal fluctuation threshold required to obtain a higher number of 'correct' than 'incorrect' estimates

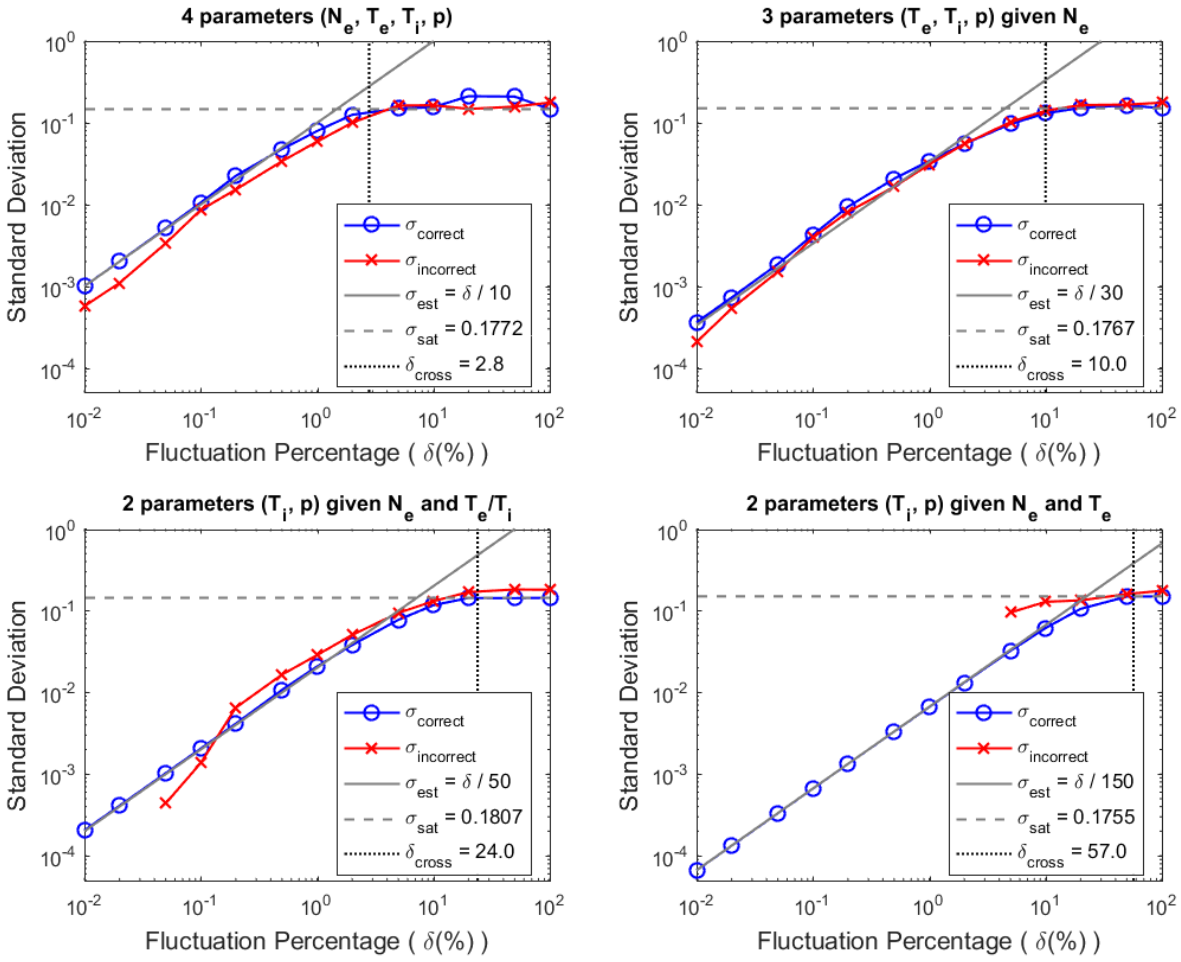

Figure 5.2. Average values of standard deviation (in logarithmical scale) of correct (blue) and incorrect (red) statistical distributions obtained by simulations of the estimation of different combinations of known a priori plasma parameters from the Plasma Line at different signal fluctuation percentages ($\delta(\%)$). Vertical dotted line represents the estimated fluctuation value (δ_{cross}) at which the standard deviation reaches its maximum value (σ_{sat}). Black line represents the estimated linear regression (σ_{est}) before arriving to saturation, and the horizontal dotted line represents the saturation value (σ_{sat}).

5.3. PROBABILITY RESULTS OF PLASMA LINE INFORMATION ADDITION

The probability of convergence of the optimization algorithm ($P_{fit\ valid}$), the probability of having a 'correct' estimation ($P_{correct}$), and their joint probability ($P_{fit\ valid\ \&\ correct}$), are shown in Figure 5.3 for different combinations of known *a priori* parameters (study cases *a*, *b*, *c*, and *d*).

Each of the different study cases shown in Figure 5.3 gradually improved $P_{correct}$, displacing the signal fluctuation threshold (δ_{th}) to higher values. This indicates that the *a priori* knowledge of different parameters provides different amounts of information to solve the TICA problem. Furthermore, this figure shows the signal variability effect on the ambiguous estimation. The signal fluctuation thresholds for obtaining a $P_{correct} \geq 95.45\%$ were found at: $\delta_{th(no\ info)} = 0.045\%$, $\delta_{th(Ne)} = 0.137\%$, $\delta_{th(Ne\ and\ Te/Ti)} = 0.568\%$, and $\delta_{th(Ne\ and\ Te)} = 7.93\%$, for study cases *a*, *b*, *c*, and *d*, respectively. This latter threshold value indicates that the addition of N_e and T_e information solved the TICA problem even for highly noisy signals, as previously indicated.

A previous study of the TICA problem [Aponte et al., 2007] simulated the estimation process at a signal fluctuation of $\delta = 0.5\%$ knowing *a priori* N_e and T_e/T_i parameters from the Plasma line. The results of [Aponte et al., 2007] were obtained without ambiguity, agreeing with the threshold found in our work. Nevertheless, the study of [Aponte et al., 2007] did not consider the convergence of the estimated results, and invalid solutions could be obtained depending on the configurations of the fitting. To account for non-convergent estimates, the probability of unambiguous estimation of this study should be determined by the value of $P_{fit\ valid\ \&\ correct}$.

The addition of *a priori* information also improved the convergence of the optimization algorithm ($P_{fit\ valid}$). The addition of N_e and T_e information (study case *d*) obtained a probability of valid convergence $P_{fit\ valid} > 99\%$ at all studied signal fluctuation values. In this case, the optimization algorithm was always able to find a minimum of the cost function independently of the signal fluctuation level. Other study cases (*a*, *b*, and *c*) obtained a reduction of $P_{fit\ valid}$ for $\delta < 0.5\%$. These reductions were assumed to be related to the effect of decreasing the estimated variance (σ^2) in the calculus of the cost function χ_r^2, previously indicated in Chapter 4.2. Therefore, at small signal fluctuation cases, 'incorrect' estimates were not convergent ($\chi^2_{r,incorrect} > \chi^2_{r,max}$) and only 'correct' solutions were obtained.

The number of computing iterations required was gradually reduced by the addition of *a priori* information at the different study cases. Figure 5.4 shows the average number of iterations required for convergence and 'correct' estimation. Computing times of simulations with *a priori* knowledge of N_e and T_e parameters (study case *d*) were approximately a quarter of those obtained without *a priori* information (study case *a*).

Furthermore, the results shown verify that the use of a large initial parameter uncertainty (i.e. β =100% in Figure 4.1) is equivalent to having initial parameters uniformly selected from the entire search range (case *a* in Figure 5.3, blue color line).

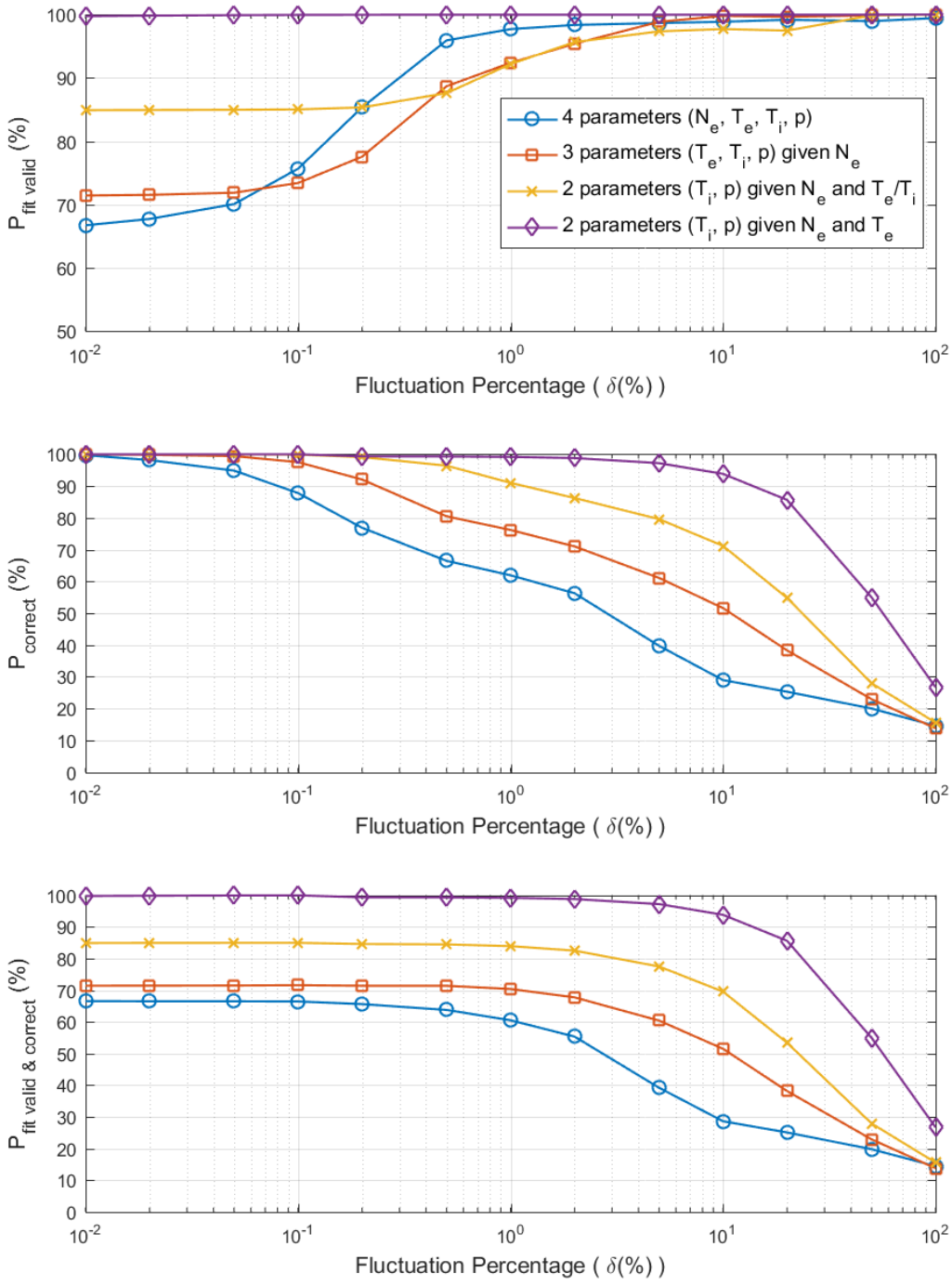

Figure 5.3. Probability of convergence ($P_{fit\ valid}$) and probability of 'correct' estimation ($P_{correct}$) (in percentage) at different signal fluctuation percentages ($\delta(\%)$) obtained by simulations of the estimation of different combinations of known a priori plasma parameters from the Plasma Line: without a priori information (blue circles), given N_e (orange squares), given N_e and T_e/T_i (yellow crosses), given N_e and T_e (purple rhombus).

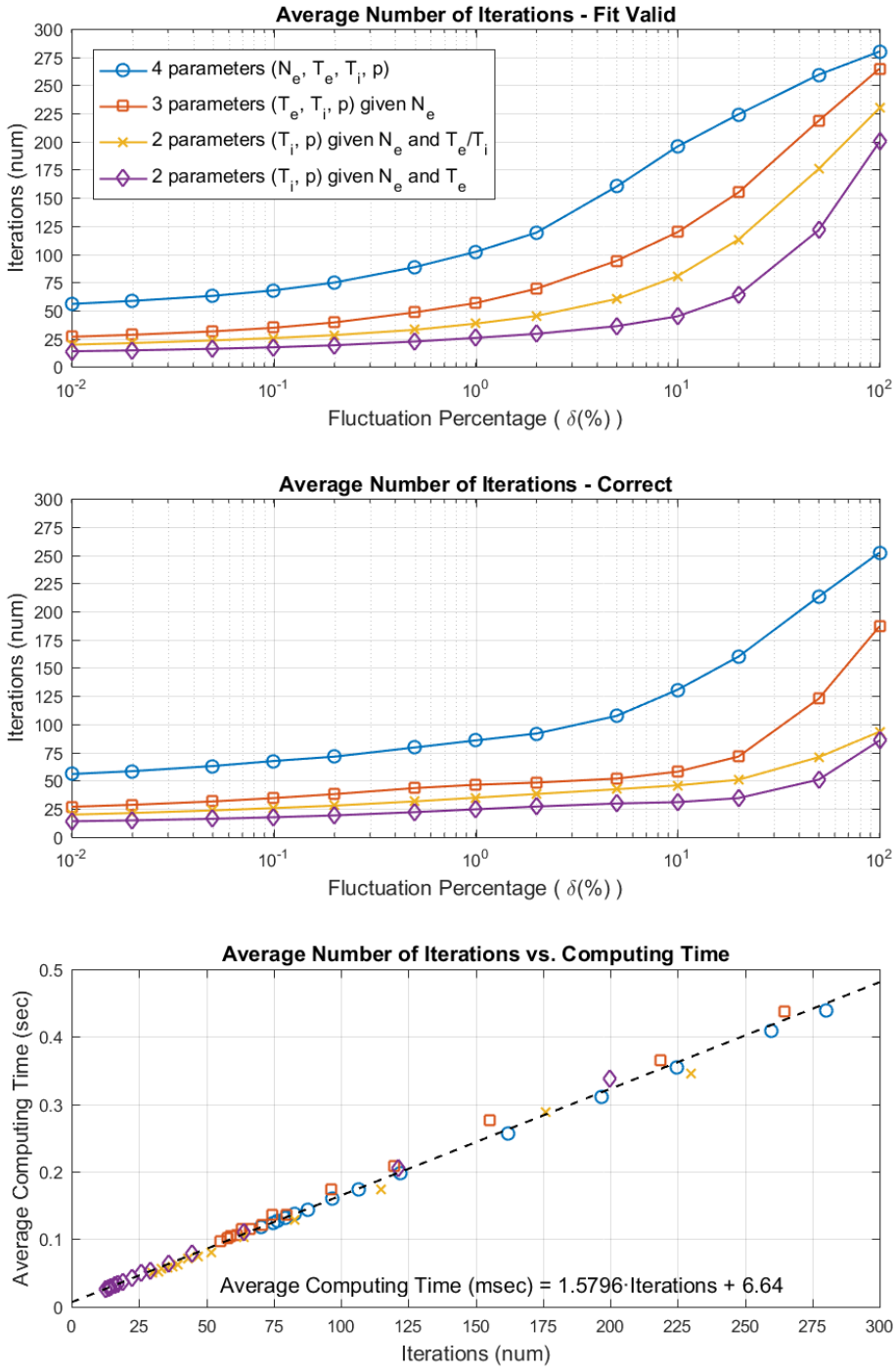

Figure 5.4. (Top) The average number of iterations required to obtain a valid convergence and (Middle) a 'correct' estimation of parameters of simulations with different a priori known parameters. (Bottom) Average computing times of each fitting can be approximated by the linear regression indicated in the graphic. Average execution times were obtained at the NLHPC supercomputer parallelizing the fitting simulations.

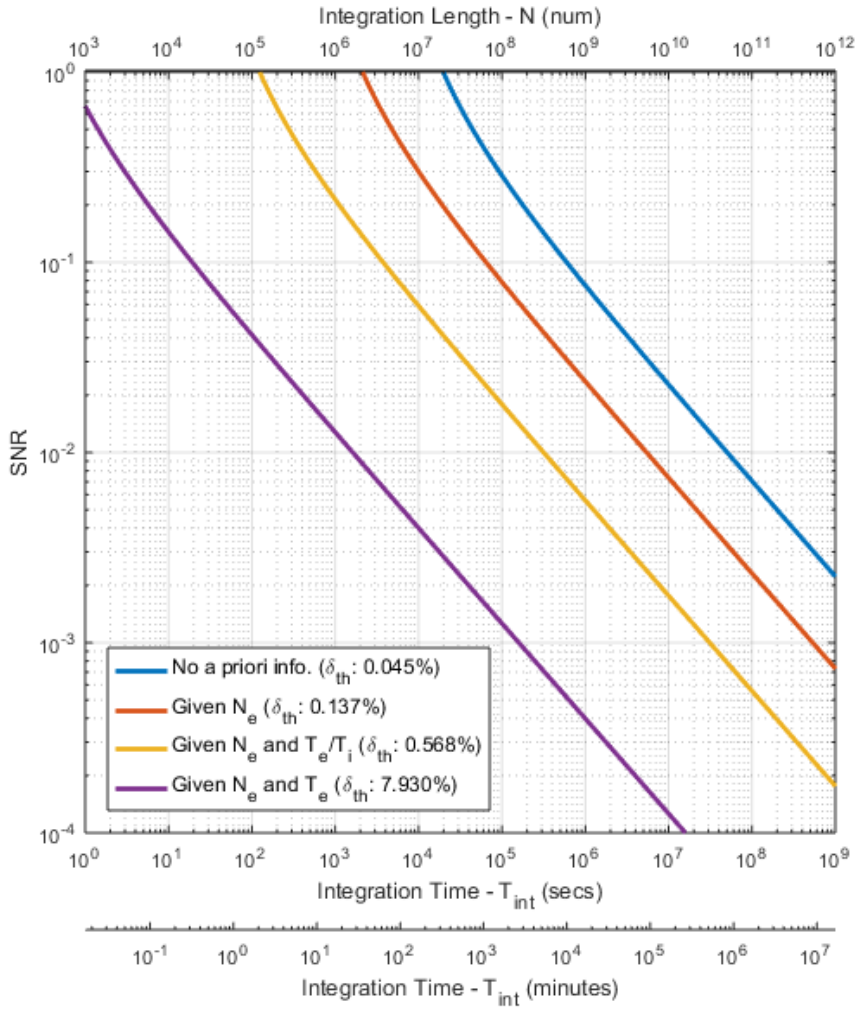

Figure 5.5. Integration times and SNR values required to obtain the signal fluctuation thresholds of unambiguous estimation of (blue) no a priori information provided, (red) given N_e, (yellow) given N_e and T_e/T_i, and (purple) given N_e and T_e parameters.

Figure 5.5 shows the integration times and corresponding SNR values required to obtain the unambiguous signal fluctuation thresholds of study cases *a*, *b*, *c*, and *d*. These values have been calculated using Equation (3) at the signal fluctuation thresholds previously calculated. Integration times have been calculated assuming an Inter-Pulse Period (IPP) of 1 millisecond and it has not been considered the simultaneous reception of signals at multiple frequencies [Sulzer 1986a]. The results shown in this figure indicate that very large integration times are required when no *a priori* information is provided or when only N_e parameter is given *a priori* (study cases *a* and *b*). In such cases, the assumption of stationary plasma conditions [Farley, 1969] may not be fulfilled due to possible variations of the plasma parameters during these long integration times. The non-stationary effect could make infeasible the unambiguous plasma parameters estimation, as indicated in Chapter 7.3. Stationarity of Unambiguous Radar Measurements. Alternatively, integration times of study cases *c* and *d* (assuming the *a priori* knowledge of N_e and T_e/T_i parameters and N_e and T_e parameters, respectively) are feasible by common IS radars under stationary plasma conditions.

5.4. RELATIONSHIPS BETWEEN ESTIMATION ERRORS OF PARAMETERS

Histograms of estimation error of plasma parameters are shown for different study cases of addition of plasma parameters information in Figure 5.6, Figure 5.7, Figure 5.8, and Figure 5.9. In the study cases *a* (Figure 5.6), *b* (Figure 5.7), and *c* (Figure 5.8), histograms show a non-linear dependence between the estimation errors of N_e, T_e, and T_i parameters and the estimation error of the ion composition ($\mathcal{E}_p$). These dependences indicate that N_e, T_e, and T_i parameters are directly related to the ion composition ambiguity. Alternatively, results of study case *d* (Figure 5.9) showed no relationships between the estimation error of T_i and $\mathcal{E}_p$, indicating that there was no estimation ambiguity in the case of knowing *a priori* N_e and T_e.

5.4.1. ESTIMATED ERRORS OF FOUR PARAMETERS (N_e, T_e, T_i, AND p)

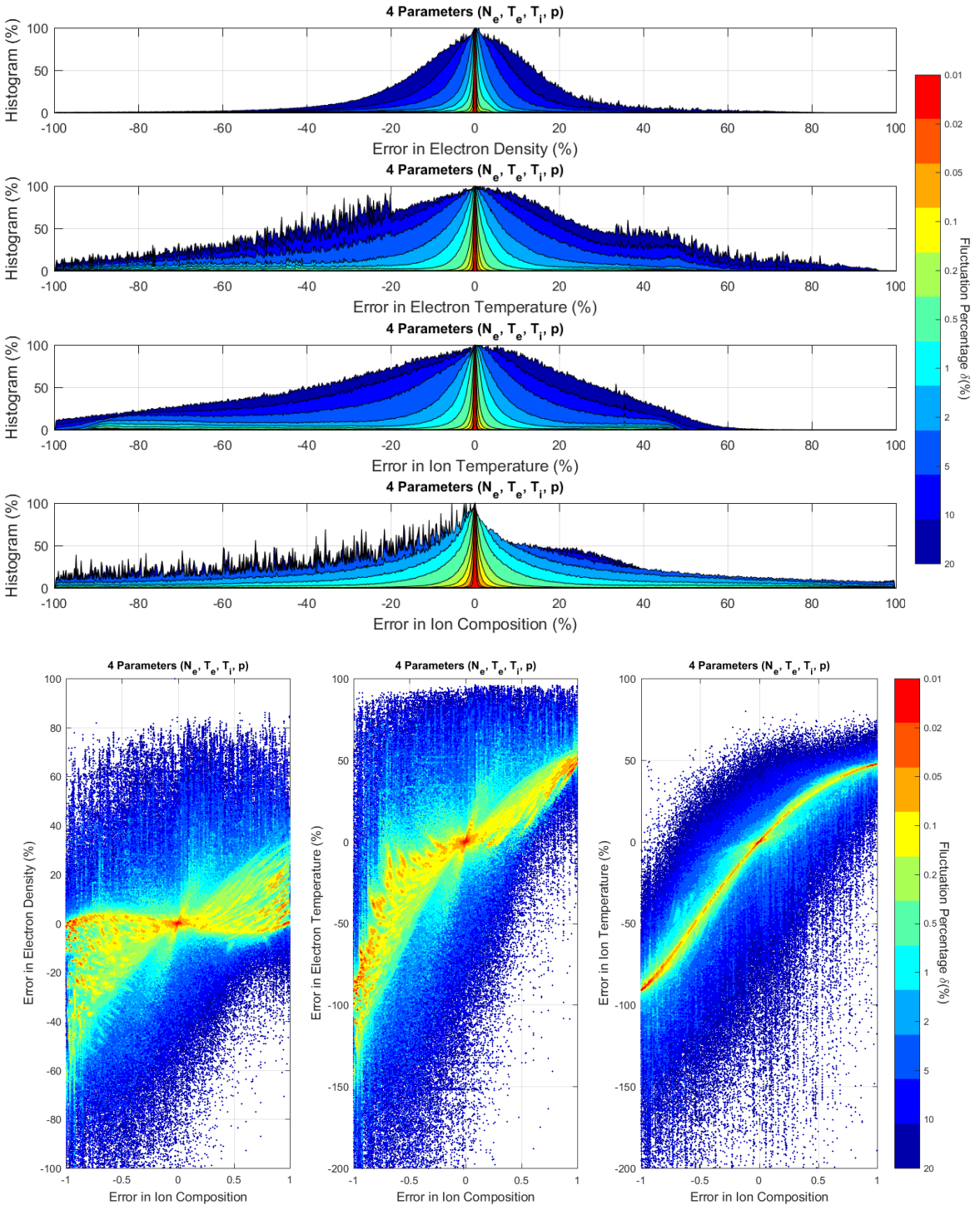

Figure 5.6. (Top) Histogram of estimated parameters errors, and (bottom) relationship between plasma parameters errors (in percentage) and the ion composition error. These results were obtained from the estimation of four plasma parameters (N_e, T_e, T_i, and p) given no a priori information at different signal fluctuation values ($\delta(\%)$).

5.4.2. ESTIMATED ERRORS OF THREE PARAMETERS (T_e, T_i, AND p) GIVEN A PRIORI N_e

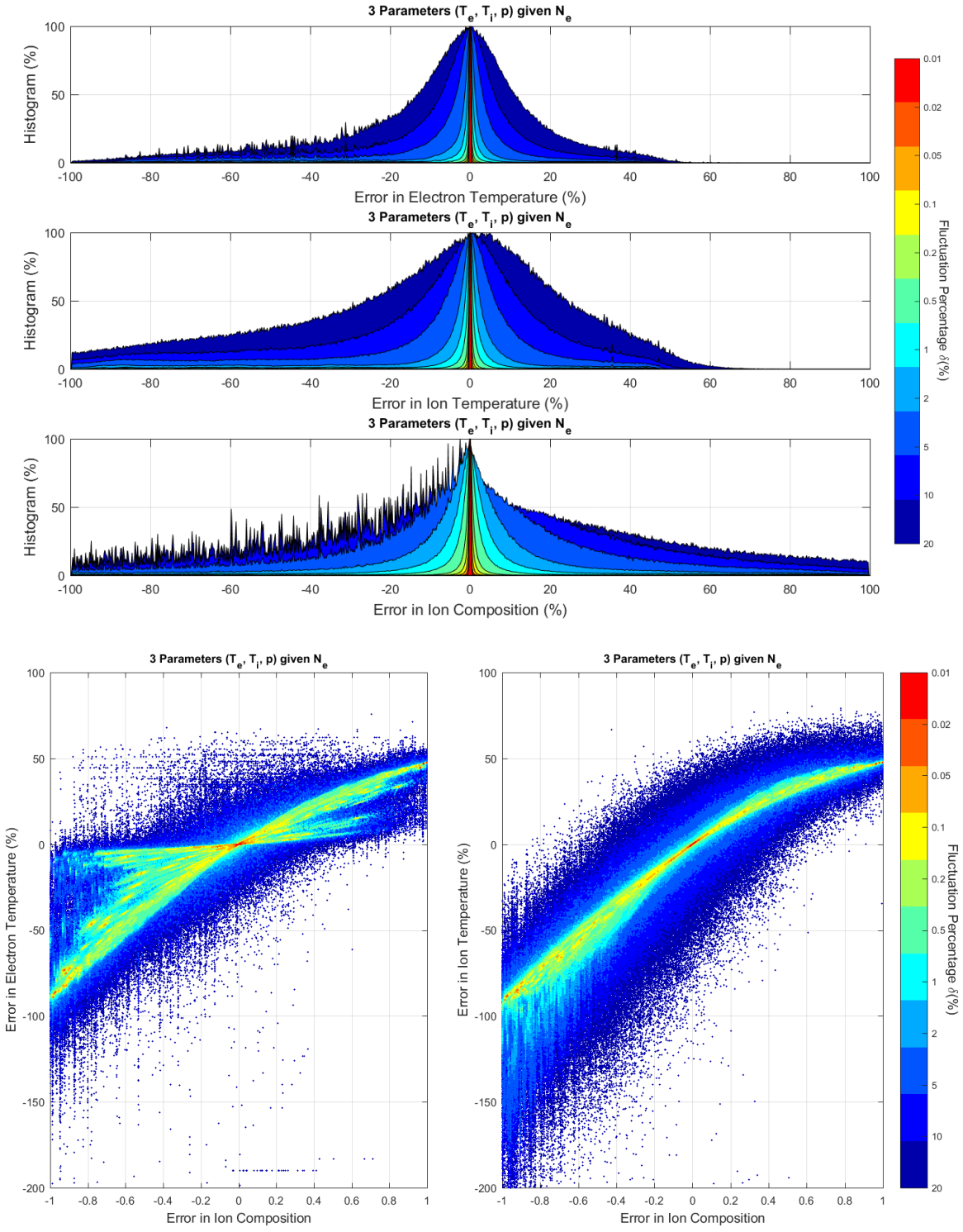

Figure 5.7. (Top) Histogram of estimated parameters errors, and (bottom) relationship between plasma parameters errors (in percentage) and the ion composition error. These results were obtained from the estimation of T_e, T_i, and p plasma parameters and knowing a priori N_e from the Plasma Line at different signal fluctuation values (δ(%)).

5.4.3. ESTIMATED ERRORS OF TWO PARAMETERS (T_i AND p) GIVEN *A PRIORI* N_e AND T_e/T_i

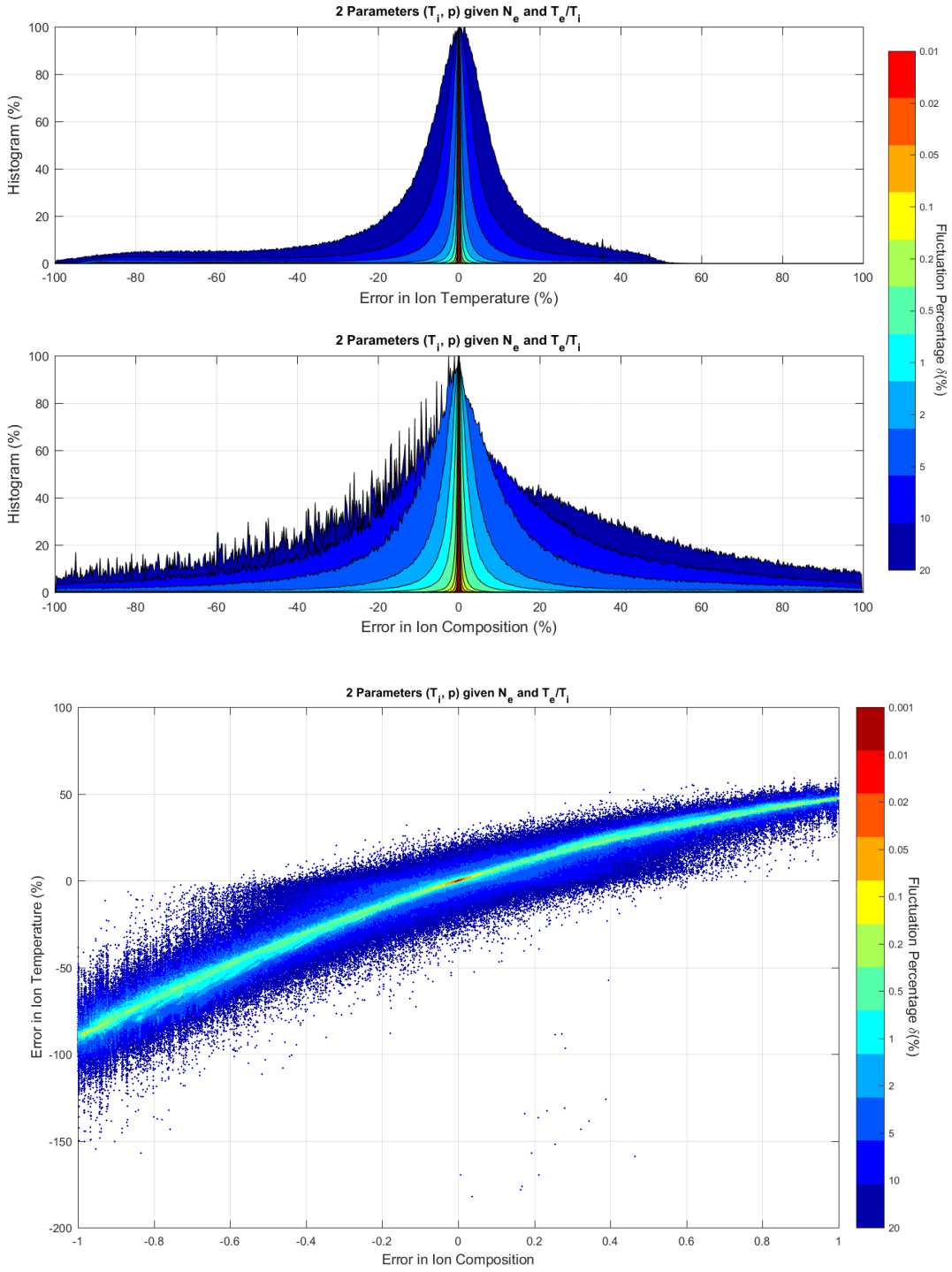

Figure 5.8. (Top) Histogram of estimated parameters errors, and (bottom) relationship between plasma parameters errors (in percentage) and the ion composition error. These results were obtained from the estimation of T_i and p plasma parameters and knowing a priori N_e and T_e/T_i from the Plasma Line at different signal fluctuation values ($\delta(\%)$).

5.4.4. ESTIMATED ERRORS OF TWO PARAMETERS (T_i AND p) GIVEN *A PRIORI* N_e AND T_e

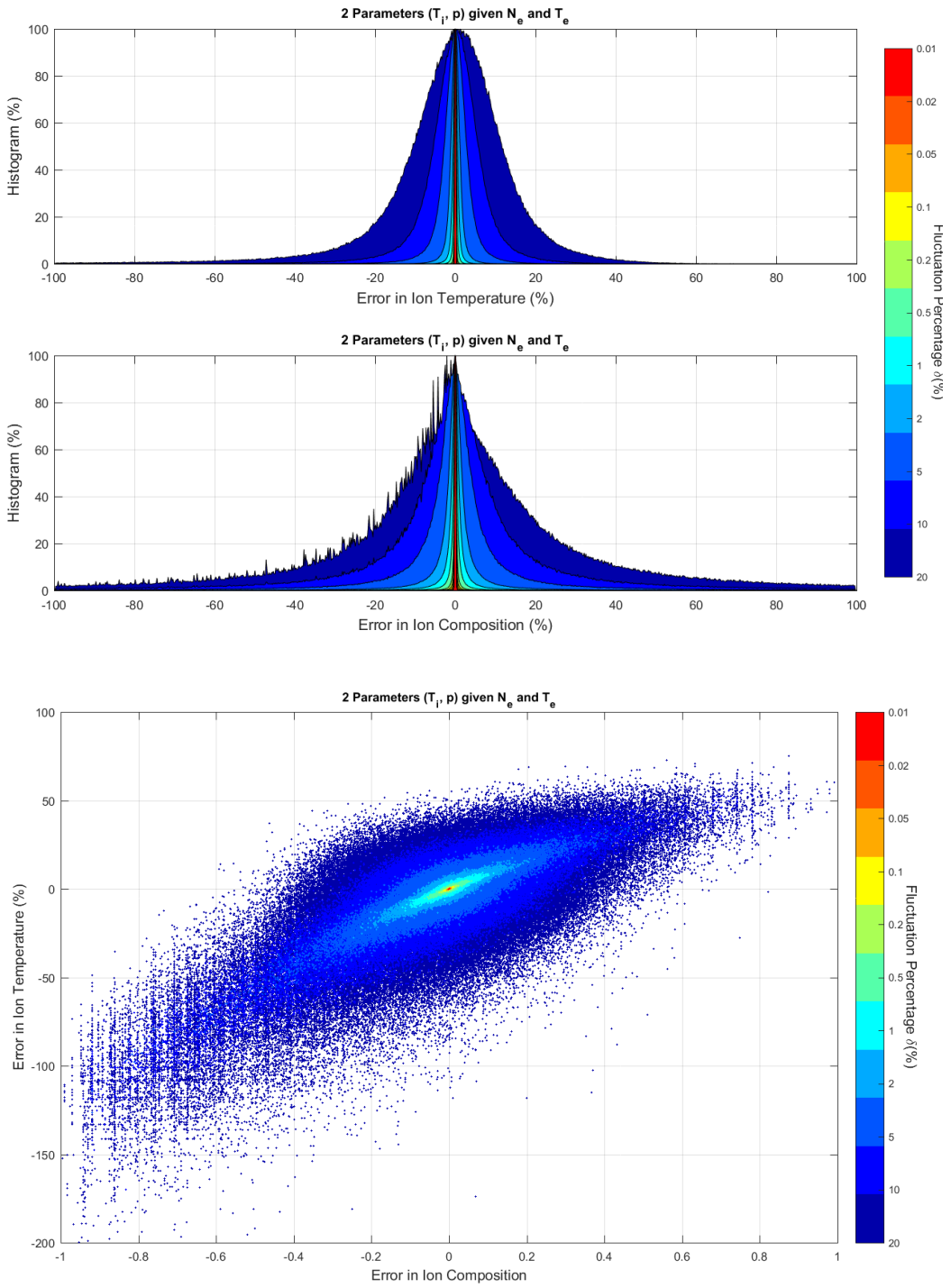

Figure 5.9. (Top) Histogram of estimated parameters errors, and (bottom) relationship between plasma parameters errors (in percentage) and the ion composition error. These results were obtained from the estimation of T_i and p plasma parameters and knowing a priori N_e and T_e from the Plasma Line at different signal fluctuation values ($\delta(\%)$).

5.4.5. VERIFICATION OF WALDTEUFEL ANALYTICAL FORMULATION

Parameters estimation error percentages ($\mathcal{E}_X(\%) = 100 \cdot (X_{true} - \hat{X})/X_{true}$) obtained from our Monte Carlo simulations were compared to the theoretical errors calculated with the analytic formulation of [Waldteufel, 1971]. Figure 5.10 shows the theoretical estimation errors obtained by the direct application of Equation (10) of [Zettergren et al., 2011]. This figure shows that both errors were similar, indicating that the approximation made by [Waldteufel, 1971] provides good error predictions. Nevertheless, electron temperature and electron density errors were much more disperse than those predicted by the analytical formulation. Regression formulas are provided in Figure 5.10 for each estimated parameter error obtained from simulations with different number of plasma parameters fitted.

The values shown in Figure 5.10 agree with the ion temperature increase calculated by [Zhang et al., 2018] in the case of positive ion composition errors ($\mathcal{E}_p > 0$). Nevertheless, negative ion composition errors obtained larger ion temperature errors. A more detailed study should be done to determine the differences found between the ISR and ionosonde f_0F_1 methods in long-term exospheric temperatures variations [Perrone & Mikhailov, 2017] [Zhang et al., 2018] [Perrone & Mikhailov, 2018].

Estimated errors of T_e and T_i parameters shown in Figure 5.10 were approximately -100% and 50% for ion composition error values of $\mathcal{E}_p = -1$ and $\mathcal{E}_p = 1$, respectively. This indicates that the maximum erroneously estimated values of temperature were approximately $\hat{T}_{incorrect}(\mathcal{E}_p = -1\,) = 2T_{correct}$ and $\hat{T}_{incorrect}(\mathcal{E}_p = 1\,) = T_{correct}/2$.

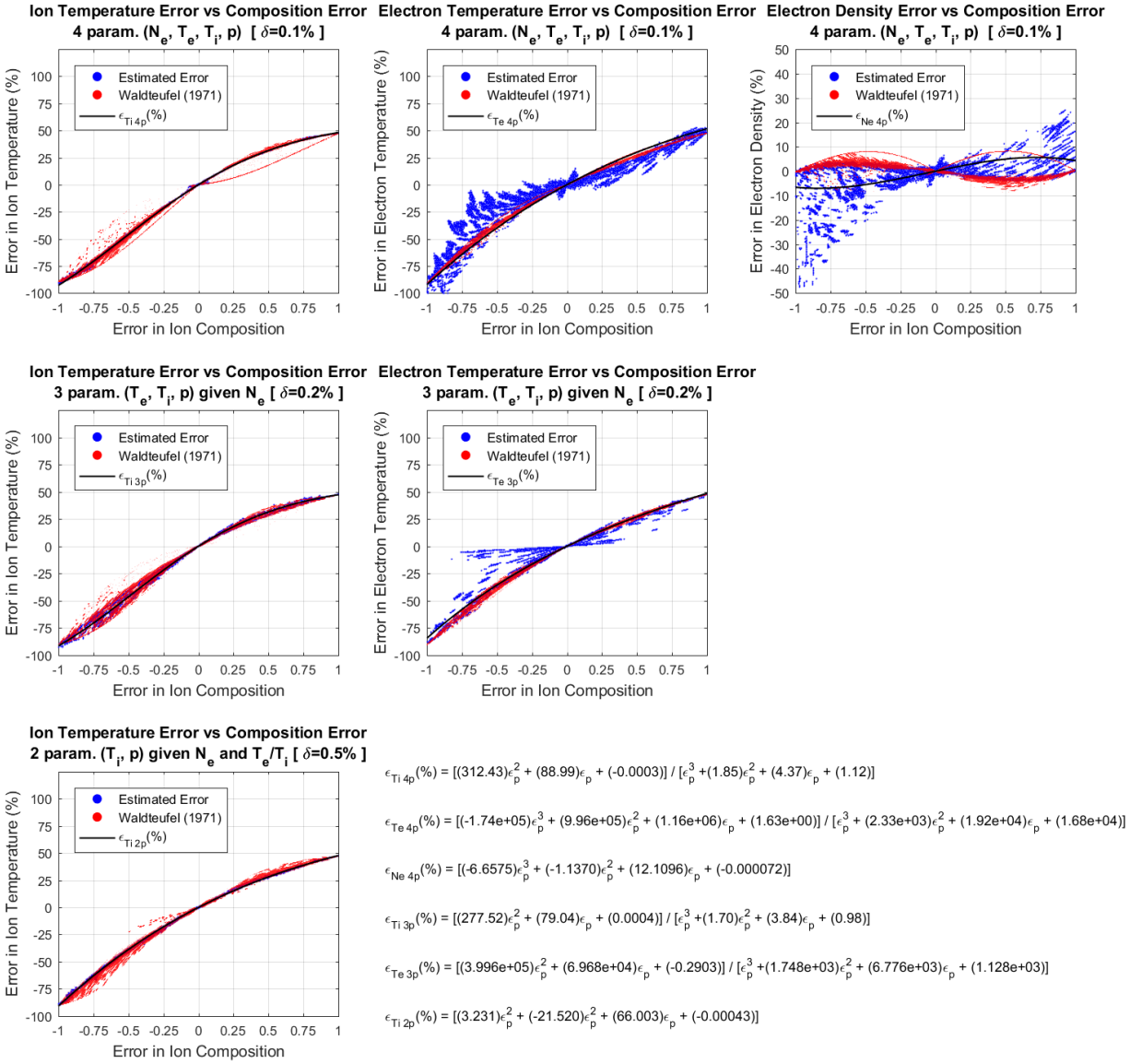

Figure 5.10. (Blue) The relationship between plasma parameters error percentage ($\mathcal{E}_x(\%) = 100(X_{true} - \hat{X})/X_{true}$) and ion composition estimation error ($\mathcal{E}_p = p_{true} - \hat{p}$), (black line) the estimated regression of errors, and (red) the theoretical error of plasma parameters errors that should be obtained by applying the formulations of [Waldteufel, 1971] to the ion composition errors obtained. Different rows show simulation results of (top) four parameters at $\delta = 0.1\%$, (middle) three parameters given N_e at $\delta = 0.2\%$, and (bottom) two parameters given N_e and T_e/T_i at $\delta = 0.5\%$. Estimated regression formulas computed using non-linear robust regression are shown at bottom row.

CHAPTER 6

DETERMINATION OF THE MOST RELEVANT A-PRIORI KNOWN PARAMETERS

To determine which *a priori* known parameter would be the most relevant to solve the ambiguity problem, the amount of information provided by each plasma parameter must be identified. Monte Carlo simulations of 2000 true input plasma parameters were done assuming the *a priori* knowledge of each plasma parameter at different signal fluctuation levels. The different study cases analyzed were:

i) 5 unknown parameters (N_e, T_i, T_e, V_i, and p) without *a priori* information;
ii) 4 unknown parameters (N_e, T_i, T_e, and p) given *a priori* V_i;
iii) 4 unknown parameters (T_i, T_e, V_i, and p) given *a priori* N_e;
iv) 4 unknown parameters (N_e, T_i, V_i, and p) given *a priori* T_e/T_i;
v) 4 unknown parameters (N_e, T_i, V_i, and p) given *a priori* T_e; and
vi) 4 unknown parameters (N_e, T_e, V_i, and p) given *a priori* T_i.

The study case *ii* (assuming V_i is known *a priori*) has been already analyzed in the Chapter 5 as study case *a*. Also, the study case *iii* (assuming N_e is known *a priori*) is similar to the study case *b* from Chapter 5, but without considering known the V_i parameter. It is relevant to review study case *iii* separately because solutions may depend on the number of parameters to estimate.

Previous studies, [Vallinkoski, 1988] and [Vallinkoski & Lehtinen, 1990] analyzed the effects of providing *a priori* plasma parameters assuming no collisions, a fixed value of ion composition, and signal fluctuations of $\delta = 1\%$. Results from these studies indicated that the addition of *a priori* parameters are useful only when parameters have highly correlated errors with ion composition (p). As a negative correlation coefficient exists between T_i and p parameters, the addition of T_i information completely solved the ambiguity. In these studies ([Vallinkoski, 1988] and [Vallinkoski & Lehtinen, 1990]), the unambiguous estimation was found to be dependent on the signal fluctuation level.

6.1. AMBIGUOUS ION COMPOSITION ESTIMATE

Figure 6.1 shows the estimated ion composition values obtained for all the study cases previously indicated. The different colors of this figure represent values estimated at different signal fluctuations (δ), similarly as in Figure 5.1.

No differences were found on the estimated ion composition values of simulations without *a priori* information and simulations assuming the ion drift (V_i) known *a priori*. This result implies that the knowledge of the V_i parameter does not modify the ion composition estimate.

The addition of ion temperature (T_i) information obtained no ambiguous estimation in any of the simulation results shown in Figure 6.1. This later result is in agreement with the results of [Vallinkoski & Lehtinen, 1990], verifying the high impact generated by information of highly correlated parameters. Also, the addition of electron temperature (T_e) information obtained most of the ion composition solutions 'correct' (i.e. in the vicinity of $\hat{p} \approx p_{true}$), but some solutions were obtained with larger uncertainty than in the case of adding T_i information.

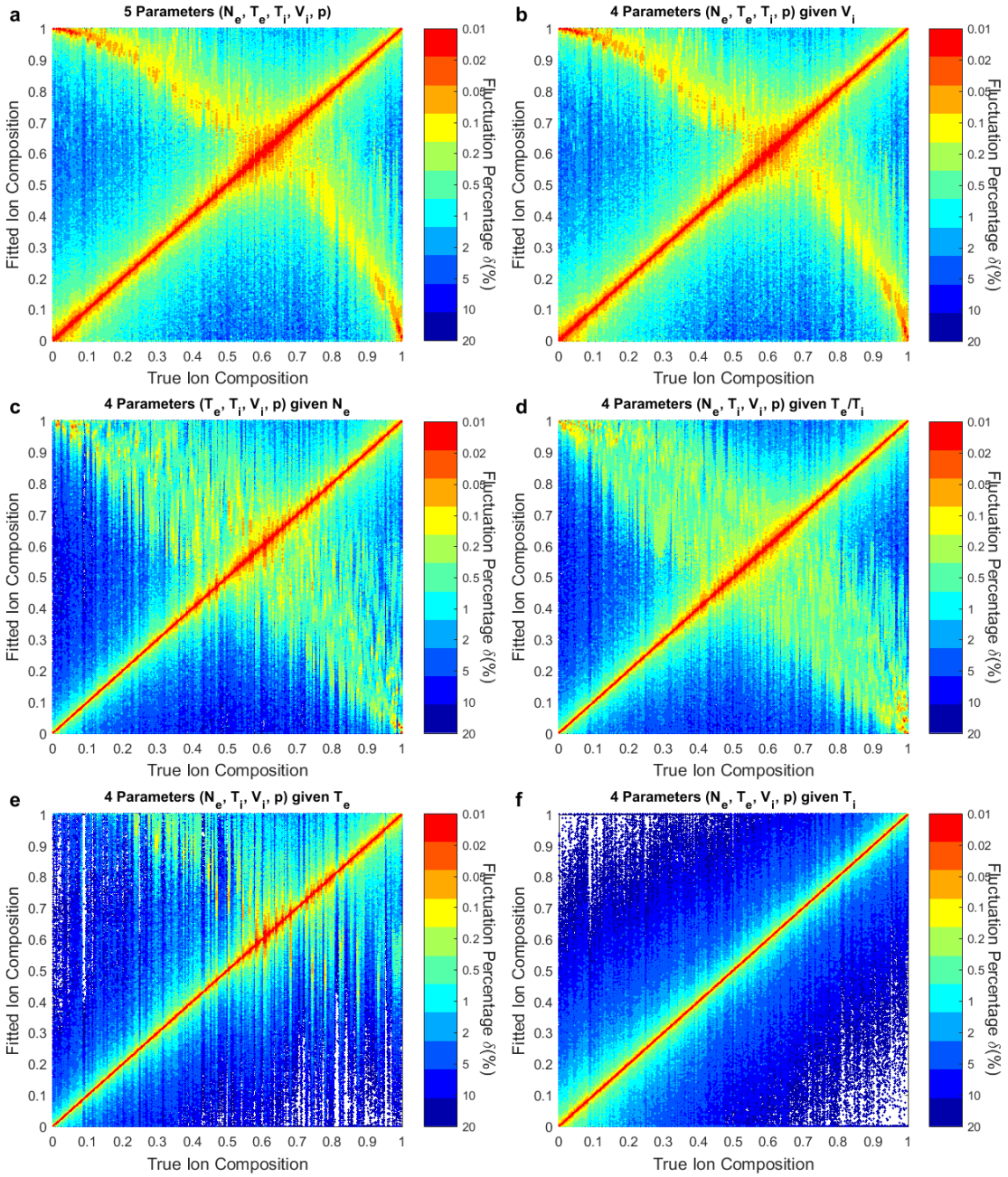

Figure 6.1. Scatter plot of estimated and input values of ion composition obtained from the analysis of plasma parameters with a parameter known a priori: (a) without a priori information, (b) given V_i, (c) given N_e, (d) given T_e/T_i, (e) given T_e, and (f) given T_i. Each color represents results obtained with simulations with a particular signal fluctuation percentage ($\delta(\%)$).

6.2. UNCERTAINTY OF THE ION COMPOSITION ESTIMATE

Figure 6.2 shows the standard deviation values obtained for all the study cases indicated. This figure shows that the estimated uncertainties of ion composition (σ_{est}) for study cases *i*, *ii*, *iii*, *iv*, *v*, and *vi* were $\delta/10$, $\delta/10$, $\delta/30$, $\delta/25$, $\delta/60$, and $\delta/60$, respectively. These values indicate that the addition of temperatures (T_e and T_i) provide a similar amount of information.

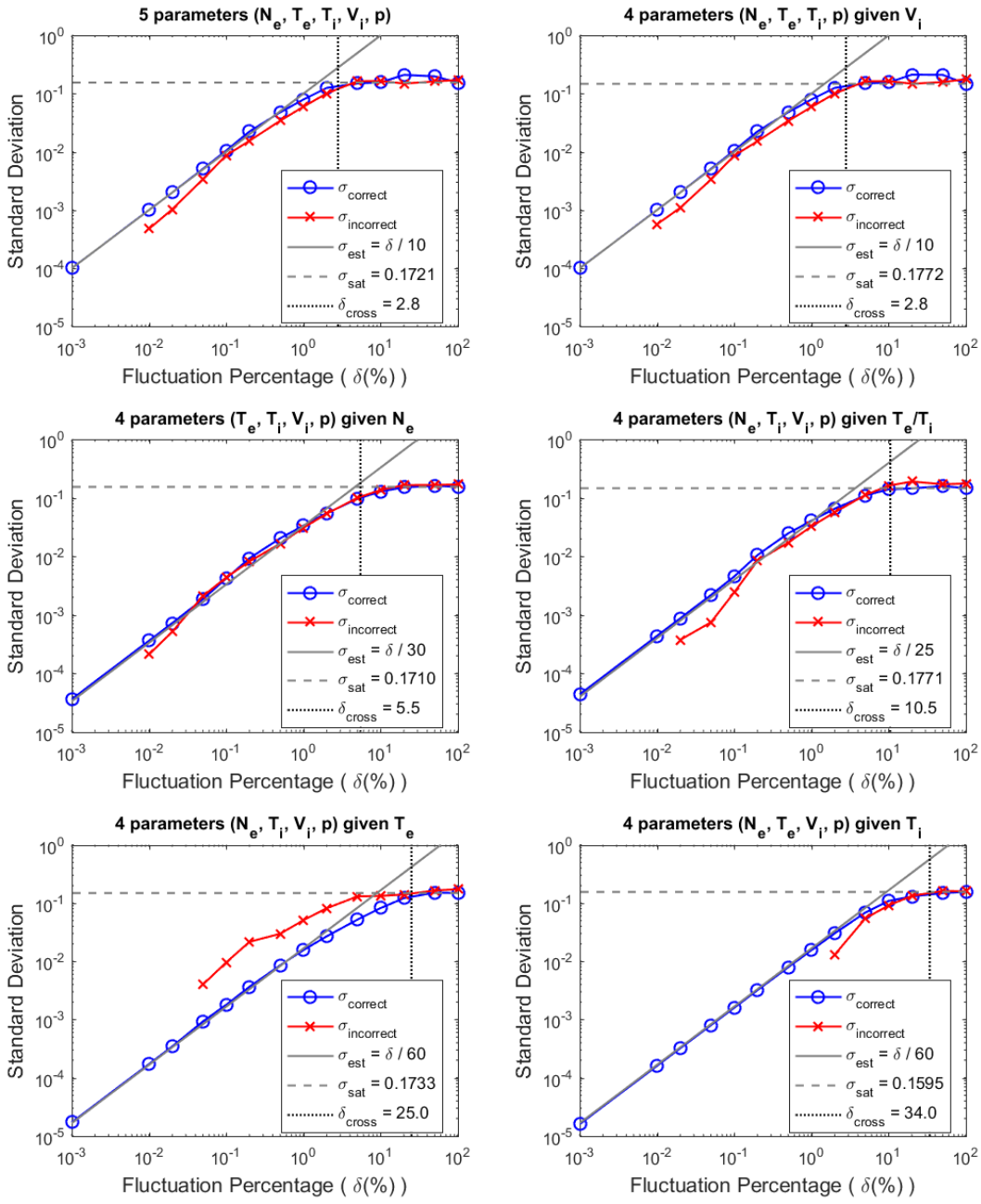

Figure 6.2. Average values of standard deviation (in logarithmical scale) of 'correct' (blue) and 'incorrect' (red) statistical distributions obtained with simulations of the estimation of different known a priori plasma parameters at different signal fluctuation percentages ($\delta(\%)$). Vertical dotted line represents the estimated fluctuation value (δ_{cross}) at which the standard deviation reaches its maximum value (σ_{sat}). Black line represents the estimated linear regression (σ_{est}) previous to saturation, and the horizontal dotted line represents the saturation value (σ_{sat}).

6.3. PROBABILITY RESULTS OF SINGLE PARAMETER ADDITION

Figure 6.3 shows the probabilities obtained with the simulations of the study cases previously indicated. The most relevant outcome is that study cases *i* and *ii* (without *a priori* information and knowing *a priori* V_i parameter, respectively) obtained almost identical probabilities. This result implies the knowledge of the V_i parameter does not affect the ion composition estimate, and consequently it does not contribute to solve the TICA problem. Nevertheless, at very high signal fluctuations ($\delta > 10\%$), small differences of $P_{correct}$ (<3%) were found. These probability differences found at very high signal fluctuations were assumed to be related to the estimate variability generated by the addition of different random noise at each simulation. Furthermore, study case *iii* (knowing *a priori* N_e) and study case *b* of Chapter 5 (knowing *a priori* N_e and V_i, shown in Figure 5.3) obtained similar probabilities. This latter result verifies that the parameter V_i does not provide information to solve the ambiguity.

Results of simulations of study case *iii* (having the *a priori* knowledge of N_e) obtained an almost constant increase of $P_{correct}$, but simulations of study case *iv* (with the *a priori* knowledge of T_e/T_i) had no increase of $P_{correct}$ at some signal fluctuation levels. This effect implies that the knowledge of N_e provided more information than the knowledge of T_e/T_i. Even so, both simulations obtained similar signal fluctuation thresholds for unambiguous estimation ($P_{correct} \geq 95.45\%$) at $\delta_{th\,(N_e)} = 0.14\%$ and $\delta_{th\,(T_e/T_i)} = 0.11\%$ for study cases *iii* and *iv*, respectively. Nevertheless, $P_{correct}$ values of study case *c* (knowing *a priori* N_e and T_e/T_i, shown in Figure 5.3) were much higher than the obtained values from study case *iii* or *iv* (knowing *a priori* N_e or T_e/T_i parameters, respectively). This latter result implies that the information provided by N_e and T_e/T_i parameters was different and complementary. Alternatively, $P_{fit\,valid}$ of study case *iv* was the smallest. It is assumed that this decrease in convergence was related to the difficulty of obtaining values of T_e and T_i that satisfy the T_e/T_i ratio imposed by the *a priori* knowledge during the estimation process.

Study cases *v* and *vi* (knowing *a priori* T_e and T_i, respectively) highly improved the values of $P_{correct}$. Signal fluctuation thresholds for those cases were $\delta_{th(T_e)} = 1.54\%$ and $\delta_{th(T_i)} = 3.06\%$. As these signal fluctuation values are commonly obtained in real ISR measurements [Vallinkoski & Lehtinen, 1990], common ISR radars could avoid the TICA problem by the direct addition of plasma temperature information. Both study cases obtained similar values of $P_{correct}$, but presented different shapes at highly noisy scenarios. The probability curve of study case *vi* had a more pronounced decay as a function of signal fluctuation. Convergences obtained were $P_{fit\,valid(T_e)} > 96\%$ and $P_{fit\,valid(T_i)} > 98\%$. Consequently, the *a priori* knowledge of T_i provided the most relevant information for solving the TICA problem, as indicated by [Vallinkoski, 1988] and [Vallinkoski & Lehtinen, 1990]. Nevertheless, the *a priori* knowledge of T_e obtained similar probabilities to the results of knowing *a priori* T_i.

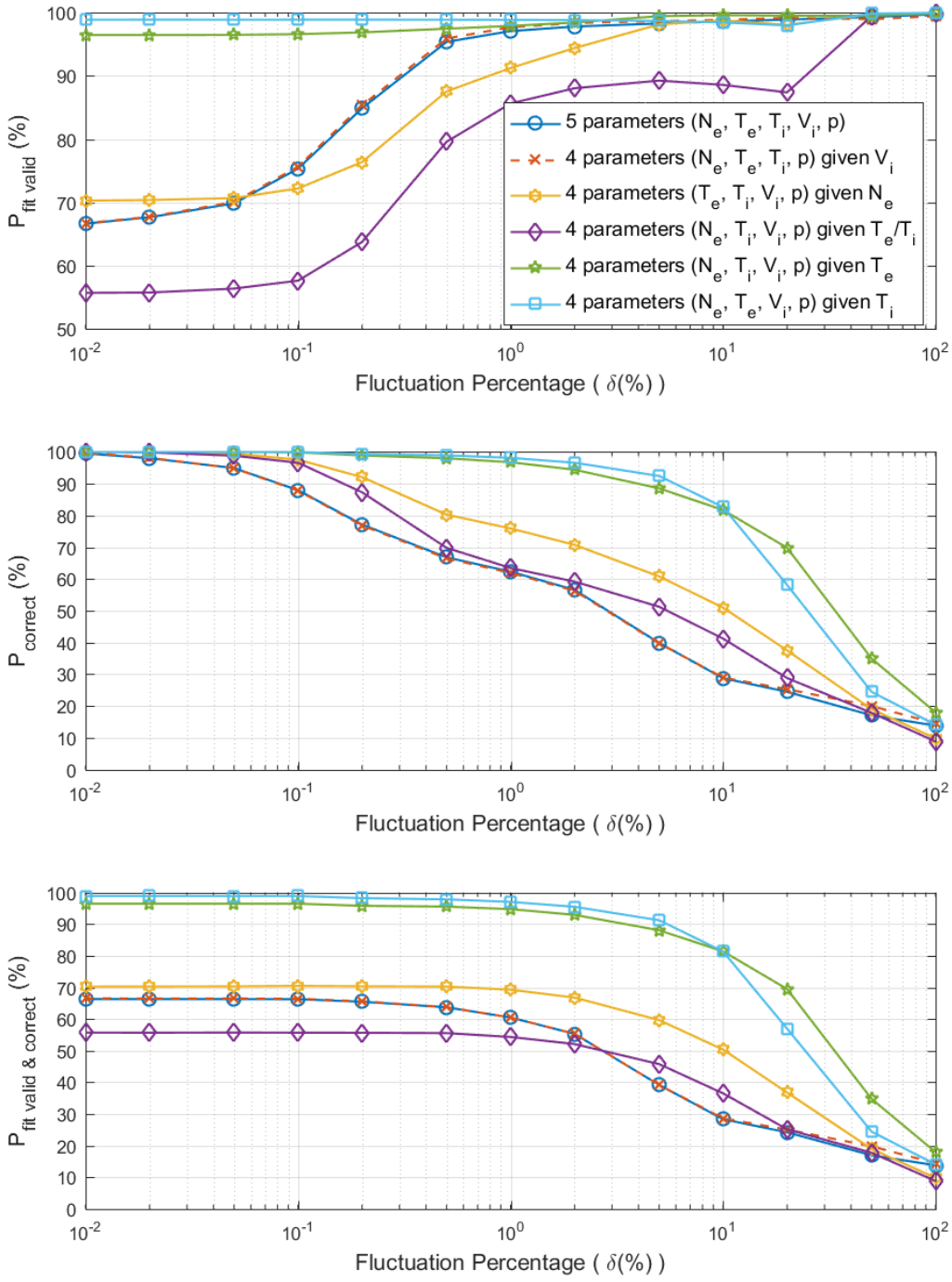

Figure 6.3. Probability of convergence ($P_{fit\ valid}$) and probability of 'correct' estimation ($P_{correct}$) (in percentage) at different signal fluctuation percentages ($\delta(\%)$) obtained from the analysis of plasma parameters without a priori information (blue circles), given V_i (orange crosses dotted line), given N_e (yellow hexagons), given T_e/T_i (purple rhombs), given T_e (green stars), and given T_i (cyan squares).

6.4. RELATIONSHIPS BETWEEN ESTIMATION ERRORS OF PARAMETERS

Histograms of parameters estimate errors and their relationships with the ion composition estimate error ($\mathcal{E}_p$) are shown in Figure 6.4, Figure 6.5, Figure 6.6, Figure 6.7, Figure 6.8, and Figure 6.9. These graphics show correlations between T_e and T_i estimation errors and the ion composition error, verifying the high impact of parameter information with highly correlated errors [Vallinkoski & Lehtinen, 1990].

6.4.1. ESTIMATED ERRORS OF FIVE PARAMETERS (N_e, T_e, T_i, V_i, p)

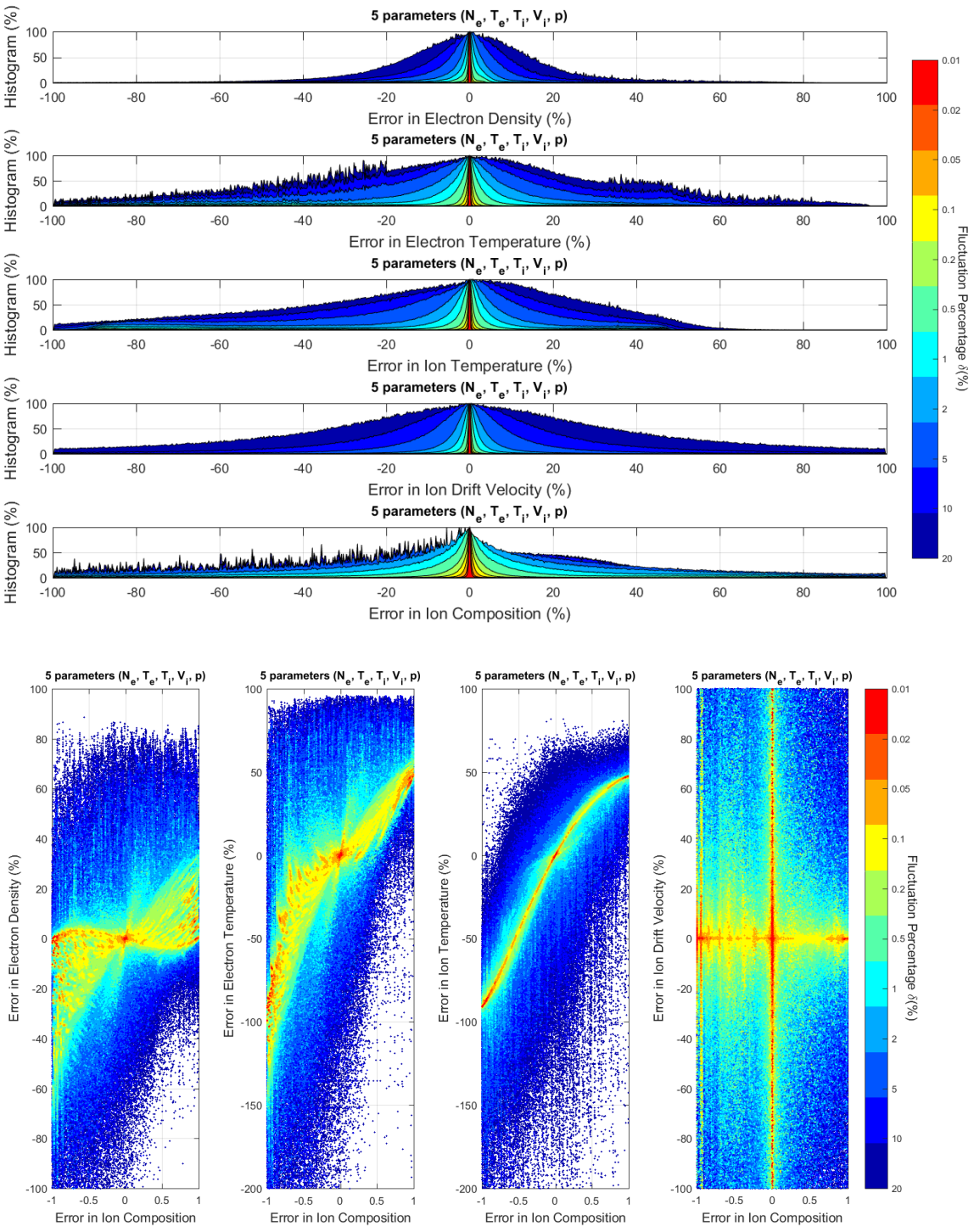

Figure 6.4. (Top) Histogram of estimated parameters errors, and (bottom) Scatter plot of parameters errors and ion composition error, obtained from the estimation of five plasma parameters (N_e, T_e, T_i, V_i, p) without a priori information, at different signal fluctuation values ($\delta(\%)$).

6.4.2. ESTIMATED ERRORS OF FOUR PARAMETERS (N_e, T_e, T_i, AND p) GIVEN A PRIORI V_i

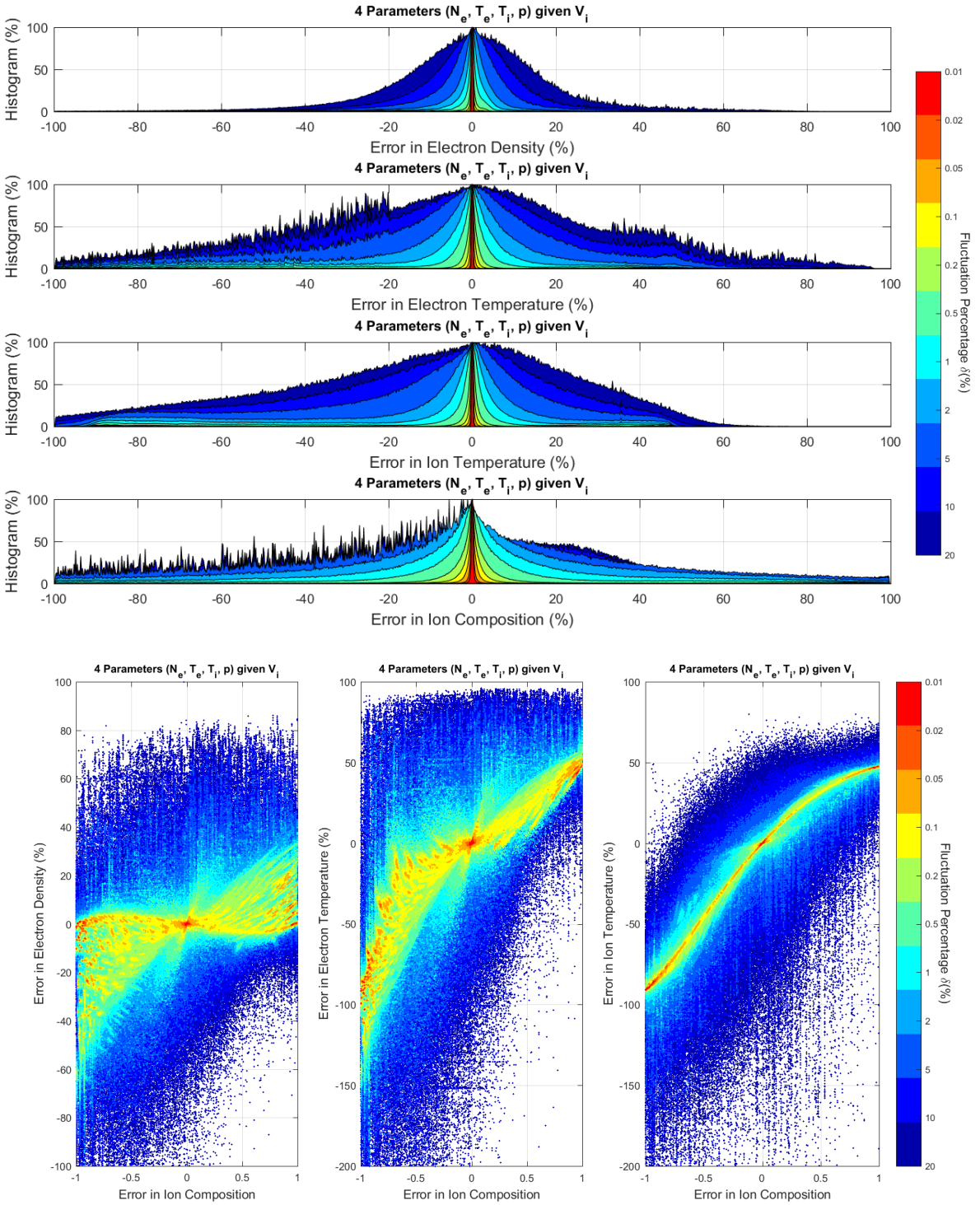

Figure 6.5. (Top) Histogram of estimated parameters errors, and (bottom) Scatter plot of parameters errors and ion composition error, obtained from the estimation of four plasma parameters (N_e, T_e, T_i, p) given a priori V_i, at different signal fluctuation values ($\delta(\%)$).

6.4.3. ESTIMATED ERRORS OF FOUR PARAMETERS (T_e, T_i, V_i, AND p) GIVEN A PRIORI N_e

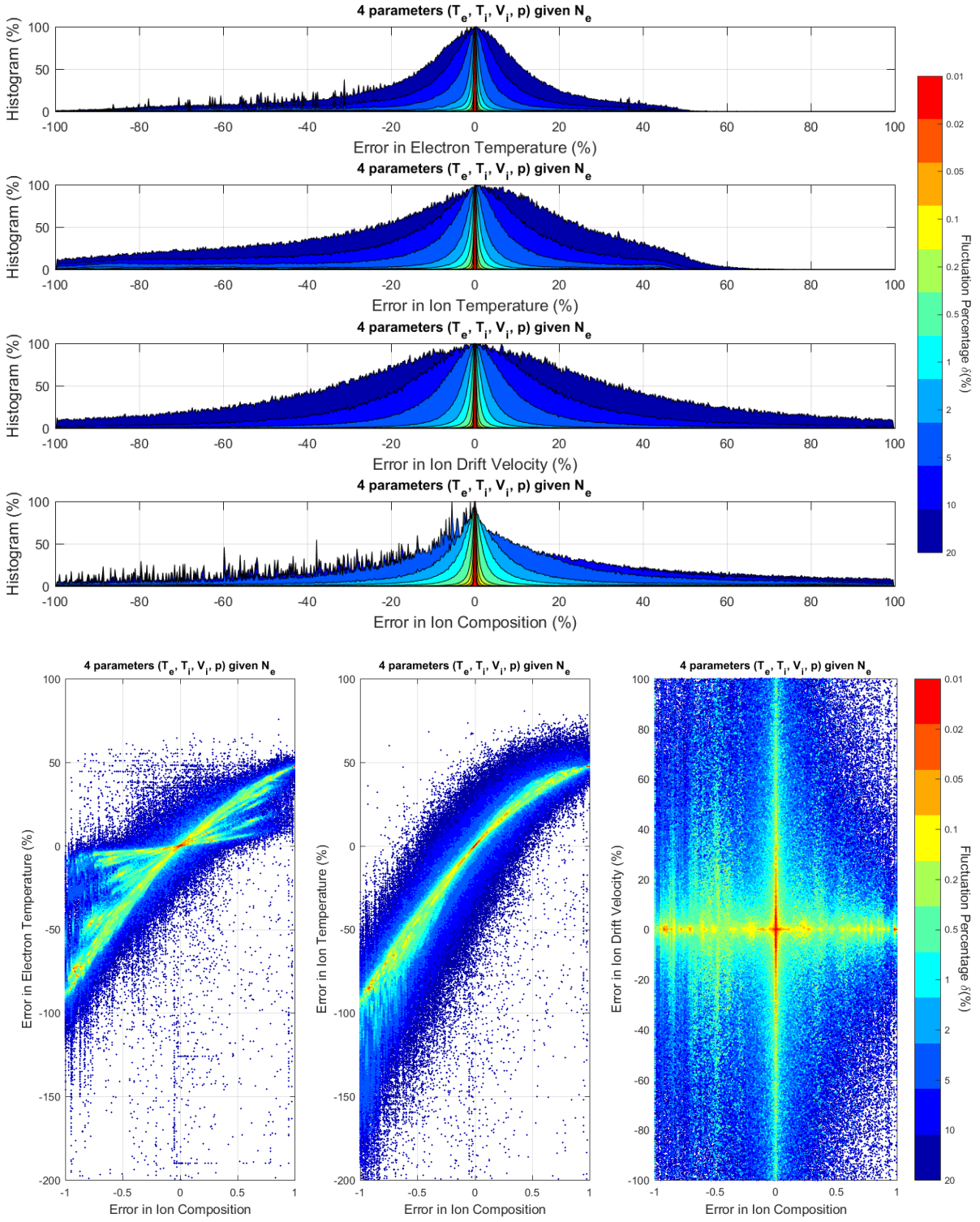

Figure 6.6. (Top) Histogram of estimated parameters errors, and (bottom) Scatter plot of parameters errors and ion composition error, obtained from the estimation of four plasma parameters (T_e, T_i, V_i, p) given a priori N_e, at different signal fluctuation values ($\delta(\%)$).

6.4.4. ESTIMATED ERRORS OF FOUR PARAMETERS (N_e, T_i, V_i, AND p) GIVEN A PRIORI T_e/T_i

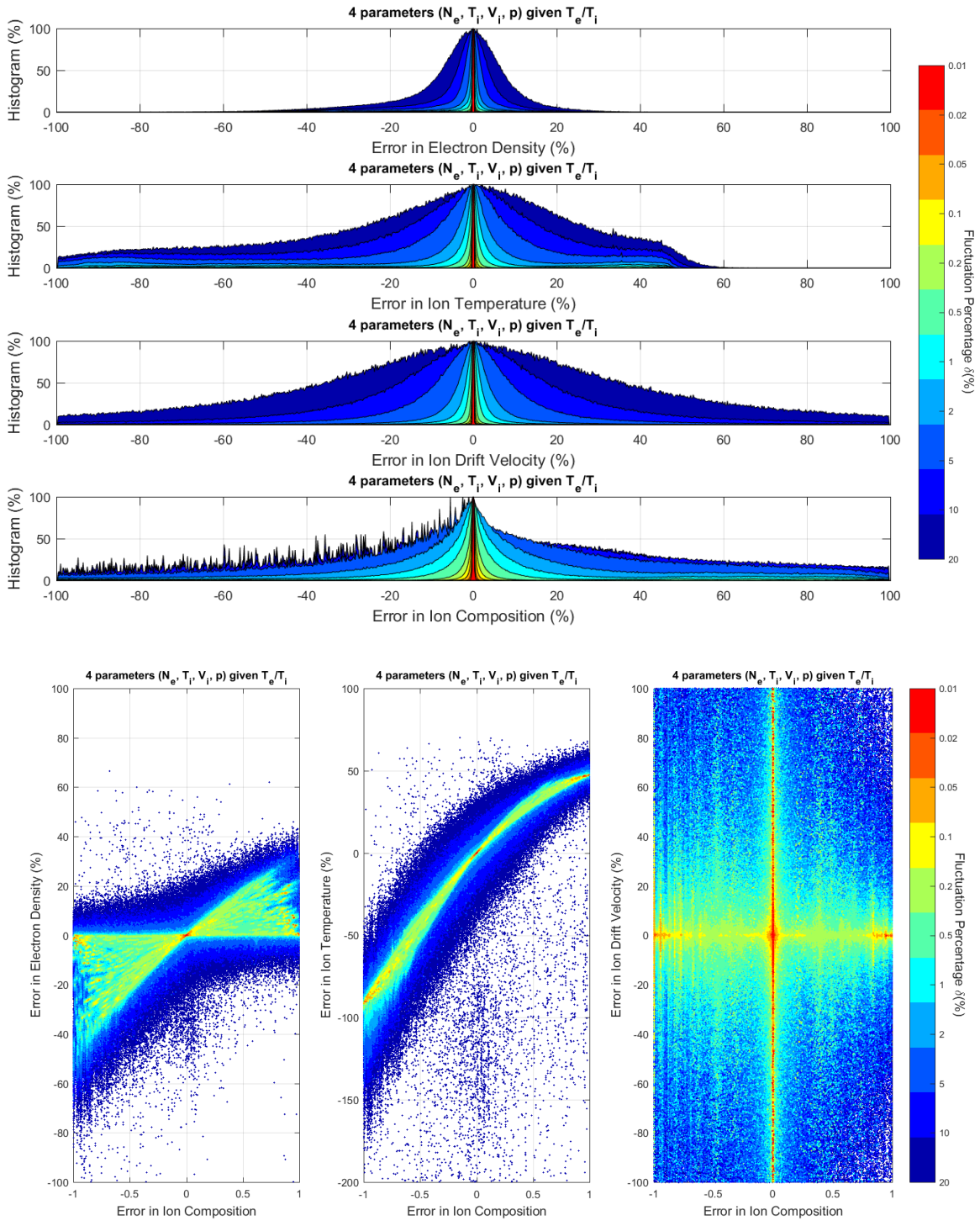

Figure 6.7. (Top) Histogram of estimated parameters errors, and (bottom) Scatter plot of parameters errors and ion composition error, obtained from the estimation of four plasma parameters (N_e, T_i, V_i, p) given a priori T_e/T_i, at different signal fluctuation values ($\delta(\%)$).

6.4.5. ESTIMATED ERRORS OF FOUR PARAMETERS (N_e, T_i, V_i, AND p) GIVEN *A PRIORI* T_e

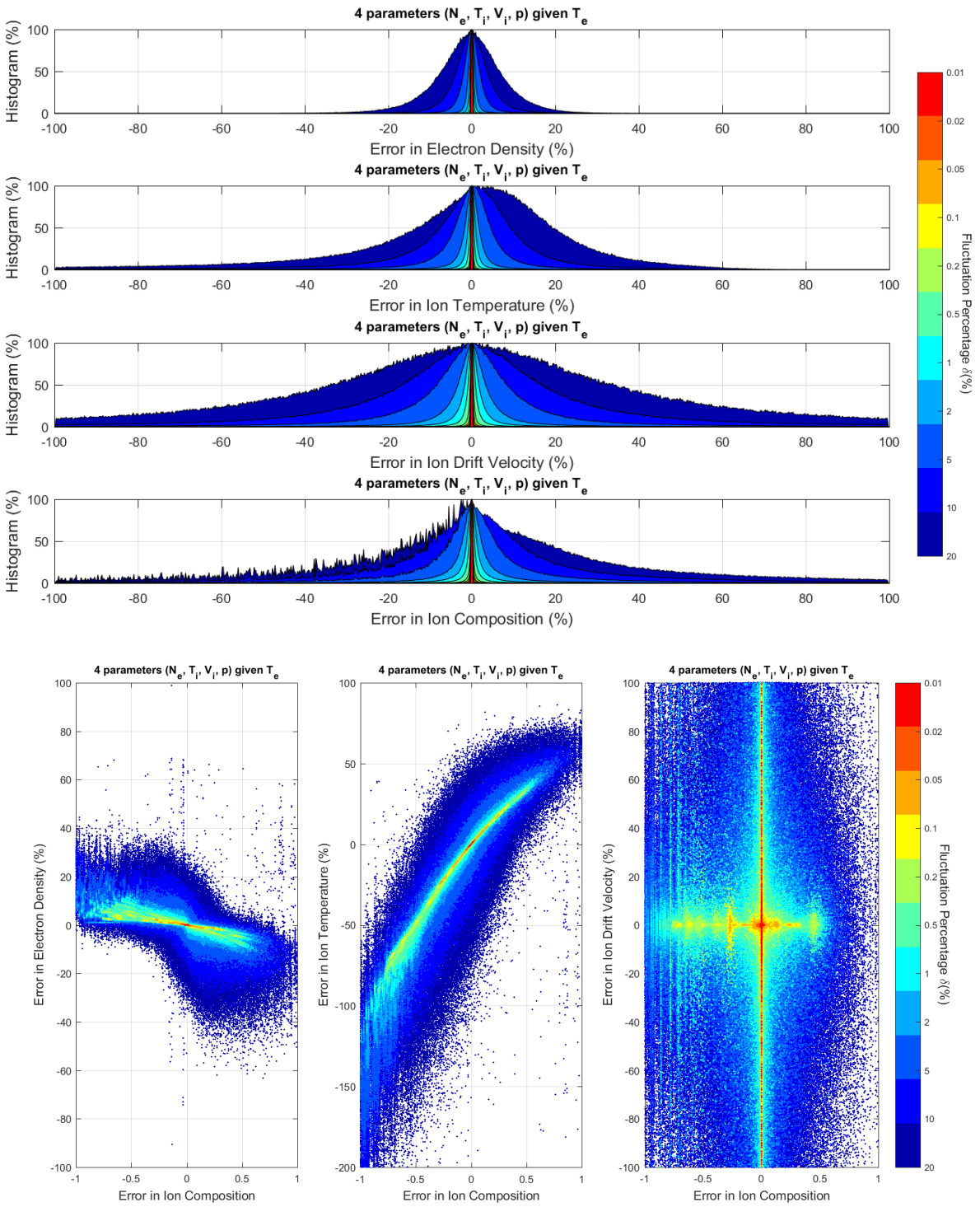

Figure 6.8. (Top) Histogram of estimated parameters errors, and (bottom) Scatter plot of parameters errors and ion composition error, obtained from the estimation of four plasma parameters (N_e, T_i, V_i, p) given a priori T_e, at different signal fluctuation values ($\delta(\%)$).

6.4.6. ESTIMATED ERRORS OF FOUR PARAMETERS (N_e, T_e, V_i, AND p) GIVEN *A PRIORI* T_i

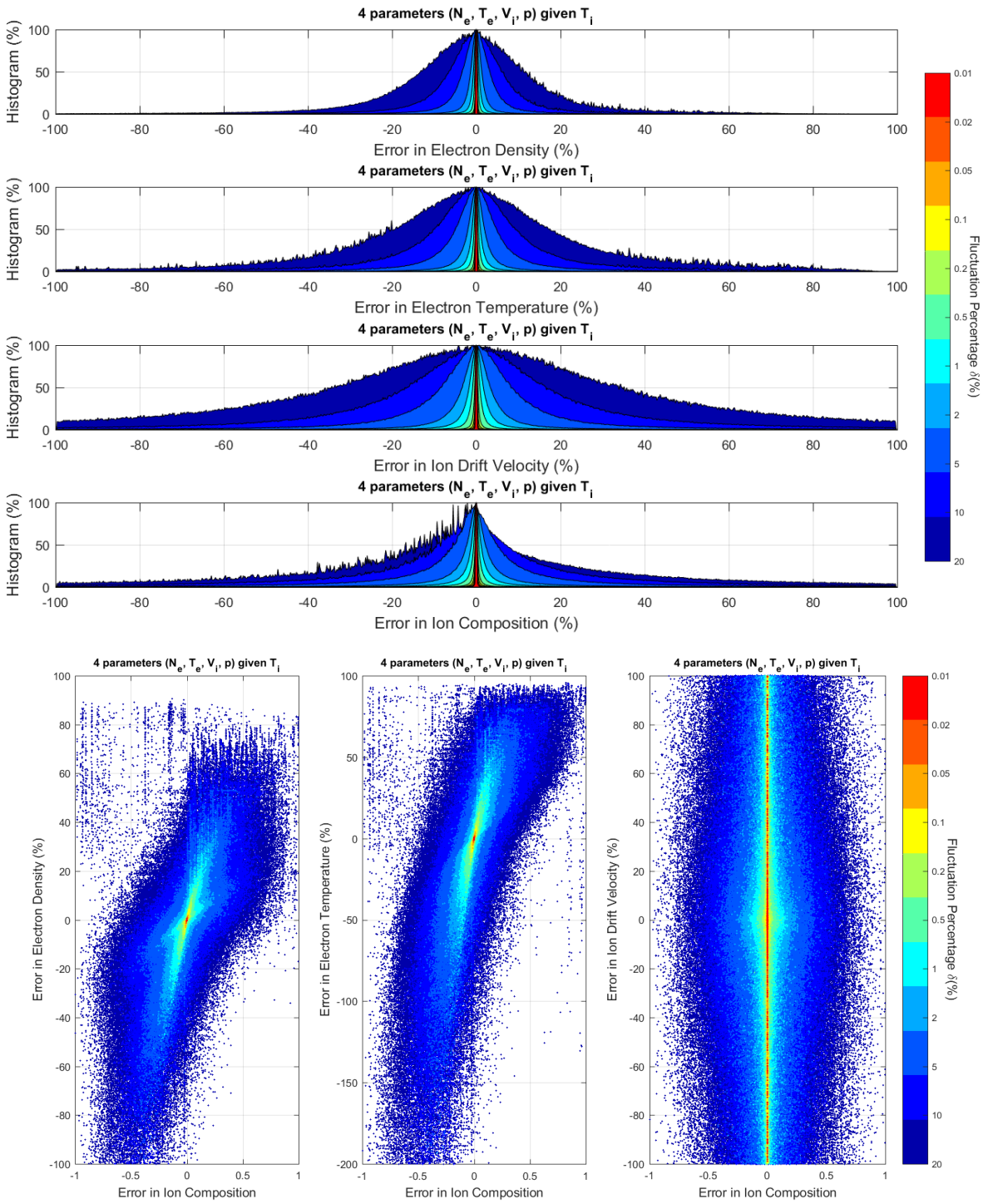

Figure 6.9. (Top) Histogram of estimated parameters errors, and (bottom) Scatter plot of parameters errors and ion composition error, obtained from the estimation of four plasma parameters (N_e, T_i, V_i, p) given a priori T_i, at different signal fluctuation values ($\delta(\%)$).

CHAPTER 7

DETERMINATION OF THE IMPACT OF THE UNCERTAINTY OF *A-PRIORI* KNOWN PARAMETERS

To provide a more realistic radar framework, different levels of uncertainty in the plasma parameters known *a priori* from the Plasma Line have been analyzed. The study of [Vallinkoski & Lehtinen, 1990] analyzed the effect of the level of uncertainty of the *a priori* given parameters. Results from this study demonstrated the dependence of the estimation error on the a priori accuracy. Uncertainty levels analyzed in our study were $\epsilon = \pm 0.05\%$, $\pm 0.1\%$, $\pm 0.5\%$, $\pm 1\%$, $\pm 5\%$, $\pm 10\%$, $\pm 25\%$, $\pm 50\%$, and $\pm 100\%$.

Monte Carlo simulations of 1000 different true input parameters were done to analyze the impact of adding Plasma Line information with uncertainty, as indicated in Chapter 3.1.9. Simulation of Uncertainty of *a-priori* Known Parameters. The different combinations of parameters studied were:

1) 3 unknown parameters (T_i, T_e, and p) given a priori N_e;
2) 2 unknown parameters (T_i and p) given a priori N_e and T_e/T_i; and
3) 2 unknown parameters (T_i and p) given a priori N_e and T_e.

7.1. PROBABILITY RESULTS OF PLASMA LINE INFORMATION ADDITION WITH UNCERTAINTY

Figure 7.1 shows the probabilities obtained from simulations with different levels of uncertainty in the addition of Plasma Line information. The probabilities of simulations without uncertainty (Figure 5.3) are shown for visual comparison in Figure 7.1 as black dotted lines. As expected, when the uncertainty level was reduced, the probabilities obtained were similar to the probabilities calculated without uncertainty in the *a priori* known parameters.

A relevant result of these simulations is that $P_{fit\ valid}$ values were similar to the results obtained without uncertainty at fluctuation percentage values larger than the uncertainty level ($\delta \geq |\epsilon|$), but at lower fluctuation levels the values of $P_{fit\ valid}$ rapidly decayed to zero when reducing the signal fluctuation value. It is assumed that simulations with $\delta \ll |\epsilon|$ did not converge because the *a priori* given parameters of these simulations were not a valid solution of the estimation problem. This result agrees with the results of [Vallinkoski & Lehtinen, 1990] that demonstrate the existence of maximum and minimum signal fluctuation levels to obtain valid solutions when providing a priori known parameters with uncertainty.

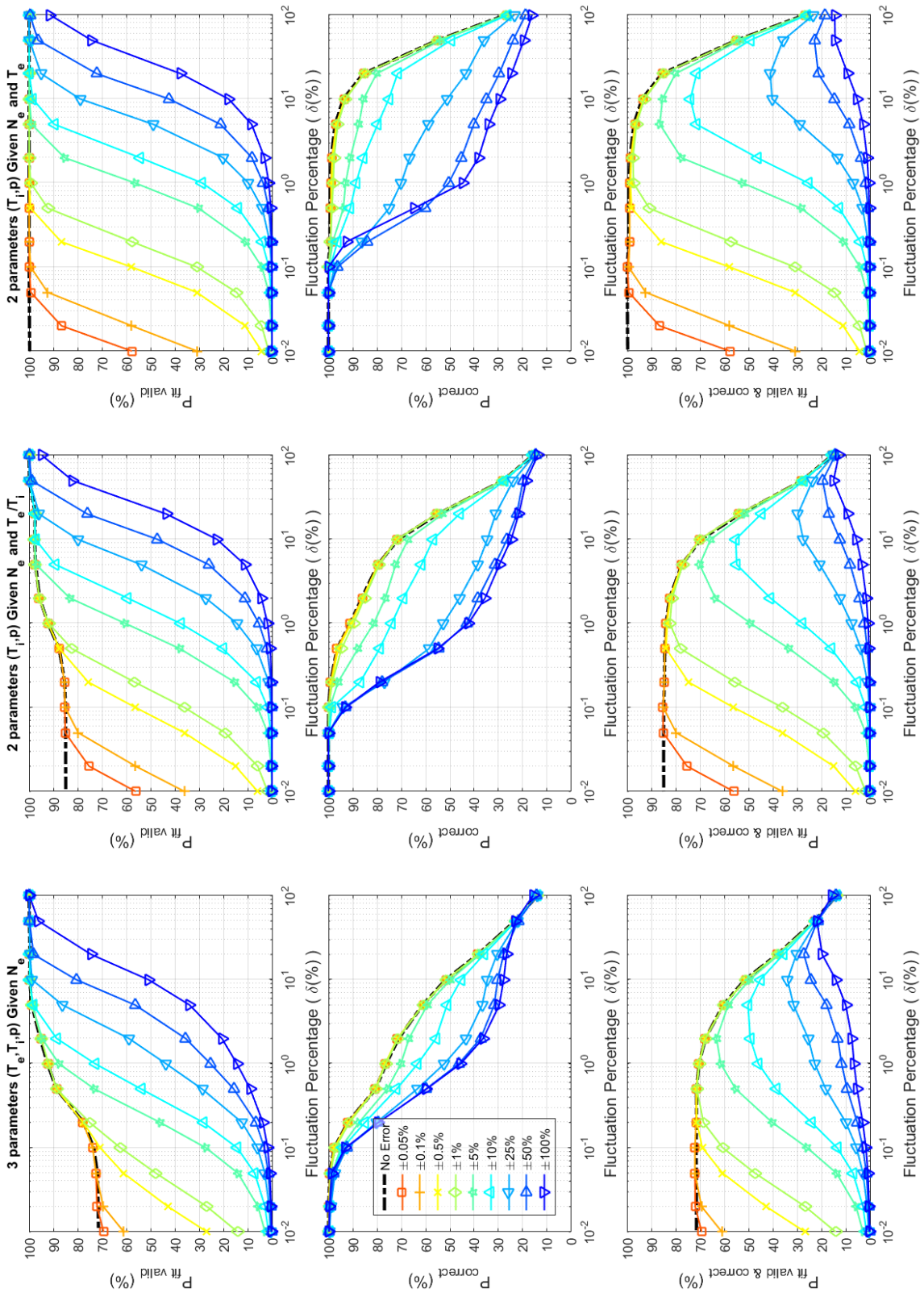

Figure 7.1. (90° Left Rotated) Probabilities obtained from the analysis of plasma parameters with a priori information having different levels of uncertainty. Columns correspond to simulations with different combinations of known a priori plasma parameters from the Plasma Line: given N_e (left), given N_e and T_e/T_i (middle), given N_e and T_e (right). Black dotted lines represent simulation results with plasma parameters known a priori without uncertainty (Figure 5.3).

Therefore, to avoid wrongly estimated results when providing *a priori* information, verifying the convergence of the solutions by using the method indicated in Chapter 3.2.1 is strongly suggested. Furthermore, verifying that the uncertainty of the known *a priori* measurements is higher or equal than the fluctuation level of the ISR signal to estimate ($\delta \geq |\epsilon|$) is also suggested.

Results of simulations with uncertainty levels $|\epsilon| \leq 1\%$ obtained values of $P_{correct}$ similar to simulations without uncertainty (black dotted lines). This suggests that knowing *a priori* information with uncertainty levels $\leq 1\%$ would solve the ambiguity problem similarly as if these *a priori* known parameters were deterministically known. Results with larger uncertainty levels ($|\epsilon| > 1\%$) obtained a decrease of $P_{correct}$ dependent on the uncertainty level.

Alternatively, for signal fluctuation levels $\delta \leq 0.05\%$, solutions were 'correct' ($P_{correct} \approx 100\%$) independently of the uncertainty and type of information provided. This latter result implies the existence of a global minimum threshold ($\delta_{th} = 0.05\%$) to solve the TICA problem, independently of the type of *a priori* information provided from the Plasma Line.

7.2. USE OF *IN SITU* SENSORS

Simulations with known *a priori* information of N_e and T_e parameters (shown in Figure 5.3 and Figure 7.1, without and with uncertainty, respectively) obtained the best unambiguous estimation performance. This result suggests that measurements from *in situ* sensors onboard satellites and rockets, such as Langmuir Probes or Retarding Potential Analyzers (RPA), could also be used as *a priori* known parameters to solve the TICA problem. The sensor type and its characteristics should be considered to determine its uncertainty and calculate the corresponding unambiguous estimation probability in Figure 7.1. Alternatively, measurements from in situ sensors could also be used to compare with the parameters estimated by the radar.

Although satellite measurements at altitudes below 300 km are scarce, present trends in nano-satellite design (i.e. Cubesat standard) are increasing the number of platforms that could be used in a future scenario for a sustained monitoring at these altitudes. These measurements could be obtained at a particular time and orbital altitude, providing information of the ion composition at a single range gate of the radar. Nevertheless, determining how these measurements could be used to unambiguously estimate the entire ion composition profile should be studied in a future work.

7.3. STATIONARITY OF UNAMBIGUOUS RADAR MEASUREMENTS

Simulation results shown in Figure 5.3 suggest that it is possible to unambiguously estimate plasma parameters from ISR signals with very small signal fluctuations without providing *a priori* information. Furthermore, the addition of *a priori* information allows increase the signal fluctuation level required to solve the ambiguity problem. Different combinations of *a priori* known parameters from the Plasma Line provide different signal fluctuation thresholds.

To obtain the signal fluctuation levels required to solve the TICA problem it must be considered the spatial and temporal stationarity of the plasma being measured. The acquisition of signals with small fluctuations (δ) require large *SNR* values or large number of integrations (N) in Equation (3), as shown in Figure 5.5. Nevertheless, assuming noise contributions independent from plasma parameters requires the reception of signals with $SNR \ll 1$, as indicated in Chapter 3.1.2. Noise Addition Scheme. Therefore, to provide small signal fluctuation levels it is mandatory to increase the number of integrations. This effect can be accomplished by increasing the integration time or by simultaneously receiving multiple signals.

Increasing the integration time to unambiguously estimate plasma parameters was previously studied in [Lathuillere et al., 1983] and [Lathuillere & Pibaret, 1992] using integration periods of 5 minutes at EISCAT. However, the integration of signals requires to assume stationary plasma conditions during the integration period [Farley, 1969] [Nikoukar et al., 2012]. Long integration periods could affect this stationary assumption if plasma parameters vary during the measurement period, finally obtaining incorrect plasma estimates. Note that the determination of the maximum integration time is dependent on particular ionospheric events and their temporal and spatial dynamics. Furthermore, the use of *a priori* information assumes that the *a priori* known parameters, obtained from the Plasma Line or from *in situ* sensors, are measured simultaneously to the radar signal measurement and also have spatial and temporal stationary conditions. The use of *a priori* known parameters different than the ionospheric parameters integrated during long periods would affect the unambiguous estimation of plasma parameters by increasing the uncertainty of the *a priori* parameters, as shown in Figure 7.1.

Alternatively, the simultaneous transmissions and reception at different frequencies was proposed by [Sulzer, 1986a]. These frequencies contribute with independent information of the same plasma, improving the final signal fluctuation level obtained. The simultaneous reception of 7 frequencies was implemented at Arecibo [Sulzer, 1986a], and 8 simultaneous frequencies were implemented at EISCAT [Folkestad et al., 1983]. Recently, a multistatic and multibeam radar estimation method has been proposed [Virtanen et al., 2014] using the tristatic facilities of EISCAT VHF and the Kilpisjärvi Atmospheric Imaging Receiver Array (KAIRA). This latter method combines the signals obtained from all receivers and obtains significant accuracy improvements of plasma parameter estimates [Virtanen et al., 2014]. In this case, different plasma parameters projections were assumed at each receiver, finally obtaining ion temperature estimates perpendicular and aligned to the magnetic field lines and tridimensional ion drifts.

It is suggested to implement the simultaneous transmission and reception at multiple frequencies and/or the multistatic reception to reduce the required integration time, helping to avoid the non-stationarity of the plasma during long measurement periods.

CHAPTER 8

CONCLUSIONS

8.1. SUMMARY OF THE THESIS WORK

This thesis has studied the TICA problem conducting Monte Carlo simulations for different plasma parameters of the ISR estimation process at different signal fluctuation values. These simulations were done using the most commonly used Non-Linear Least Squares (NLLS) optimization algorithm: the Levenberg-Marquardt (L-M) algorithm. The convergence of the estimated solution was determined using the statistical distribution of the Reduced Chi-Squared (χ_r^2) cost function. Results were clustered using the Expectation Maximization (EM) algorithm to determine the 'correctness' of the result depending on the distance of the ion composition estimate error to the 'correct' and 'incorrect' probability distribution functions.

The probability of convergence of the optimization algorithm ($P_{fit\ valid}$) was computed as the ratio between the number of convergent solutions and the total number of simulations. The probability of 'correct' estimation ($P_{correct}$) was computed as the ratio between the number of 'correct' solutions and the number of convergent solutions. Finally, the probability of valid convergence and 'correct' estimation ($P_{fit\ valid\ \&\ correct}$) was computed as the product of previous probabilities. Simulations were done at different fluctuation levels (δ) to determine the maximum fluctuation threshold required to avoid the TICA problem. These thresholds (δ_{th}) were selected to have a probability of 'correct' estimation of 2σ ($P_{correct} = 95.45\%$).

Ionospheric models are commonly used in the ISR estimation process to determine the initial parameters of the fitting. To study the effects of an inaccurate initial guess of parameters, simulations with different values of uncertainty on the initial parameters (β) were performed. Results indicate that for very small signal fluctuations ($\delta \leq \delta_{th} = 0.05\%$) almost all parameters were estimated 'correctly' independently of the uncertainty of the initial guess. For noisier cases, the value of $P_{correct}$ was highly dependent on the distance between the initial guess and the true input parameters. Results show that even with an accurate initial guess ($\beta = 1\%$), estimations are ambiguous ($P_{correct} < 95.45\%$) for signal fluctuation levels higher than $\delta_{th(\beta=1\%)} = 0.54\%$. These results might suggest changes in the typical estimation process of ISRs.

Probabilities of valid convergence ($P_{fit\ valid}$) and of 'correct' estimation ($P_{correct}$) were computed for different combinations of plasma parameters known *a priori* from the Plasma Line. Results indicate that the combination of *a priori* information of N_e and T_e solved the ambiguity problem even in highly noisy scenarios ($\delta_{th(Ne\ and\ Te)} = 7.93\%$). To further determine the information provided by each plasma parameter, simulations knowing *a priori* a single parameter were done. Results indicate that plasma temperatures (T_e and T_i) provide the most information. This agrees with previous studies ([Vallinkoski, 1988] and [Vallinkoski &

Lehtinen, 1990]) that indicate that highly correlated error parameters have the greatest impact on solving the ion composition ambiguity.

The effect of increasing the uncertainty of the *a priori* known parameters (ϵ) was also studied. Solutions were not convergent for signal fluctuation levels much smaller than the uncertainty value ($\delta \ll |\epsilon|$). The value of $P_{correct}$ was found dependent on the uncertainty level. For uncertainties $|\epsilon| \leq 1\%$, $P_{correct}$ values were similar to the results of simulations without uncertainty. Furthermore, results suggest that a global minimum threshold ($\delta_{th} = 0.05\%$) solves the TICA problem independently of the amount of uncertainty and type of information provided from the Plasma Line. Given the importance of a priori estimates of N_e and T_e parameters for unambiguously solve the estimation problem, it is suggested that *in situ* sensors onboard satellites or rockets could be used to obtain these a priori estimates at particular ranges. Finally, it is suggested to use multiple transmission frequencies [Sulzer 1986a] and/or a multistatic and multibeam reception [Virtanen et al., 2014] to increase the number of integrations without affecting the stationary plasma assumption.

8.2. IMPROVEMENTS TO THE ISR TECHNIQUE

The results obtained in this study suggest several operational improvements of the ISR estimation process:

1) To verify the solution convergence. It is suggested to use the Reduced Chi-Square (χ_r^2) cost function to normalize the estimation results (as shown in Chapter 3.1.3), determine a maximum cost function value ($\chi_{r,max}^2$) that corresponds to a 4σ probability criterion depending on the Degrees of Freedom (DoF) of the estimation problem (as shown in Section 3.2.1), and then determine the convergence of the estimated solutions by selecting $\chi_r^2 \leq \chi_{r,max}^2$.

2) To obtain results independent of the initial guess accuracy (shown in Chapter 4.3). It is suggested to execute several fittings with different initial parameters uniformly selected from the range of parameters, and then select the most frequently obtained solution. Due to the ambiguity problem the most frequent solution could be the 'incorrect' solution, as shown in the example of Figure 3.7. The probability to select the 'correct' solution can be obtained as the value of $P_{correct}$ shown in Figure 5.3 and Figure 7.1.

3) To increase the probability of unambiguous estimation when adding *a priori* information (shown in Chapter 7). It is suggested to determine the probability of unambiguous estimation as the value of $P_{correct}$ shown in Figure 7.1 depending on the fluctuation level of the radar signal (δ) and the uncertainty of the *a priori* known parameters (ϵ). When the solution convergence is not verified, the probability of unambiguous estimation should be approximated by $P_{fit\ valid\ \&\ correct}$ in Figure 7.1. It is also suggested to ensure the convergence of the solutions by selecting a signal fluctuation $\delta \geq |\epsilon|$.

8.3. FUTURE WORK

Several studies are planned as a direct continuation of the results of this thesis work:

- Study of the ion composition ambiguity at the 2017 Eclipse ISR data of the MIT Haystack Observatory, in collaboration with MIT researchers (L. Goncharenko et al.).

- Study of the effect of using different radar frequencies in the ion composition estimation ambiguity, as suggested by a reviewer of the manuscript.

- A review of current estimation methods at ISR facilities, as suggested by a reviewer of the manuscript.

- Verification of the capability of the Particle Swarm Optimization (PSO) algorithm to solve the TICA problem at different signal fluctuation levels, in collaboration with IDS Research Group of the Universidad de Chile (F. Jaramillo, M. Orchard, and J. Silva).

- Implementation of a Bayesian Estimator to provide optimally information of plasma parameters from the Plasma Line, in collaboration with IDS Group.

GLOSSARY

ACF – Auto-Correlation Function

AMISR – Advanced Modular ISR

AO – Arecibo Observatory (Puerto Rico, USA)

AWGN – Additive White Gaussian Noise

DoF – Degrees of Freedom

EISCAT – European Incoherent SCATter Radar

EISCAT_3D – New EISCAT holographic radar design

ERFI – Imaginary Error Function

ESR – EISCAT Svalbard Radar

f_0F_2 – Plasma Frequency at F_2-layer Maximum

f_r – Radar transmission frequency

f_{PL} – Plasma Line frequency

GN – Gauss-Newton search method

IDS – Information and Decision Systems Group

IRI – International Reference Ionosphere

ISR – Incoherent Scatter Radar

ISS – Incoherent Scatter Spectrum

IPP – Inter-Pulse Period

JRO – Jicamarca Radio Observatory (Jicamarca, Peru)

KAIRA – Kilpisjärvi Atmospheric Imaging Receiver Array

L-M – Levenberg-Marquardt algorithm

LSE – Least Squares Estimator

MAP – Maximum A Posteriori Estimator

MC – Monte Carlo simulation

MLE – Maximum Likelihood Estimator

NLLS – Non-Linear Least Squares

MSIS – Mass Spectrometer Incoherent Scatter Radar

RCS – Radar Cross-Section

RF – Radio Frequency

RMS – Root Mean Square Error

SD – Steepest Descent search method

SDR – Software Defined Radio

SNR – Signal to Noise Ratio

TICA – Temperature-Ion Composition Ambiguity

UHF – Ultra High Frequency

VHF – Very High Frequency

WSS – Wide-Sense Stationary process

BIBLIOGRAPHY

Akbari, H., Bhatt, A., La Hoz, C., and Semeter, J. L. (2017). "Incoherent Scatter Plasma Lines: Observations and Applications". Space Science Reviews, 1-46, doi:10.1007/s11214-017-0355-7

Alker, H.-J. (1978). "Measurement principles in the EISCAT system". EISCAT Technical Note No. 78/5, Kiruna, Sweden. Available at: https://www.eiscat.se/groups/Documentation/BlueBooks/Technical_%20note_1978_5_200dpi_R.pdf/view (Last Access: 26/05/2016)

Andrae, R., Schulze-Hartung, T., and Melchior, P., (2010). "Dos and don'ts of reduced chi-squared". e-print arXiv:1012.3754 [astro-ph.IM]

Aponte, N., Sulzer, M. P., Nicolls, M. J., Nikoukar, R., and González, S. A. (2007). "Molecular ion composition measurements in the F1 region at Arecibo". Journal of Geophysical Research: Space Physics (1978–2012), 112(A6).

Aster, R. C., Borchers, B., and Thurber, C. H. (2012). "Parameter Estimation and Inverse Problems". 2nd Edition, Academic Press, ISBN: 9780123850485

Balsley, B. B., and Gage, K. S. (1980). "The MST radar technique: Potential for middle atmospheric studies". Pure and Applied Geophysics, 118(1), 452-493.

Bauer, P. (1975). "Theory of waves incoherently scattered". Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences, The Royal Society, 280, 167-191

Bauer, P., Waldteufel, P., and Vialle, C. (1974). "The French quadristatic incoherent scatter facility". Radio Science, 9(2), 77-83.

Bevington, P. R., and Robinson, D. K. (2003). "Data Reduction and Error Analysis for the Physical Sciences". 3rd ed., Ed. McGraw-Hill, ISBN: 0-07-247227-8

Beynon, W. J. G., and Williams, P. J. S. (1978). "Incoherent scatter of radio waves from the ionosphere". Reports on Progress in Physics, 41(6), 909.

Bilitza, D., Altadill, D., Truhlik, V., Shubin, V., Galkin, I., Reinisch, B., and Huang, X. (2017). "International Reference Ionosphere 2016: From ionospheric climate to real-time weather predictions". Space Weather, 15, 418–429, doi:10.1002/2016SW001593

Bjørnå, N. (1989). "Derivation of ion-neutral collision frequencies from a Combined Ion Line/Plasma Line Incoherent Scatter Experiment". Journal of Geophysical Research: Space Physics, 94(A4), 3799-3804.

Bjørnå, N., and Kirkwood, S. (1988). "Derivation of ion composition from a combined ion line/plasma line incoherent scatter experiment". Journal of Geophysical Research: Space Physics, 93(A6), 5787-5793.

Blelly, P. L., Alcaydé, D., and van Eyken, A. P. (2010). "A new analysis method for determining polar ionosphere and upper atmosphere characteristics from ESR data: Illustration with IPY period". Journal of Geophysical Research: Space Physics, 115(A9).

Bowles, K. L. (1958). "Observations of vertical incidence scatter from the ionosphere at 41 Mc/sec". Phys. Rev. Lett., vol. 1, pp. 454-455.

Bowles, K. L. (1961). "Incoherent scattering by free electrons as a technique for studying the ionosphere and exosphere: Some observations and theoretical considerations". Journal of Research NBS, vol. 65D, p p . 1-13.

Bowles, K. L., Ochs, G. R., and Green, J. L. (1962). "On the absolute intensity of incoherent scatter echoes from the ionosphere". J. Res. NBS D, 66, 395-407.

Cabrit, B., and Kofman, W. (1996). "Ionospheric composition measurement by EISCAT using a global fit procedure". In Annales Geophysicae (Vol. 14, No. 12, pp. 1496-1505). Springer-Verlag.

Cohen, M. H. (2009). "Genesis of the 1000-foot Arecibo Dish". Journal of Astronomical History and Heritage, 12, 141-152.

Dempster A. P., Laird, N. M., and Rubin, D. B. (1977). "Maximum Likelihood from Incomplete Data via the EM Algorithm". Journal of the Royal Statistical Society. Series B (Methodological), 39, 1, 1-38

Diaz, M. A., Semeter, J. L., Oppenheim, M., and Zettergren, M. (2008). "Particle-in-cell simulation of the incoherent scatter radar spectrum". Radio Science, 43(1).

Ding, Z., Wu, J., Xu, Z., Xu, B., and Dai, L. (2018). "The Qujing incoherent scatter radar: system description and preliminary measurements". Earth, Planets and Space, 87, 70, 1, doi: 10.1186/s40623-018-0859-8

Djuth, F. T., Sulzer, M. P., and Elder, J. H. (1994). "Application of the coded long-pulse technique to plasma line studies of the ionosphere". Geophysical research letters, 21(24), 2725-2728.

Erickson, P. J. (1998). "Observations of light ions in the midlatitude and equatorial topside ionosphere". PhD Dissertation, Cornell University, May 1998

Erickson, P., Gonzalez, S., Heinselman, C., Hysell, D., Kelly, J., McCready, M., Robinson, R., Ruohoniemi, J. M., Tech, V., and Van Eyken, A.P. (2008). "The National Science Foundation's Upper Atmospheric Facilities: Integrating Management, Operations, and Science". Office of Publications and Marketing at Cornell University, 12/08 500 CDS 080306, Available at: http://landau.geo.cornell.edu/papers/UAFreport.pdf (Last Access: 25/06/2016)

Evans, J. V. (1969). "Theory and practice of ionosphere study by Thomson scatter radar". Proceedings of the IEEE, 57(4), 496-530.

Evans, J. V., and Oliver, W. L. (1972). "The study of E-region ion concentration and composition by incoherent scatter radar". Radio Science, 7(1), 103-112.

Farley, D. T. (1969). "Incoherent Scatter Correlation Function Measurements". Radio Science, 4(10), 935-953, DOI:10.1029/RS004i010p00935.

Folkestad, K., Hagfors, T., and Westerlund, S. (1983), ''EISCAT: An updated description of technical characteristics and operational capabilities'', Radio Science, 18(6), 867– 879, doi:10.1029/RS018i006p00867.

Gavin, H. (2017). "The Levenberg-Marquardt method for nonlinear least squares curve-fitting problems". Department of Civil and Environmental Engineering, Duke University. Retrieved from: http://people.duke.edu/~hpgavin/lm.pdf

González, S. A., and Sulzer, M. P. (1996). "Detection of He+ layering in the topside ionosphere over Arecibo during equinox solar minimum conditions". Geophysical research letters, 23(18), 2509-2512.

González, S. A., Sulzer, M. P., Nicolls, M. J., and Kerr, R. B. (2004). "Solar cycle variability of nighttime topside helium ion concentrations over Arecibo". Journal of Geophysical Research: Space Physics, 109(A7).

Gordon, W. E. (1958). "Incoherent scattering of radio waves by free electrons with applications to space exploration by radar". Proc. IRE, vol. 46, no. 11, pp. 1824-1829.

Grydeland, T., Lind, F. D., Erickson, P. J., and Holt, J. M. (2005). "Software radar signal processing". In Annales Geophysicae (Vol. 23, No. 1, pp. 109-121).

Hoffman, J. H., Johnson, C. Y., Holmes, J. C., and Young, J. M. (1969). "Daytime midlatitude ion composition measurements". Journal of Geophysical Research, 74(26), 6281–6290. doi: 10.1029/ja074i026p06281

Holt, J. M., Rhoda, D. A., Tetenbaum, D., and Eyken, A. P. (1992). "Optimal analysis of incoherent scatter radar data". Radio science, 27(3), 435-447.

Holt, J. M., Erickson, P. J., Gorczyca, A. M., and Grydeland, T. (2000). "MIDAS-W: a workstation-based incoherent scatter radar data acquisition system". In Annales Geophysicae (Vol. 18, No. 9, pp. 1231-1241). Springer-Verlag.

Huuskonen, A., and Lehtinen, M. S. (1996). "The accuracy of incoherent scatter measurements: error estimates valid for high signal levels". Journal of Atmospheric and Terrestrial Physics, 58(1), 453-463.

Hysell, D. L., Rodrigues, F. S., Chau, J. L., and Huba, J. D. (2008). "Full profile incoherent scatter analysis at Jicamarca". In Annales geophysicae: atmospheres, hydrospheres and space sciences (Vol. 26, No. 1, p. 59).

Hysell, D. L., Chau, J. L., and Huba, J. D. (2009). "Topside measurements at Jicamarca during solar minimum". In Annales geophysicae: atmospheres, hydrospheres and space sciences (Vol. 27, No. 1, p. 427).

Jain, A. K., Murty, M. N., and Flynn, P. J. (1999). "Data clustering: a review". ACM computing surveys (CSUR), 31(3), 264-323, doi:10.1145/331499.331504

Janches, D., and Nicolls, M. J. (2007). "Diurnal variability of the gyro resonance line observed with the Arecibo incoherent scatter radar at E-and F1-region altitudes". Geophysical research letters, 34(1).

Johnson, S. G. (2012). "Faddeeva Package". MIT License, Massachusetts Institute of Technology. Available at: http://ab-initio.mit.edu/wiki/index.php/Faddeeva_Package (Last Access: 27/06/2016)

Kelly, J. D., and Wickwar V. B. (1981). "Radar measurements of high-latitude ion composition between 140 and 300 km altitude". J. Geophys. Res., 86(A9), 7617–7626, doi: 10.1029/JA086iA09p07617

Kennedy, J., and Eberhart, R. (1995). "Particle swarm optimization". Neural Net-works, Proceedings., IEEE International Conference on , vol.4, no., pp.1942-1948 vol.4,

Kirkpatrick, S., Gelatt, C.D. and Vecchi, M.P. (1983). "Optimization by Simulated Annealing". Science, 220, 671-680.

Kudeki, E. (2010). "Applications of Radiowave Propagation". Course ECE 458 Lecture Notes. Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, January 2006, edited in Jan. 2010. Available at: https://courses.engr.illinois.edu/ece458/458sp10.pdf (Last Access: 20/07/2016)

Kudeki, E., Bhattacharyya, S., and Woodman, R. F. (1999). "A new approach in incoherent scatter F region E× B drift measurements at Jicamarca". Journal of Geophysical Research: Space Physics, 104(A12), 28145-28162.

Kudeki, E. and Milla, M. A. (2011). "Incoherent scatter spectral theories-Part I: A general framework and results for small magnetic aspect angles". IEEE Transactions on Geoscience and Remote Sensing, 49(1), 315-328.

Lathuillere, C., Lejeune, G., and Kofman, W. (1983). "Direct measurements of ion composition with EISCAT in the high-latitude F1 region". Radio science, 18(6), 887-893.

Lathuillere, C., and Pibaret, B. (1992). "A statistical model of ion composition in the auroral lower F region". Advances in Space Research, 12(6), 147-156, doi: 10.1016/0273-1177(92)90048-3

Leadabrand, R. L., Baron, M. J., Petriceks, J., and Bates, H. F. (1972). "Chatanika, Alaska, Auroral-Zone Incoherent-Scatter Facility". Radio Science, 7(7), 747-756.

Lehtinen, M. S. (1986). "Statistical theory of incoherent scatter measurements". EISCAT Tech. Note 86/45. PhD Dissertation, University of Helsinki.

Lehtinen, M. S., and Huuskonen, A. (1996). "General incoherent scatter analysis and GUISDAP". Journal of Atmospheric and Terrestrial Physics, 58(1), 435-452.

Lehtinen, M. S., Huuskonen, A., and Pirttilä, J. (1996). "First experiences of full-profile analysis with GUISDAP". In Annales Geophysicae (Vol. 14, No. 12, pp. 1487-1495). Springer-Verlag.

Lehtinen, M. S., Huuskonen, A., and Markkanen, M. (1997). "Randomization of alternating codes: Improving incoherent scatter measurements by reducing correlations of gated autocorrelation function estimates". Radio Science, 32(6), 2271-2282.

Levenberg, K. (1944). "A Method for the Solution of Certain Non-Linear Problems in Least Squares". Quarterly of Applied Mathematics 2: 164-168.

Litvine, A., Kofman, W., and Cabrit, B. (1998). "Ion composition measurements and modelling at altitudes from 140 to 350 km using EISCAT measurements". In Annales Geophysicae (Vol. 16, No. 10, pp. 1159-1168). Springer-Verlag.

Longley, W. J., Oppenheim, M. M., Fletcher, A. C., and Dimant, Y. S. (2018). "ISR spectra simulations with electron-ion Coulomb collisions". Journal of Geophysical Research: Space Physics, 123, 2990–3004, doi:10.1002/2017JA025015

Lourakis, M. I. (2005). "A brief description of the Levenberg-Marquardt algorithm implemented by levmar". Foundation of Research and Technology, Hellas (FORTH) 4, 1-6.

Lu, Z., Yao, M., and Deng, X. (2016). "An Effective Method for Incoherent Scattering Radar's Detecting Ability Evaluation". Radio Science, 51, DOI:10.1002/2015RS005827.

Madsen, K., Bruun, H., and Tingleff, O. (2004). "Methods for non-linear least squares problems". 2nd Edition, Informatics and Mathematical Modelling, Technical University of Denmark

Marquardt, D. (1963). "An Algorithm for Least-Squares Estimation of Nonlinear Parameters". SIAM Journal on Applied Mathematics 11 (2): 431-441.

Mathews, J. D., Breakall, J. K. and Ganguly, S. (1982). "The measurement of diurnal variations of electron concentration in the 60–100 km ionosphere at Arecibo". Journal of Atmospheric and Terrestrial Physics, 44(5), 441-448.

McCrea, I., Aikio, A., Alfonsi, L., Belova, E., Buchert, S., Clilverd, M., ..., and Kosch, M. (2015). "The science case for the EISCAT_3D radar". Progress in Earth and Planetary Science, 2(1), 1.

McCready, M. A., and Heinselman, C. J. (2013). "The Chatanika and Sondrestrom Radars–a brief history". History of Geo-and Space Sciences, 4(1), 1-6.

McKay, D., and McCrea, I. (2009). "EISCAT_3D. The next generation European Incoherent Scatter radar system. Final design study report". Work Package 11. Deliverable D11.1 Version 1.00, Date: 05-Jun-2009. Available at: https://eiscat3d.se/sites/default/files/e3d_docs/WP11%20deliverables/Deliverable_D11_1.pdf (Last Access: 25/06/2016)

McKay-Bukowski, D., Vierinen, J., Virtanen, I. I., Fallows, R., Postila, M., Ulich, T., ..., and Gerbers, M. (2015). "KAIRA: The Kilpisjärvi Atmospheric Imaging Receiver Array-System Overview and First Results". IEEE Transactions on Geoscience and Remote Sensing, 53(3), 1440-1451.

Milla, M. A., and Kudeki, E. (2011). "Incoherent Scatter Spectral Theories-Part II: Modeling the Spectrum for Modes Propagating Perpendicular to B". IEEE Transactions on Geoscience and Remote Sensing, 49(1), 329-345.

Milla, M.A., Kudeki, E., Reyes, P.M., and Chau, J.L. (2013). "A multi-beam incoherent scatter radar technique for the estimation of ionospheric electron density and Te/Ti profiles at Jicamarca". J. Atmos. Solar Terr. Phys. 105, 214–229

Moon, T. K. (1996). "The expectation-maximization algorithm". IEEE Signal processing magazine, 13(6), 47-60, doi:10.1109/79.543975

Nicolls, M. J., Sulzer, M. P., Aponte, N., Seal, R., Nikoukar, R., and González, S. A. (2006). "High-resolution electron temperature measurements using the plasma line asymmetry". Geophysical research letters, 33(18).

Nikoukar, R. (2010). "Near-Optimal Inversion of Incoherent Scatter Radar Measurements: Coding Schemes, Processing Techniques, and Experiments". PhD Thesis Dissertation, University of Illinois at Urbaba-Champaign. Available at: https://www.ideals.illinois.edu/bitstream/handle/2142/16990/Nikoukar_Romina.pdf?sequence=11 (Last Access: 27/06/2016)

Nikoukar, R., Kamalabadi, F., Kudeki, E., and Sulzer, M. (2008). "An efficient near-optimal approach to incoherent scatter radar parameter estimation". Radio Science, 43(5).

Nikoukar, R., Kamalabadi, F., Kudeki, E., and Sulzer, M. (2012). "On resolution/error trade-offs in incoherent scatter radar measurements". Radio Science, 47(1).

Oliver, W. L. (1979). "Incoherent scatter radar studies of the daytime middle thermosphere". In Annales de Geophysique (Vol. 35, pp. 121-139).

Perkins, F. W., Salpeter, E. E., and Yngvesson, K. O. (1965). "Incoherent scatter from plasma oscillations in the ionosphere". Physical Review Letters, 14(15), 579.

Perrone, L., and Mikhailov, A. V. (2017). "Long-term variations of exospheric temperature inferred from foF1 observations: A comparison to ISR Ti trend estimates". J. Geophys. Res. Space Physics, 122, 8883–8892, doi:10.1002/2017JA024193

Perrone, L., and Mikhailov, A. V. (2018). "Reply to comments by S. Zhang, J.M. Holt, P.J. Erickson, and L.P. Goncharenko on the paper "Long-term variations of exospheric temperature inferred from foF1 observations: A comparison to ISR Ti trend estimates" by L. Perrone and A.V. Mikhailov". Journal of Geophysical Research: Space Physics, 123, doi:10.1029/2017JA025039

Picone, J. M., Hedin, A. E., Drob, D. P., and Aikin, A. C. (2002). "NRLMSISE-00 empirical model of the atmosphere: Statistical comparisons and scientific issues". J. Geophys. Res., 107(A12), 1468, doi:10.1029/2002JA009430

Sato, K., Tsutsumi, M., Sato, T., Nakamura, T., Saito, A., Tomikawa, Y., Nishimura, K., Kohma, M., Yamagishi, H., and Yamanouchi, T. (2014). "Program of the Antarctic Syowa MST/IS radar (PANSY)". Journal of Atmospheric and Solar-Terrestrial Physics, 118A, 2-15, doi: 10.1016/j.jastp.2013.08.022.

Scherliess, L., Fejer, B. G., Holt, J., Goncharenko, L., Amory-Mazaudier, C., and Buonsanto, M. J. (2001). "Radar studies of midlatitude ionospheric plasma drifts". J. Geophys. Res., 106(A2), 1771–1783, doi:10.1029/2000JA000229.

Sharma, D. K., Aggarwal, M., and Bardhan, A. (2016). "Variability of ionospheric parameters during solar minimum and maximum activity and assessment of IRI model". Advances in Space Research, 60, 2, 435-443, doi: 10.1016/j.asr.2016.11.027

Sheffield, J., Froula, D., Glenzer, S. H., and Luhmann Jr, N. C. (2011). "Plasma scattering of electromagnetic radiation: theory and measurement techniques". Academic press. Second Edition, 520 pages ISBN: 978-0-12-374877-5

Sulzer, M. P. (1986a). "A phase modulation technique for a sevenfold statistical improvement in incoherent scatter data-taking". Radio science, 21(4), 737-744.

Sulzer, M. P. (1986b). "A radar technique for high range resolution incoherent scatter autocorrelation function measurements utilizing the full average power of klystron radars". Radio Science, 21(6), 1033-1040.

Swoboda, J., Semeter, J., Zettergren, M., and Erickson, P. J. (2017). "Observability of Ionospheric Space-Time Structure with ISR: A Simulation Study". Radio Sci., 52, 215–234, doi:10.1002/2016RS006182

Taran, V. I. (1988). "Contribution of incoherent scatter facilities to ionospheric informatics". Advances in Space Research, 8(4), 39-48.

Taylor, J. R. (1997). "An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements". 2nd edition, University Science Books, Sausalito, ISBN: 0-935702-42-3

Tsui, P.C., and Boedigheimer, M. (2006). "Fast EM_GM, An expectation maximization algorithm for learning a multi-dimensional Gaussian mixture". Version 1.0. Open source code, Available at: https://www.mathworks.com/matlabcentral/fileexchange/?term=authorid%3A20228

Valentic, T., Buonocore, J., Cousins, M., Heinselman, C., Jorgensen, J., Kelly, J., ..., and Van Eyken, A. (2013). "AMISR the advanced modular incoherent scatter radar". In Phased Array Systems & Technology, 2013 IEEE International Symposium on (pp. 659-663). IEEE.

Vallinkoski, M. (1988). "Statistics of incoherent scatter multiparameter fits". Journal of atmospheric and terrestrial physics, 50(9), 839-851.

Vallinkoski, M., and Lehtinen, M. S. (1990). "The effect of a priori knowledge on parameter estimation errors with applications to incoherent scatter". Journal of Atmospheric and Terrestrial Physics, 52(6), 675-685.

Vandekerckhove, J. (2008). "General simulated annealing algorithm" MATLAB MathWorks® File Exchange, 28 Mar 2006 (Updated 02 Jun 2008), Available at: http://www.mathworks.com/matlabcentral/fileexchange/10548-general-simulated-annealing-algorithm (Last Access: 27/06/2016)

Vierinen, J. (2012). "On statistical theory of radar measurements". PhD Dissertation, Aalto University

Virtanen, I. I., Lehtinen, M. S., Nygrén, T., Orispaa, M., and Vierinen, J. (2008). "Lag profile inversion method for EISCAT data analysis". In Annales geophysicae: atmospheres, hydrospheres and space sciences (Vol. 26, No. 3, p. 571).

Virtanen, I. I., D. McKay-Bukowski, J. Vierinen, A. Aikio, R. Fallows, and L. Roininen (2014), "Plasma parameter estimation from multistatic, multibeam incoherent scatter data", J. Geophys. Res. Space Physics, 119, 10,528–10,543, doi:10.1002/2014JA020540.

Waldteufel, P. (1971). "Combined incoherent scatter F1-region observations". Journal of Geophysical Research, 76(28), 6995-6999.

Walker, J. C. (1979). "Radar measurement of the upper atmosphere". Science, 206(4415), 180-189.

Wand, R. H. (1970). "Electron-to-ion temperature ratio from radar Thomson scatter observations". Journal of Geophysical Research, 75(4), 829-838.

Wannberg, U. G., Andersson, H., Behlke, R., Belyey, V., Bergqvist, P., Borg, J., ..., and Gustavsson, B. (2010). "EISCAT_3D: A Next-Generation European Radar System for Upper-Atmosphere and Geospace Research". The Radio Science Bulletin No 332

Williams, P.J.S., and Taylor, G.N. (1974). "The UK incoherent scatter radar". Radio Science, Volume 9, Number 2, pages 85-88, February 1974

Wu, L. L., Zhou, Q. H., Chen, T. J., Liang, J. J., and Wu, X. (2015). "Application of particle swarm optimization method to incoherent scatter radar measurement of ionosphere parameters". Journal of Geophysical Research: Space Physics, 120(9), 8096-8110.

Yao, M., Deng, X., Lu, Z., Bai, B., and Li, H. (2014a). "Low power software defined incoherent scatter radar system design concept for continuous sounding the earth's ionosphere". IET Radar, Sonar & Navigation, 8(9), 1026-1034.

Yao, M., Zhang, L., Deng, X., Bai, B., Wu, W., Zhou, M., Li, H., and Lu, Z. (2014b). "System Design of the Prototype Incoherent Scatter Radar at Nanchang University". IEEE Geoscience and Remote Sensing Letters, 11(1), 352-356.

Yngvesson, K. O., and Perkins, F.W. (1968). "Radar Thomson scatter studies of photoelectrons in the ionosphere and Landau damping". J. Geophys. Res., 73(1), 97–110.

Zettergren, M., Semeter, J., Heinselman, C., and Diaz, M. (2011). "Incoherent scatter radar estimation of F region ionospheric composition during frictional heating events". Journal of Geophysical Research: Space Physics, 116(A1).

Zhang, S.-R., Holt, J. M., Erickson, P. J., and Goncharenko, L. P. (2018). "Comments on "Long-term variations of exospheric temperature inferred from foF1 observations: A comparison to ISR Ti trend estimates" by Perrone and Mikhailov". Journal of Geophysical Research: Space Physics, 123, 4467–4473, doi:10.1029/2017JA024948

ANNEXES

ANNEX 1. THE INCOHERENT SCATTER SPECTRUM

The signals received by ISRs are generated by the Thomson Scatter effect on ionospheric electrons. An electron located at a distance r from the transmitter would get excited by an incident wave (E_i). Then, an electric field amplitude (E_s) would be backscattered [Kudeki and Milla, 2011]:

$$E_s = -\frac{r_e}{r}\sin(\theta)E_i e^{-jkr} \tag{22}$$

where k ($= 2\pi f_r/c$) is the wavenumber, r_e is the classical electron radius ($r_e = e^2/4\pi\epsilon_0 m_e c^2 \approx 2.818 \times 10^{-15}[m]$), and θ is the polarization angle of the scattered signal that indicates the backscatter reception angle with respect to the transmission.

A "monostatic" radar system uses the same antenna ($\theta = \pi/2$) for the transmission of the signal pulse and the reception of the backscatter signal. In the case of using different antennas to receive the scattering, the radar system can be "bi-static" (two antennas are used) or "multi-static" (more than two antennas are used). In this work we focus on the study of monostatic radar systems because are the most common ISR implementation.

The radar antenna receives the backscattering from a volumetric profile of ionosphere equivalent to the solid angle of the antenna beam. A particular sub-volume ΔV of the beam will generate the backscatter signal from the ionosphere given by [Kudeki and Milla, 2011]:

$$E_s(t) \approx -\frac{r_e}{r}E_i n_e\left(\boldsymbol{k}, t - \frac{r}{c}\right) \tag{23}$$

where the density $n_e(\boldsymbol{k}, t)$ is the 3D spatial Fourier transform of the electron number density $n_e(\boldsymbol{r}, t)$ that relates the electron motions and density wave effects on the backscattering volume ΔV. The propagation time to receive the signal at the radar antenna is considered in the r/c time offset.

As the electron positions and movements of the plasma are random processes, it is required to study these processes using statistical measures. The frequency behavior of a Wide-Sense Stationary (WSS) process can be described using the Power Spectral Density (PSD), or its inverse Fourier transform, the Auto-Correlation Function (ACF), thanks to the Wiener-Khinchin theorem:

$$\langle |E_s(\omega)|^2 \rangle = \int \langle E_s^*(t)E_s(t+\tau)\rangle e^{-j\omega\tau}\, d\tau \tag{24}$$

$$\langle E_s^*(t)E_s(t+\tau)\rangle = \int \frac{1}{2\pi}\,\langle |E_s(\omega)|^2 \rangle e^{+j\omega\tau} d\omega$$

The radar spectrum signal can be computed using the Fourier transform over time lags of the ACF of the received backscattered signals. This spectrum is related to the frequency fluctuations of the ionospheric electron density [Kudeki and Milla, 2011]:

$$\langle |E_s(\omega)|^2 \rangle = \frac{r_e^2}{r^2} |E_i|^2 \langle |n_e(\boldsymbol{k}, \omega)|^2 \rangle \tag{25}$$

The electron density frequency fluctuations $\langle |n_e(\boldsymbol{k}, \omega)|^2 \rangle$ are obtained as the Fourier transform over time lags of the ACF of the random density $n_e(\boldsymbol{k}, t)$:

$$S(\omega) = \langle |n_e(\boldsymbol{k}, \omega)|^2 \rangle = \int \langle n_e^* \left(\boldsymbol{k}, t - \frac{r}{c} \right) n_e \left(\boldsymbol{k}, t - \frac{r}{c} + \tau \right) \rangle \, e^{-j\omega\tau} \, d\tau \tag{26}$$

Many authors have developed the theoretical framework of the Incoherent Scatter Spectra (ISS) using different approaches with identical results [Evans, 1969], verifying the interaction between electrons and ions in the spectrum. For a detailed explanation of the theoretical spectrum refer to [Sheffield et al., 2011].

In this work we use the ISR spectrum formulation of [Kudeki and Milla, 2011]:

$$\langle |n_e(\boldsymbol{k}, \omega)|^2 \rangle = \frac{|j\omega\epsilon_0 + \sigma_i|^2 \langle |n_{te}(\boldsymbol{k}, \omega)|^2 \rangle}{|j\omega\epsilon_o + \sigma_e + \sigma_i|^2} + \frac{|\sigma_e|^2 \langle |n_{ti}(\boldsymbol{k}, \omega)|^2 \rangle}{|j\omega\epsilon_o + \sigma_e + \sigma_i|^2} \tag{27}$$

where the $\langle |n_{te}(\boldsymbol{k}, \omega)|^2 \rangle$ and $\langle |n_{ti}(\boldsymbol{k}, \omega)|^2 \rangle$ are independent thermal density spectra of electrons and ions, respectively. These density parameters are obtained considering a plasma in thermodynamic equilibrium with a Maxwellian velocity distribution and in the absence of collective interactions. Also, σ_e and σ_i are the electron and ion conductivities, respectively.

The thermal electron an ion densities ($\langle |n_{te}(\boldsymbol{k}, \omega)|^2 \rangle$ and $\langle |n_{ti}(\boldsymbol{k}, \omega)|^2 \rangle$, respectively) and the corresponding electron and ion conductivities (σ_e and σ_i) are computed using the following formulation [Kudeki and Milla, 2011]:

$$\langle |n_{ts}(\boldsymbol{k}, \omega)|^2 \rangle = 2N_0 Re\{J_s(\omega_s)\} \tag{28}$$

$$\sigma_s(\boldsymbol{k}, \omega) = j\omega\epsilon_0 \frac{1 - j\omega_s J_s(\omega_s)}{k^2 \lambda_{Ds}^2}$$

where N_0 is the total electron density in the volume of study ΔV, λ_{Ds}^2 is the Debye length, and ω_s is the Doppler-shifted frequency ($\omega_s = \omega - \boldsymbol{k} \cdot \boldsymbol{V_s}$) of the specie generated by the bulk velocity $\boldsymbol{V_s}$ (drift velocity). Finally, the parameter $J_s(\omega)$ denotes the Gordeyev integral of species in a plasma in thermal equilibrium calculated as [Kudeki and Milla, 2011]:

$$J_s(\omega) = \int_0^{+\infty} \langle e^{j\boldsymbol{k} \cdot \Delta \mathbf{r_s}} \rangle e^{-j\omega\tau} d\tau \tag{29}$$

This integral considers all the possible particle displacements $\Delta\mathbf{r}_s$ without external collective effects (i.e. no macroscopic fields). The expectation $\langle e^{j\boldsymbol{k}\cdot\Delta\mathbf{r}_s}\rangle$, called "single-particle ACF", represent the characteristic function of the particle displacement in a thermalized plasma. Considering the simplest case, in which the plasma has no collisions and it is no magnetized, the single-particle ACF can be obtained as [Kudeki and Milla, 2011]:

$$\langle e^{j\boldsymbol{k}\cdot\Delta\mathbf{r}_s}\rangle = e^{-\frac{1}{2}k^2C_s^2\tau^2} \tag{30}$$

where $C_s = \sqrt{k_B T_s/m_s}$ is the thermal speed of species.

The Gordeyev integral can be calculated using the identity [Kudeki and Milla, 2011]:

$$jZ(\theta) = \int_0^{+\infty} e^{-\frac{t^2}{4}} e^{-j\theta t} dt = \sqrt{\pi}\left(e^{-\theta^2}\right) - 2j\left(e^{-\theta^2}\right)\int_0^{\theta} e^{t^2} dt \tag{31}$$

Assuming no collisions and no magnetic field, we can consider $\theta = \omega_s/\sqrt{2}kC_s$ and $t = \sqrt{2}kC_s\tau$. The Dawson integral can be used to compute $\left(e^{-\theta^2}\right)\int_0^{\theta} e^{t^2} d\tau$ [Kudeki and Milla, 2011].

The formulation of the spectral density $\langle|n_e(\boldsymbol{k},\omega)|^2\rangle$ can be divided in two separated bands:

a) The first term is the *Langmuir mode*, also known as *Electron-Line* or *Plasma-Line* [Akbari et al., 2017]. It is a high frequency band (in the MHz range) centered on the electron plasma frequency ($f_p = (2\pi)^{-1}\sqrt{N_e e^2/m_e\epsilon_0}[Hz] \approx 8.97\sqrt{N_e(cm^{-3})}[kHz]$).

b) The second term is the *Ion-Acoustic mode* or *Ion-Line*. It is a low frequency band (tens of kHz) that contains most of the signal power received. Its shape depends upon electron and ion temperature ratio, electron densities, ion drift velocity, and ion composition.

Figure Annex 1 (Left) shows separately the contribution of these two bands (i.e. Ion Acoustic and Plasma Line) in logarithmic scale. In this figure, the Ion Line contribution (in red) is responsible of most of the power received. Alternatively, the Plasma Line (in blue) shows the characteristic frequency peak in the MHz range.

The ISR spectra at the Ion Acoustic frequency band is shown in Figure Annex 1 (Right). This figure shows the typical bi-Maxwellian shape of the Ion Acoustic band, formed by two Gaussians joined together smoothly at the center. This characteristic "double-humped shape" spectrum is created by the Landau damping effect on the plasma. In thermal equilibrium ($T_e = T_i$), the peak frequencies of the spectrum are related to the Ion-Acoustic velocities $v_{IA} = \sqrt{k_B 2T_i/m_i}$ Doppler shifted by the radar frequency ($f_{IA} = kv_{IA} = 2f_r v_{IA}/c$).

In Figure Annex 1 the central frequency of the spectrum is located at the radar transmission frequency (f_r), but for a visual simplification it has been set to zero. This representation resembles the frequency displacement done in the down-conversion mixer of a radar system to process radar signals in base-band.

Figure Annex 2 shows the normalized real part of the ACF calculated as the inverse Fourier transform of the Ion Line spectrum of Figure Annex 1 (Right). Radars measure amplitude and phase information of backscatter signals, being the data acquisition system an analog quadrature detection scheme, and therefore the obtained signals are complex [Alker, 1978] [Holt et al., 2000]. Consequently, the resulting ACF is an imaginary (complex) measure [Farley, 1969]. The imaginary part of the ACF is often very small comparing to the real part, but information of ion velocities is embedded into the phase of the ACF complex signal [Erickson, 1998].

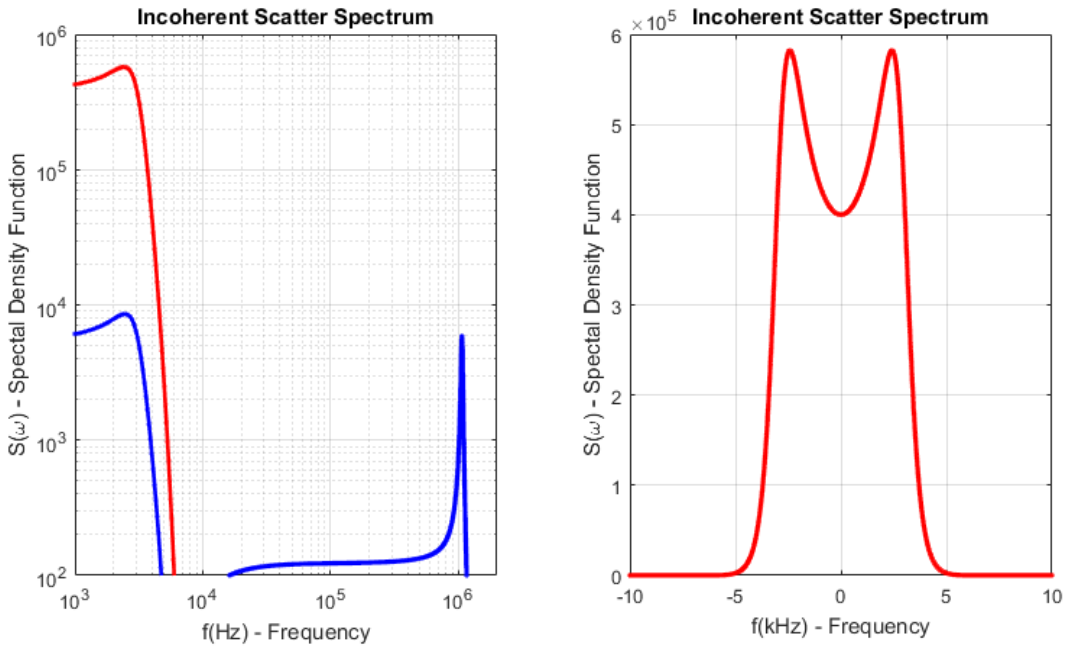

Figure Annex 1. Incoherent Scatter density spectrum showing (Left) the Ion Line in red and Plasma Line in blue frequency bands, and (Right) the total spectrum at the Ion-Acoustic frequency band. The spectrum calculus has been done with $N_e = 1e10[m^{-3}]$, $T_e = 500[^{o}K]$, $T_i = 300[^{o}K]$, *and radar frequency* $f_r = 500\ [MHz]$

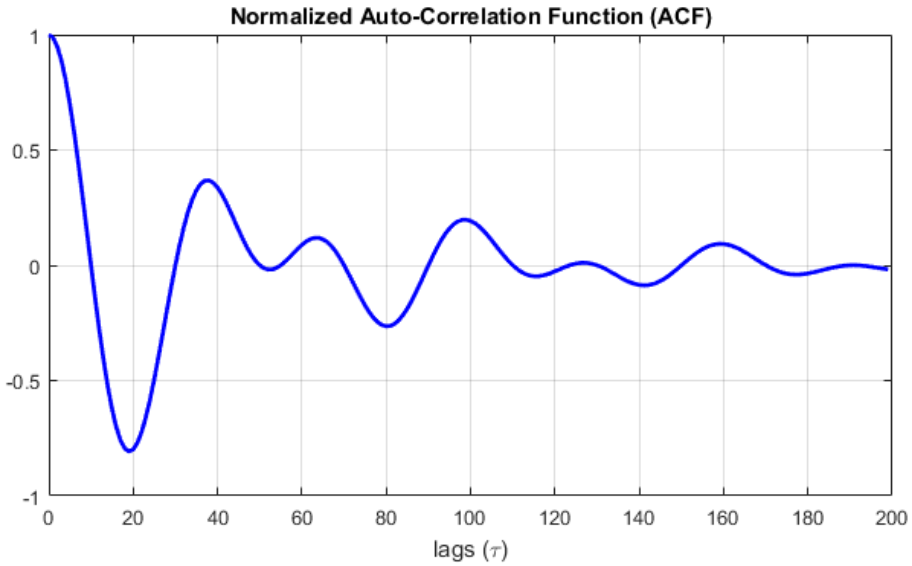

Figure Annex 2. Real part of the Auto Correlation Function (ACF) of the spectrum shown in Figure Annex.1.

The ISR spectrum is a non-linear function of the plasma parameters. Several spectrum simulations have been done to show the effects of varying the input parameters. Figure Annex 3 shows the effect of changing (Left) the electron to ion temperature ratio (T_e/T_i) and (Right) the value of the electron temperature (with $T_e = T_i$). Figure Annex 4 shows the effects of varying (Left) the electron density (N_e) and (Right) the composition percentage of molecular ions ($p = n(M^+)/N_e$). Finally, Figure Annex 5 shows the effect of (Left) using a different transmission frequency (f_r) and (Right) varying the horizontal speed of ions (V_i). All these figures have been done without considering magnetic field effect, assuming the antenna beam not perpendicular to the field lines. Furthermore, these simulations have been done without coulomb collisions between species. Collision effects are considered negligible in the upper ionosphere above 120 km. Simulation have been considered for a plasma composition with atomic oxygen ions (O^+) and otherwise noted.

The impact of varying plasma parameters in the ISR spectrum seen in Figure Annex 3, Figure Annex 4, and Figure Annex 5 could be summarized as: 1) The increase in temperature ratio (T_e/T_i) sinks the spectrum center valley. 2) The increase of electron temperature (T_e) spreads out the spectrum to higher frequencies and lowers the peak power. 3) The increase of electron density (N_e) in an ionospheric volume improves exponentially the power received. 4) The change of ion composition (p) modifies both height and width of the spectrum. 5) The use of a higher radar frequency (f_r) spreads the spectrum over higher frequencies and decreases the backscatter power peak received. And 6) the increase of the horizontal ion drift velocity (V_i) changes the center frequency of the spectrum, indicating a Doppler effect due to the collective plasma drift movement.

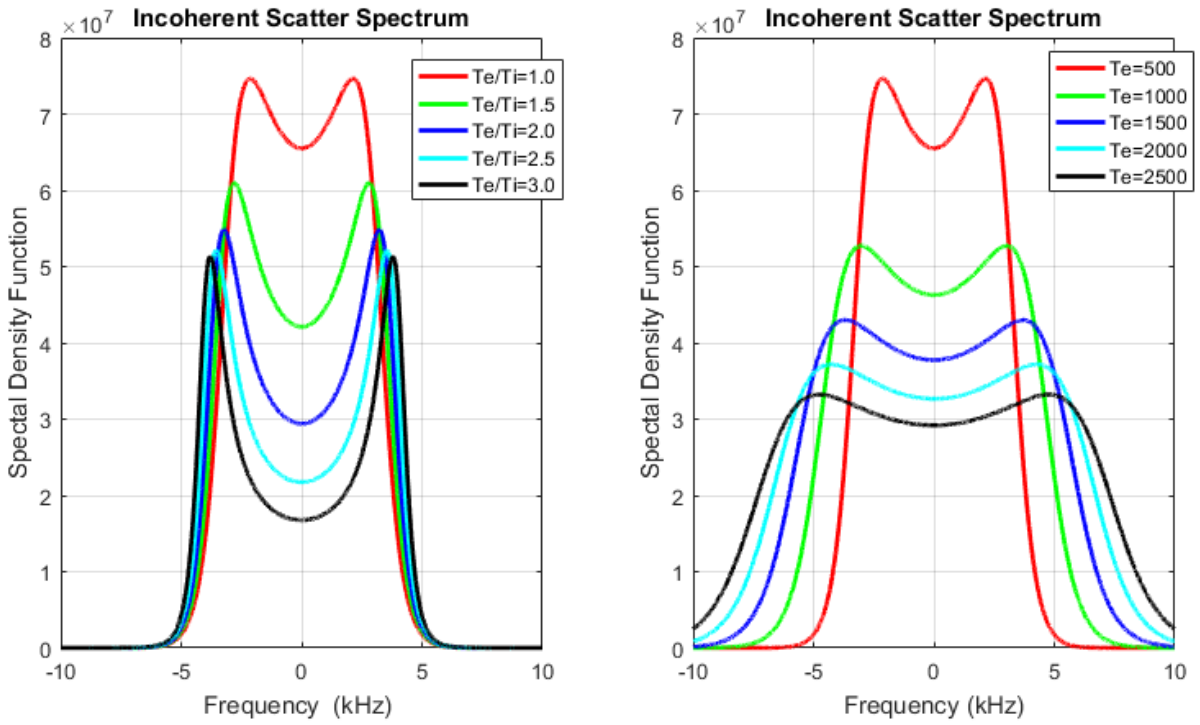

Figure Annex 3. Incoherent Scatter theoretical spectrum varying (Left) the electron to ion temperature ratio (T_e/T_i) and (Right) the electron temperature (T_e) for $T_e = T_i$. Simulations have been done considering a radar frequency $f_r = 450$ $[MHz]$ and N_e=$1e12$ $[m^{-3}]$.

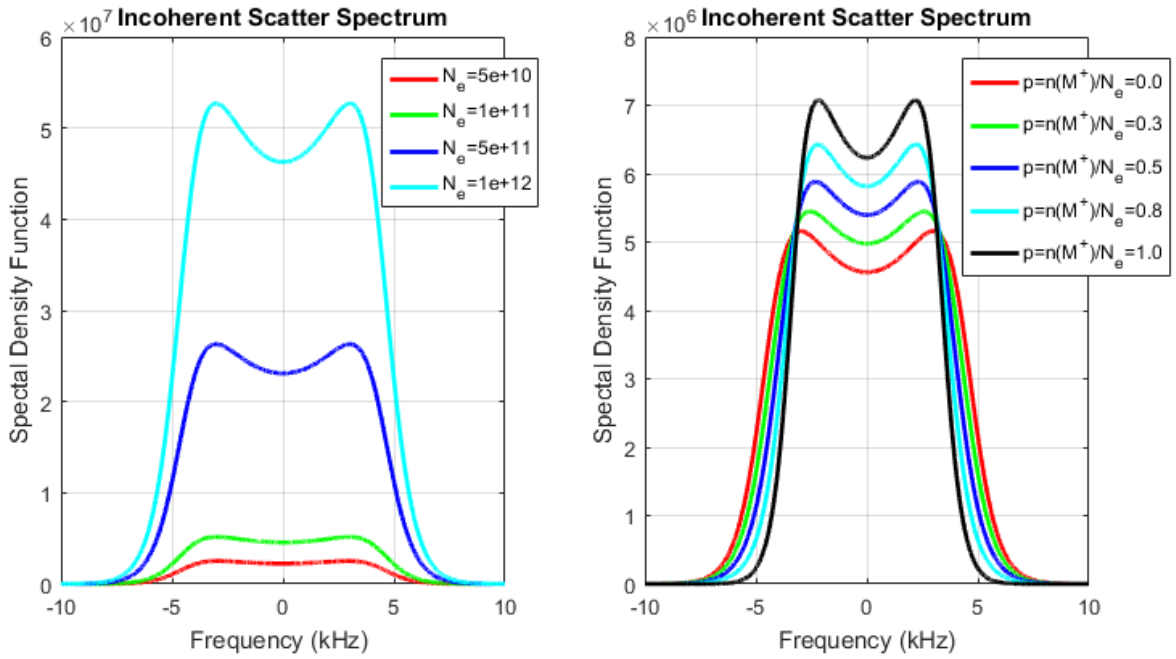

Figure Annex 4. Incoherent Scatter theoretical spectrum varying (Left) the electron density (N_e) and (Right) the molecular composition percentage ($p = n(M^+)/N_e$). Simulations have been done considering a radar frequency $f_r = 450\ [MHz]$ and $T_e = T_i = 1000\ [ºK]$.

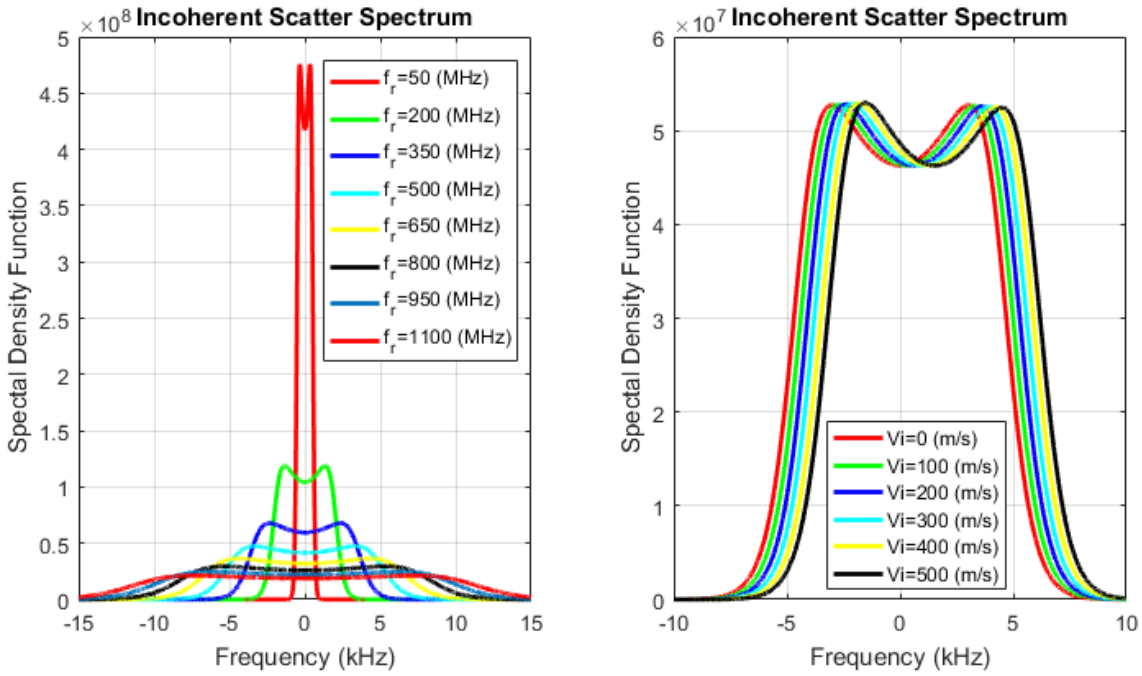

Figure Annex 5. Incoherent Scatter theoretical spectrum varying (Left) the radar transmission frequency (f_0) and (Right) the horizontal velocity of ions (V_i). Simulations have been done with $f_r = 450\ [MHz]$, N_e=1e12 $[m^{-3}]$ and $T_e = T_i = 1000\ [ºK]$.

The ambient magnetic field ($\boldsymbol{B}$) also affects the shape of the spectrum depending on the aspect angle (α) that forms the backscatter wave-vector (k) with magnetic field lines. The plasma particles that interact with wave components parallel ($k_{\parallel}$) and perpendicular ($k_{\perp}$) to the magnetic field suffer different displacements. In particular, the parallel contribution is not being affected by the magnetic field, but the perpendicular contribution is modified depending on the particle gyrofrequency ($\Omega = qB/m$) [Kudeki and Milla, 2011] [Milla and Kudeki, 2011]. This effect is increased almost exponentially when the radar aspect angle is pointed perpendicular to the magnetic lines. This effect can be visualized in the Incoherent Scatter spectra showed in Figure Annex 6, obtained from [Kudeki et al., 1999], that have been computed (Left) for $\alpha = 30º, 60º, 90º$ and (Rigth) for $\alpha = 0.005º, 0.01º, 0.015º, 0.02º$. In this figure, the spectrum has almost no variation for aspect angles higher than 20º, but for very small angles the spectrum gets stretched and tolled almost exponentially.

The Jicamarca Radio Observatory (JRO) located at Jicamarca (Peru) was constructed near the Earth magnetic equator, obtaining aspect angles almost equal to zero. This ISR obtains an extremely narrowed spectrum, shown in Figure Annex 7 from [Kudeki et al., 1999]. This characteristic spectrum allows a very precise Doppler measurements of ion drift velocities.

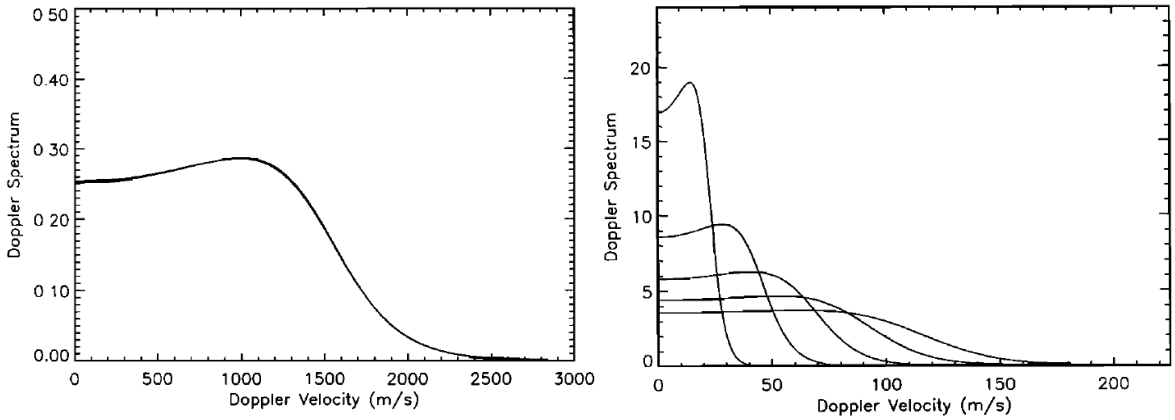

Figure Annex 6. Incoherent Scatter Doppler spectra for (Left) α=30º,60º,90º and (Rigth) α=0.005º,0.01º,0.015º,0.02º. The curves are obtained for an O+ plasma with $Te = Ti = 1000$ [ºK] and a radar frequency $f_r = 50$ MHz, from [Kudeki et al., 1999].

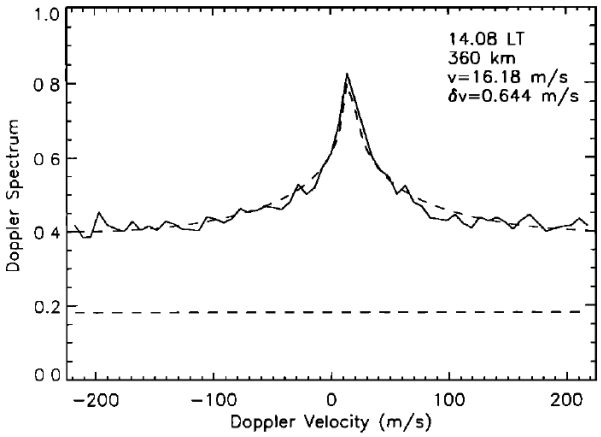

Figure Annex 7. Incoherent Scatter Doppler spectra measured at Jicamarca on September 29, 1996. The spectrum represents 5 minutes of time integration at 15 km height resolution. The dashed curve represents nonlinear least squares fit, and the bottom dashed line indicates the system noise level, from [Kudeki et al., 1999].

ANNEX 2. STANDARD ISR ESTIMATION TECHNIQUE

In the standard estimation process of the Range-Gate analysis method of ISRs, the received signals ($\boldsymbol{m}$) are assumed to be describable as a function ($f(\boldsymbol{x})$) that depends on the plasma parameter vector ($\boldsymbol{x}$) and an additional noise contribution ($\boldsymbol{\varepsilon}$) [Vallinkoski, 1988]:

$$\boldsymbol{m} = f(\boldsymbol{x}) + \boldsymbol{\varepsilon} \tag{32}$$

where $f(\boldsymbol{x})$ function represents the backscatter signal spectra (indicated in Annex 1. The Incoherent Scatter Spectrum), or its inverse Fourier transform (i.e. the ACF, Equation (2)). Both the theoretical signal ($f(\boldsymbol{x})$) and the noise contribution ($\boldsymbol{\varepsilon}$) are vectors of data points with values distributed at the different frequencies of the spectra or at different time lags of the ACF.

The integration of many pulses of the signal provides Gaussian characteristics according to the Central Limit Theorem [Vallinkoski, 1988]. Therefore, the noise ($\boldsymbol{\varepsilon}$) was assumed to be an Additive Gaussian White Noise (AWGN) with a variance-covariance matrix $\mathbf{C}_m = \langle \boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\boldsymbol{T}} \rangle$ [Lehtinen & Huuskonen, 1996]. The diagonal elements of the covariance matrix ($\boldsymbol{\Sigma}_m$) are the measurement variances σ_i^2 and the non-diagonal elements are $\sigma_{ij} = \sigma_i \sigma_j \rho_{ij}$, where ρ_{ij} is the correlation coefficient of measurement errors.

The conditional probability densities ($D(\boldsymbol{m}|\boldsymbol{x})$) of measurements $\boldsymbol{m}$ supposing known parameters $\boldsymbol{x}$ are then given by [Vallinkoski, 1988] [Nikoukar, 2010]:

$$D(\boldsymbol{m}|\boldsymbol{x}) = \frac{1}{(2\pi)^{M/2}|\mathbf{C}_m|^{1/2}} \exp\left(-\frac{1}{2}\big(\boldsymbol{m} - f(\boldsymbol{x})\big)^T \boldsymbol{C}_m^{-1}\big(\boldsymbol{m} - f(\boldsymbol{x})\big)\right) \tag{33}$$

where M is the number of elements of the measurement vector $\boldsymbol{m}$, and $|\boldsymbol{C}_m|$ is the determinant of the measurement error covariance matrix $\boldsymbol{C}_m$.

The "*a posteriori*" density ($D_{post}(\boldsymbol{x})$) as function of parameters $\boldsymbol{x}$ can be calculated from the Bayes theorem as [Vallinkoski, 1988]:

$$D_{post}(\boldsymbol{x}) = D(\boldsymbol{x}|\boldsymbol{m}) = K\, D_{prior}(\boldsymbol{x})\, D(\boldsymbol{m}|\boldsymbol{x}) \tag{34}$$

where K is a constant independent of $\boldsymbol{x}$ ($K = 1/\int D(\boldsymbol{x}, \boldsymbol{m}) d\boldsymbol{x}$), and the "*a priori*" distribution is defined as $D_{prior}(\boldsymbol{x}) = \int D(\boldsymbol{x}, \boldsymbol{m}) d\boldsymbol{m}$.

Therefore, the un-normalized "*a posteriori*" distribution for a parameter vector $\boldsymbol{x}$ given measured values $\bar{\boldsymbol{m}}$ is [Vallinkoski, 1988]:

$$D_{post}(\boldsymbol{x}) = D(\boldsymbol{x}|\bar{\boldsymbol{m}}) = K'\, D_{prior}(\boldsymbol{x})\ \exp\left(-\frac{1}{2}(f(\boldsymbol{x}) - \bar{\boldsymbol{m}})^T \boldsymbol{C}_m^{-1}(f(\boldsymbol{x}) - \bar{\boldsymbol{m}})\right) \tag{35}$$

If the "*a priori*" distribution ($D_{prior}(\boldsymbol{x})$) is considered uniform (i.e. no information provided), the "*a posteriori*" distribution ($D_{post}(\boldsymbol{x})$) will be maximized when the exponent of the exponential function gets maximized. Then, the set of parameters $\boldsymbol{x}$ with maximum probability corresponds to minimize the function $S(\boldsymbol{x})$ [Vallinkoski, 1988] [Nikoukar, 2010]:

$$S(\boldsymbol{x}) = (f(\boldsymbol{x}) - \bar{\boldsymbol{m}})^T \mathbf{C}_m^{-1} (f(\boldsymbol{x}) - \bar{\boldsymbol{m}}) \tag{36}$$

The Maximum Likelihood Estimator (MLE) corresponds to the parameter ($\boldsymbol{x}_{ML}$) that minimizes $S(\boldsymbol{x})$ [Vallinkoski, 1988], and it is equivalent to the Least Squares Estimator (LSE) when considering Gaussian noise additions:

$$\boldsymbol{x}_{ML} = argmin_{\boldsymbol{x}}\{S(\boldsymbol{x})\} \tag{37}$$

For strong radar SNR, the stochastic nature of the plasma dominates the variability of the signal [Huuskonen & Lehtinen, 1996]. Therefore, signals obtained with a strong backscatter from the Ionosphere, due to high electron densities or obtained using high power transmissions, have a signal-dependent noise ($\boldsymbol{\varepsilon}$) [Nikoukar et al., 2008]. Alternatively, for weak SNR conditions, the thermal and sky noises are responsible for the signal variability [Sulzer, 1986a].

In weak SNR conditions the noise ($\boldsymbol{\varepsilon}$) is independent from the signal ($f(\boldsymbol{x})$). Therefore, the matrix $\boldsymbol{C_m}$ is diagonal and only populated by measurement variances (σ_i^2) [Vallinkoski, 1988] [Erickson, 1998]. Previous studies of the TICA problem [Aponte et al., 2007] [Wu et al., 2015] considered those weak SNR and no range smearing effects (i.e. Range-Lag ambiguity [Lehtinen & Huuskonen, 1996]). Those assumption can be accomplished in Multi-Pulse or Alternating Code experiments [Vallinkoski, 1988].

Considering no correlation between measurement errors (i.e. assuming $\boldsymbol{C_m}$ diagonal, with $\sigma_{ij} = 0 \; \forall i \neq j$), the minimization of $S(\boldsymbol{x})$ can be reduced to a minimization of the squared sum of differences weighted by the variances of the measurement errors (i.e. the Chi-Squared cost function) [Erickson, 1998] [Nikoukar, 2010]:

$$\chi^2 = \sum_{i=1}^{M} \frac{\left(\bar{m}_i - f_i(\boldsymbol{x})\right)^2}{\sigma_i^2} \tag{38}$$

This minimization process is identified as an Unconstrained Weighted Least-Squares fitting process [Erickson, 1998], and it is typically computed using a Non-Linear Least Squares (NLLS) optimization algorithm. The Levenberg-Marquardt (L-M) algorithm is the most commonly used NLLS optimization algorithm of ISR analyses, both for Range-Gate [Erickson, 1998] [Swoboda et al., 2017] and for Full Profile [Hysell et al., 2008] [Nikoukar et al., 2008] analyses. The L-M algorithm is explained in more detail in Annex 3. The Levenberg-Marquardt Optimization Algorithm.

ANNEX 3. THE LEVENBERG-MARQUARDT OPTIMIZATION ALGORITHM

Due to the non-linear characteristics of the ISR model (see Annex 1. The Incoherent Scatter Spectrum), a non-linear optimization algorithm is required to estimate plasma parameters from the ISR signal by the minimization of a cost function (i.e. Equation (38) or Equation (13)). This minimization process is a Non-Linear Least Squares (NLLS) optimization problem.

Many optimization algorithms approximate the results of a non-linear function using the first order Taylor expansion. Such algorithms use the Jacobian matrix of the cost function to determine the subsequent step of the iterative descent search. This approximation provides linear or quadratic convergences to the optimum solution [Madsen et al., 2004]. One of the most commonly used optimization methods that uses this criterion is the Levenberg-Marquardt (L-M) optimization algorithm [Levenberg, 1944] [Marquardt, 1963].

The L-M algorithm has become a standard optimization method for NLLS unconstrained problems thanks to its very fast computation performance. This algorithm is a line search in the descent direction that combines the Steepest Descent (SD) method (or gradient method) and the Gauss-Newton (GN) method to accelerate the solution to the optimum minimum [Gavin, 2017] [Madsen et al., 2004] [Lourakis, 2005]. The SD method is a general minimization algorithm which considers the minimum search in the direction opposite to the gradient of the function. Alternatively, the GN method assumes a second order Taylor approximation of the function near the optimum solution. The L-M method is a damped method that adaptively varies the parameter update step between SD and GN methods [Gavin, 2017]. Initially, it is assumed that the optimum solution is far from the initial parameters selected. In this case, the parameter update should be large, and the SD method is then used. When the parameters are near the local minimum, the fine step is computed using the GN method, reaching the minimum in a smaller number of steps. The pseudo-code of the L-M algorithm is shown in Figure Annex 8.

In this work, the implementation of the L-M algorithm to minimize the cost function (χ_r^2) was based on the library created by [Gavin, 2017]. The search range of the parameters ($\boldsymbol{x}_{search\,min} \leq \hat{\boldsymbol{x}} \leq \boldsymbol{x}_{search\,max}$) was configured to be equal to the input parameter ranges of the Monte Carlo simulation (explained in Chapter 3.1.6). To ensure that the two possible ambiguous solutions of temperature were obtained into the search range, electron and ion temperatures search ranges were configured broader than the input parameter ranges, with values ranging from 200 to 6000 ºK. Resolution problems arise in the calculus of the Jacobian matrix due to differences of orders of magnitude of plasma parameters. To avoid this problem, electron density parameter (N_e) values were converted to a logarithmic scale ($log_{10}(N_e)$), as in [Cabrit & Kofman, 1996].

To determine the number of iterations required by the L-M optimization algorithm to obtain the maximum number of convergent and 'correct' solutions, several Monte Carlo simulations were done with different configurations of the optimization algorithm. Simulations were done fitting five plasma parameters (N_e, T_e, T_i, V_i, and p) for different fluctuation levels to ensure that the optimization algorithm would be able to solve the most difficult estimation case (i.e. without considering *a priori* information). Therefore, it is assumed that this configuration would be able to estimate optimally when assuming *a priori* information.

Input: A vector function $f : \mathcal{R}^m \rightarrow \mathcal{R}^n$ with $n \geq m$, a measurement vector $\mathbf{x} \in \mathcal{R}^n$ and an initial parameters estimate $\mathbf{p}_0 \in \mathcal{R}^m$.
Output: A vector $\mathbf{p}^+ \in \mathcal{R}^m$ minimizing $||\mathbf{x} - f(\mathbf{p})||^2$.
Algorithm:
$k := 0$; $\nu := 2$; $\mathbf{p} := \mathbf{p}_0$;
$\mathbf{A} := \mathbf{J}^T\mathbf{J}$; $\epsilon_\mathbf{p} := \mathbf{x} - f(\mathbf{p})$; $\mathbf{g} := \mathbf{J}^T\epsilon_\mathbf{p}$;
stop:=($||\mathbf{g}||_\infty \leq \varepsilon_1$); $\mu := \tau * \max_{i=1,...,m}(A_{ii})$;
while (not stop) and ($k < k_{max}$)
  $k := k + 1$;
  repeat
    Solve $(\mathbf{A} + \mu\mathbf{I})\delta_\mathbf{p} = \mathbf{g}$;
    if ($||\delta_\mathbf{p}|| \leq \varepsilon_2||\mathbf{p}||$)
      stop:=true;
    else
      $\mathbf{p}_{new} := \mathbf{p} + \delta_\mathbf{p}$;
      $\rho := (||\epsilon_\mathbf{p}||^2 - ||\mathbf{x} - f(\mathbf{p}_{new})||^2)/(\delta_\mathbf{p}^T(\mu\delta_\mathbf{p} + \mathbf{g}))$;
      if $\rho > 0$
        $\mathbf{p} = \mathbf{p}_{new}$;
        $\mathbf{A} := \mathbf{J}^T\mathbf{J}$; $\epsilon_\mathbf{p} := \mathbf{x} - f(\mathbf{p})$; $\mathbf{g} := \mathbf{J}^T\epsilon_\mathbf{p}$;
        stop:=($||\mathbf{g}||_\infty \leq \varepsilon_1$) or ($||\epsilon_\mathbf{p}||^2 \leq \varepsilon_3$);
        $\mu := \mu * \max(\frac{1}{3}, 1 - (2\rho - 1)^3)$; $\nu := 2$;
      else
        $\mu := \mu * \nu$; $\nu := 2 * \nu$;
      endif
    endif
  until ($\rho > 0$) or (stop)
endwhile
$\mathbf{p}^+ := \mathbf{p}$;

Figure Annex 8. Pseudo-code of Levenberg-Marquardt algorithm, from [Lourakis, 2005]

Simulations to determine the optimum value of $MaxIter$ were configured with a zero tolerance of the cost function ($\chi^2_{r,tol}$=0), a tolerance of the gradient of parameters $GradTol = 1e - 12$, and a tolerance for parameters values of $ParTol = 1e - 16$. These tolerances are restrictive enough to ensure finding a minimum of the fitting function. A zero tolerance of the cost function ($\chi^2_{r,tol} = 0$) is required to ensure that the fitting achieves a minimum of the cost function. Simulations with a less restrictive cost function tolerance ($\chi^2_{r,tol}$>0) obtained estimated parameters biased. In particular, simulations done with a tolerance equal to the convergence limit ($\chi^2_{r,tol} = \chi^2_{r,max}$) obtained the electron density parameter highly biased at high signal fluctuations.

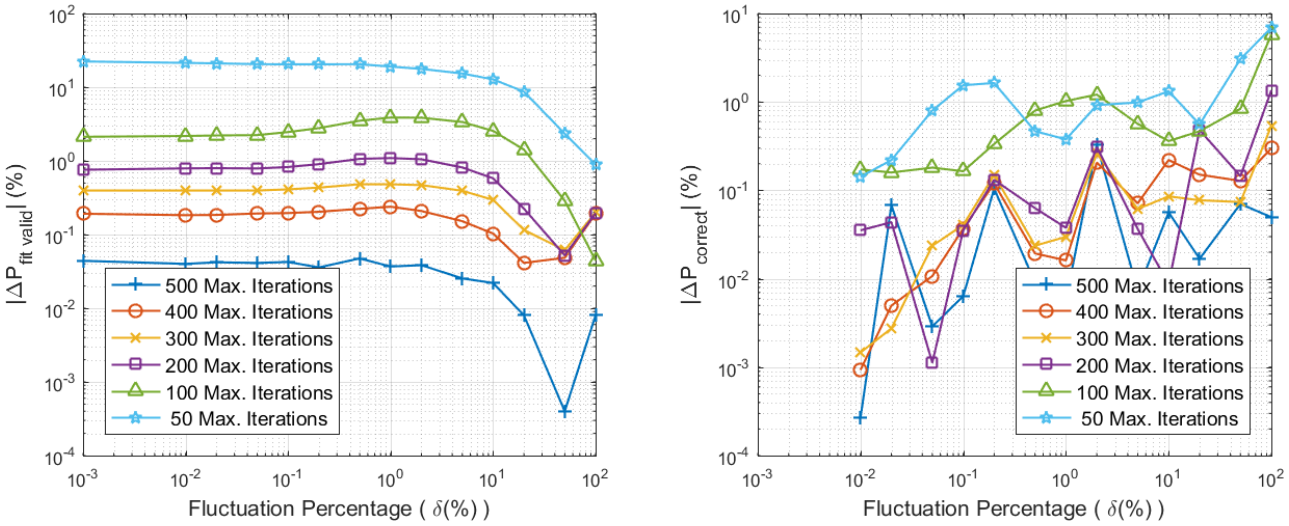

Figure Annex 9. Probability differences of valid ($|\Delta P_{fit\ valid}|$) and 'correct' ($|\Delta P_{correct}|$) estimation for different configurations of maximum number iterations of the L-M algorithm ($MaxIter$). These differences were calculated against results obtained with $MaxIter$=1000.

Probabilities of convergence ($P_{fit\ valid}$) and of 'correct' estimation ($P_{correct}$) were calculated with different values of $MaxIter$. Simulations were done with $MaxIter$ values of 50, 100, 200, 300, 400, 500 and 1000. The absolute values of the differences of those probabilities ($|\Delta P_{fit\ valid}|$ and $|\Delta P_{correct}|$) were obtained to determine the optimum value of maximum iterations ($MaxIter$). The probability differences were calculated comparing each simulation result against the probability results with the highest configuration value (i.e. $MaxIter = 1000$). Figure Annex 9 shows the probability differences in a logarithmical scale. In this figure, $|\Delta P_{fit\ valid}|$ was found highly dependent of the $MaxIter$ configuration value, but $|\Delta P_{correct}|$ obtained very small differences for $MaxIter \geq 200$. The configuration value $MaxIter = 300$ was selected because it obtained probability differences smaller than 1% in all cases of signal fluctuations with a reduced number of iterations. A small number of iterations is required to reduce the computation time required for estimation. A probability difference value of 1% was considered enough to estimate the probabilities with high precision because the random noise added at each estimation generates an estimate variability, and therefore it is not possible to obtain identical probabilities.

Further tests have been also done to determine the optimum tolerances of the L-M algorithm when using $MaxIter$ =300. Results of simulations for different values of Tolerance of Parameters ($ParTol$) are shown in Figure Annex 10. Simulations were done with $ParTol$ values of 1e-16, 1e-12, 1e-9, 1e-6, and 1e-3. This figure (Top) shows the average number of iterations required to obtain convergent and 'correct' results at different fluctuation levels and (Bottom) the absolute values of the differences of the probabilities of those results ($|\Delta P_{fit\ valid}|$ and $|\Delta P_{correct}|$). The probability differences were calculated comparing each simulation result against the probability results with the most restrictive configuration value ($ParTol = 1e-16$). A configuration of $ParTol = 1e-9$ was selected because the number of iterations obtained was smaller than 300 iterations and obtained almost identical probabilities than simulations done with more restrictive tolerances.

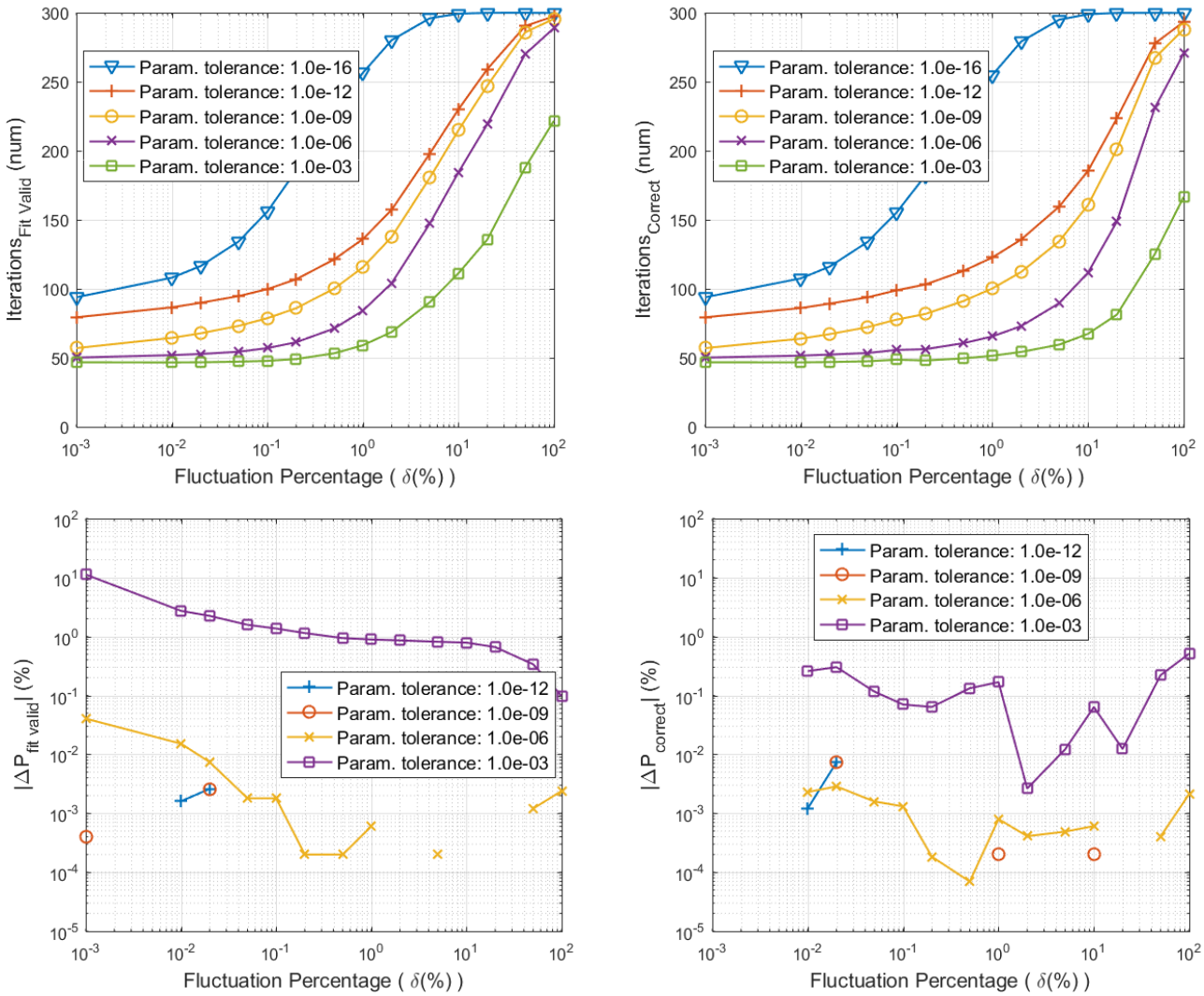

Figure Annex 10. (Top) Average number of iterations of valid and 'correct' results and (Bottom) probability differences of valid ($|\Delta P_{fit\ valid}|$) and 'correct' ($|\Delta P_{correct}|$) estimation obtained with different tolerances of parameters ($ParTol$).

Results of simulations for different values of Tolerance of the Gradient of Parameters ($GradTol$) are shown at Figure Annex 11. Simulations were done with $GradTol$ values of 1e-16, 1e-12, 1e-9, 1e-6, and 1e-3. This figure is similar to Figure Annex 10 but for changes of a different configuration parameter. The probability differences were calculated comparing each simulation result against the probability results with the most restrictive configuration value ($GradTol = 1e-16$). In this case, a value of $GradTol = 1e-9$ was selected because simulations with $GradTol = 1e-12$ and $1e-16$ required very similar number of iterations. This indicates that almost no improvement was obtained with more restrictive tolerances.

Therefore, the configuration values selected for the L-M optimization algorithm were $MaxIter = 300$, $ParTol = 1e-9$, and $GradTol = 1e-9$.

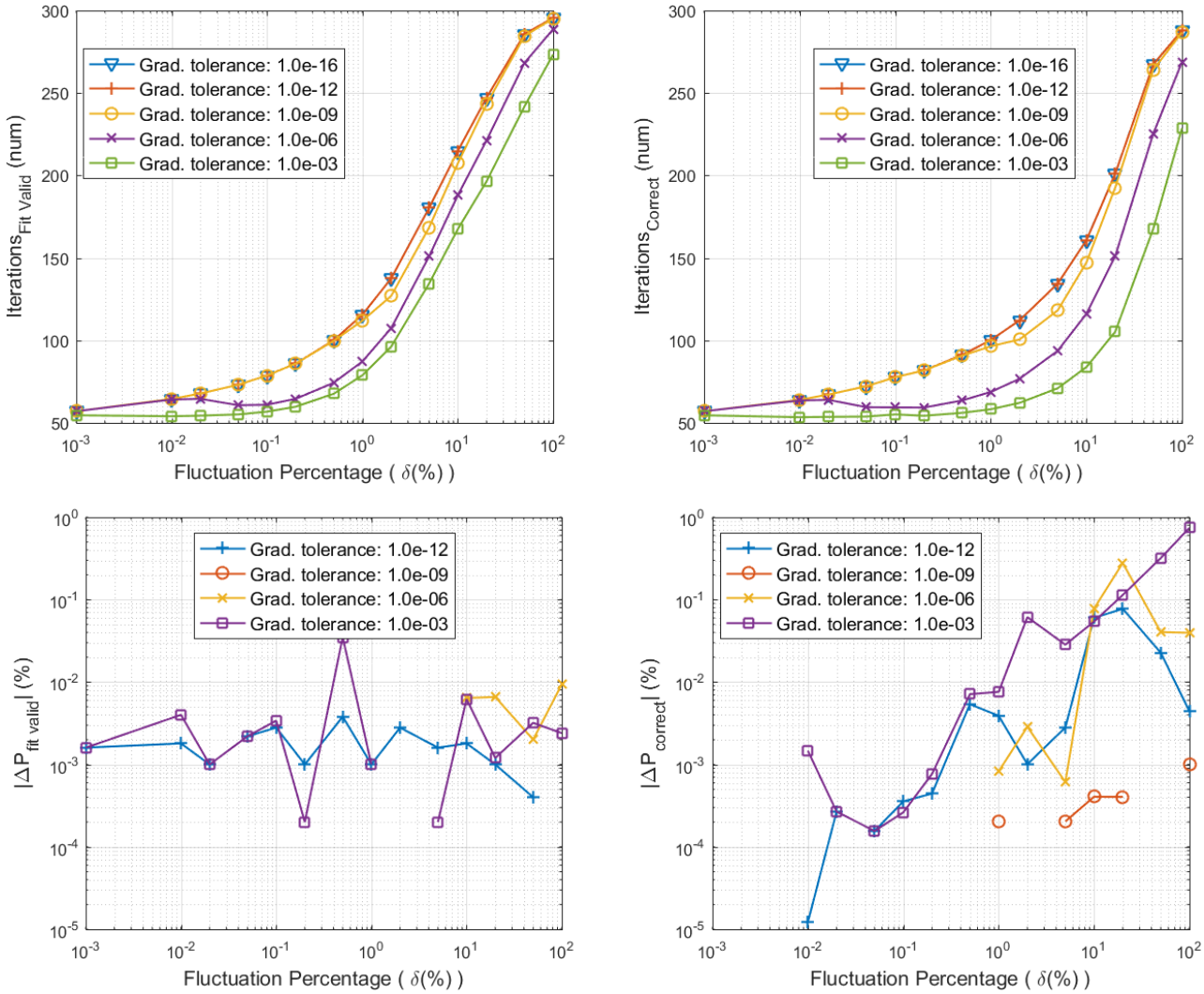

Figure Annex 11. (Top) The average number of iterations of valid and 'correct' results and (Bottom) probability differences of valid ($|\Delta P_{fit\ valid}|$) and 'correct' ($|\Delta P_{correct}|$) estimation obtained with different gradients of parameters ($GradTol$).

ANNEX 4. THE EXPECTATION MAXIMIZATION ALGORITHM

In this work, the Expectation Maximization (EM) was used to identify solutions as 'correct' or 'incorrect' (i.e. selecting the global or local minimum, respectively). Our implementation of the EM algorithm uses the library of [Tsui and Boedigheimer, 2006]. The EM algorithm [Depmster et al., 1977] obtains the Maximum Likelihood (ML) estimates of PDFs statistical parameters [Moon, 1996]. The process first determines the PDFs of the GMM of the ion composition error estimate ($\mathcal{E}_p = p_{true} - \hat{p}$) of each simulation (see Equation (18)). Once these statistical distributions are obtained, a clustering of is done based on the distance of each solution to these distributions.

The EM algorithm is commonly used in signal processing application when some data is missing or cannot be directly accessed [Moon, 1996]. The convergence of the EM algorithm solution is guaranteed because it increases the likelihood function at every iteration [Depmster et al., 1977] [Moon, 1996]. Nevertheless, this algorithm suffers from the local maximum selection problem in cases of likelihood functions with multiple maxima [Moon, 1996].

The EM algorithm is based on two different phases [Depmster et al., 1977] [Moon, 1996]:

1) an Expectation step (*E-step*) in which the expected value of the likelihood function is calculated assuming a particular set of statistical parameters; and
2) a Maximization step (*M-step*) that selects the statistical parameters that maximize the likelihood function.

The algorithm iterates these two steps (*E-step* and *M-step*) until reaches the configured likelihood convergence threshold or the maximum number of iterations [Moon, 1996]. In this work, the EM algorithm was configured with a maximum tolerance of the likelihood function step of 1e-16, and a maximum of 500 iterations.

The EM algorithm requires an initial guess of the statistical parameters of the probability density functions (μ_i and σ_i^2) and mixture weight parameter (α). The mean of the 'correct' distribution was fixed to zero in all cases ($\mu_0 = 0$). Due to the local maximum selection problem [Moon, 1996], it was required to configure the initial set of statistical parameters near the expected solution. The initial parameters of the 'correct' and 'incorrect' distributions were configured as $(\mu_0, \sigma_0^2)_{initial} = (0, (\delta/10)^2)$ and $(\mu_1, \sigma_1^2)_{initial} = (2p_{true} - 1, (\delta/10)^2)$, respectively. The initial search parameter of the mixture weight was set to $\alpha_{initial} = 1 - (log_{10}(\delta) + 3)/10)$.

The optimization algorithm was not allowed to search parameters outside of the configured search range. Therefore, wrongly estimated solutions were obtained near the limits of the search space range (i.e. $p = 0$ or $p = 1$). To avoid a wrong determination of the statistical distributions, estimated values found near of the maximum and minimum ion composition values were not considered in the EM algorithm analysis (i.e. $\hat{p} \leq \epsilon$ and $\hat{p} \geq 1 - \epsilon$, with $\epsilon = 0.0025$).

Wrong estimates of the statistical parameters could occur when only one distribution is present. Therefore, if both distributions obtained similar statistical means (i.e. $|\mu_0 - \mu_1| \leq 0.2$) it was assumed that the global distribution could be represented by a single 'correct' Gaussian distribution. In this case, the estimation is repeated fixing the GMM mixture weight to $\alpha = 1$.