Determination of the Signal Fluctuation Threshold of the Temperature-Ion Composition Ambiguity Problem Using Monte Carlo Simulations

Miguel Martínez-Ledesma[1] 🆔 and **Marcos Andrés Díaz Quezada**[1] 🆔

[1]Electrical Engineering Department, Faculty of Physical and Mathematical Sciences, University of Chile, Santiago, Chile

Abstract 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 in almost noiseless scenarios, not commonly feasible during routine monitoring. In this study, 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 nonlinear least squares optimization algorithm. Results are shown as probability curves of *valid convergence* and *correct estimation*. We 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.

1. Introduction

Incoherent scatter radar (ISR) is a ground-based sounding technique used to estimate the most relevant plasma parameters at different altitudes of the ionosphere (Evans, 1969). Secondary ionospheric parameters can also be inferred from directly estimated parameters (i.e., electric fields, conductivities, neutral winds, and temperatures; Beynon & Williams, 1978). This makes ISR one of the most powerful remote sounding methods to study the ionospheric profile from altitudes typically between 80 and ~1,000 km.

Ionospheric parameter estimations are obtained from the analysis of the ISR backscattered signal autocorrelation function (ACF) or its Fourier transform, the incoherent scatter spectra (ISS). The most common method of ISR analysis is the "range gate" (or "height-by-height") technique, in which a nonlinear least squares (NLLS) optimization algorithm is used independently at each altitude range to fit the measured signal with a theoretical ISR model (Lehtinen & Huuskonen, 1996). Alternatively, the "full profile" technique computes the entire altitude profile by treating the incoherent scatter method as an inversion problem (Holt et al., 1992; Hysell et al., 2008; Lehtinen et al., 1996; Nikoukar et al., 2008). This latter method provides a more accurate parameter estimation solution. Nevertheless, this method is more sensitive to noise and interferences and highly dependent on the initial guess of plasma variables to converge to the correct solution (Hysell et al., 2009). Furthermore, solutions may not be unique or even become unstable, requiring the application of regularization techniques (Aster et al., 2013). These techniques apply theoretical models and assumptions about the behavior of the plasma variables in the ionosphere that may be inappropriate in certain cases.

Estimating plasma parameters is not always directly solvable, and several problems arise from the analysis of the signal backscatter (Nikoukar et al., 2008). Of particular relevance is the "temperature-ion composition ambiguity" (TICA) effect, which is generated by the mixture of different ions in the ionosphere (Oliver, 1979). This ambiguity is related to the dependency of the theoretical ISR spectrum on the thermal speed

of different ion species, which is, in turn, proportional to the square root of the ratio of ion temperature (T_i) and mass (m_i; Oliver, 1979; Vallinkoski, 1988). Different combinations of molecular and atomic ion compositions and temperatures generate similar values of $\sqrt{T_i/m_i}$, hindering the correct discrimination of signals (Aponte et al., 2007). This effect is currently a relevant issue under discussion that affects the determination of long-term trends of ionospheric and thermospheric variables (Perrone & Mikhailov, 2017; Perrone & Mikhailov, 2018; Zhang et al., 2018).

The TICA problem has been widely studied in the literature using different methods. 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 plasma parameters (Evans & Oliver, 1972; Waldteufel, 1971). 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). Oliver (1979) developed a widely used parametric model of ion composition and temperature based on rocket and satellite measurements. More recently, models of ion composition have been created to apply the full profile inversion method unambiguously (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 using theoretical models is to provide complementary information extracted from the plasma line (Akbari et al., 2017). Both the plasma line and the ion acoustic frequency bands of the ISR provide complementary and simultaneous information of ionospheric plasma parameters (Aponte et al., 2007; Waldteufel, 1971; Wand, 1970). As most of the signal backscatter power is contained in the ion acoustic frequency band, plasma parameters have typically been extracted from the analysis of this band. Alternatively, the plasma line is a narrow frequency peak in the megahertz range, mainly dependent on electron density, electron temperature, and magnetic aspect angle (Yngvesson & Perkins, 1968).

Several methods that combine information measured from plasma line and the ion acoustic spectral bands have been successfully implemented. For a radar transmission frequency similar to that used at Arecibo Radio Observatory (i.e., 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). Using the electron density measurement, it is possible to obtain the electron-to-ion temperature ratio (T_e/T_i) analyzing the total power received (Pr) at the antenna (Aponte et al., 2007; Wand, 1970). Using the asymmetry of the upshifted and downshifted resonance frequencies of the plasma line, it is also possible to estimate the electron temperature (Nicolls et al., 2006). Alternatively, when the radar transmission frequency is sufficiently elevated (e.g., 933 MHz of the European Incoherent Scatter, EISCAT, UHF radar system), the plasma line resonance frequency depends also on the electron temperature. In this case, it is possible to estimate unambiguously the ion composition by fitting together the ion acoustic band and the resonance frequency (Bjørnå & Kirkwood, 1988).

Another method was proposed by Oliver (1979) for unambiguous estimation of ISR parameters. From the analysis of signals with very small fluctuations, two possible solutions of ion composition can be obtained. 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 obtained only partial success mainly because it required the use of almost noiseless signals (Oliver, 1979). There are a variety of sources that add noise-like variability to the ISR signal (Lehtinen & Huuskonen, 1996): the intrinsic stochastic nature of plasma particles that contribute randomly to the radar backscatter, the internal thermal noise of the radar receiver, and the sky noise. Statistical variability of ISR signals can be reduced by averaging many radar signals. This can be accomplished using multiple radar frequencies simultaneously (Sulzer, 1986a) or by postintegrating backscattered signals during long time periods (Lathuillere et al., 1983). However, the integration of signals coming from different radar pulses requires the assumption of stationary plasma conditions during the integration period (Farley, 1969). For long integration periods, this assumption can smooth and hide the dynamic phenomenon occurring in the ionospheric plasma. Lathuillere et al. (1983) and Lathuillere and Pibaret (1992) successfully demonstrated the efficacy of Oliver's (1979) method with experimental data. These studies reduced the fluctuation of signals by postintegrating multiple radar pulses with integration times of 5 min.

Recently, 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 high signal-to-noise ratio (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. Wu et al. (2015) 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. The PSO algorithm obtained much better estimations than the standard NLLS algorithm commonly used in ISRs. Results from Lathuillere et al. (1983) and Wu et al. (2015) demonstrate the feasibility of unambiguously estimating plasma parameters of signals with very small fluctuations. Nevertheless, no previous study has assessed the required signal fluctuation level to unambiguously estimate the ion composition and temperature parameters in the TICA problem.

Here we present a theoretical framework to quantify the ambiguity problem and determine the signal fluctuation threshold for which the TICA problem is unambiguously solved. We apply the Monte Carlo technique to the plasma parameter estimation of an NLLS optimization algorithm at different signal fluctuation levels. The results are presented as probability curves, which provide the *probability of valid convergence* and the *probability of correct determination of the global minimum* as a function of signal noise level. The impact of using a model to set the initial parameters of the optimization algorithm search is also analyzed by increasing the distance between the initial guess and the correct solution. To evaluate the effect of adding a priori information on the ambiguity threshold, different combinations of plasma parameters were analyzed. Plasma parameters known a priori are assumed to be obtained from the analysis of the plasma line frequency band and the total power received at the antenna. To determine which would be the most relevant a priori parameter to solve the ambiguity, the information provided by each plasma parameter was evaluated. Finally, to provide a realistic application, the effect of adding parameters with different levels of uncertainty was studied. The results of this study could be used to unambiguously analyze data obtained from previous experiments, properly configure future radar campaigns, and determine the design requirements for future ISR observatories.

This article is organized in five sections. Section 2 describes our simulation methods. Section 3 explains the statistical analysis performed. Section 4 includes the results and discussion, and section 5 indicates our conclusions with suggestions for operational improvements of the ISR estimation process.

## 2. Simulation Methods

2.1. Implementation of the ISR Model

To perform the estimation process, an ISR spectrum model was implemented based on the formulation of Kudeki and Milla (2011). The magnetic field aspect angle effect (Milla & Kudeki, 2011) was not included in the 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). 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 (Kudeki & Milla, 2011) was computed using the imaginary error function implemented in the Faddeyeva Dawson function 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 each fitting comparison, estimation analysis was done by directly comparing the difference between simulated and theoretical spectrums at each range (Wu et al., 2015). Theoretical ion acoustic spectra were obtained at frequencies between ±10 kHz.

2.2. Ion Composition Parameter

In this work two different ion species were considered (M^+ and O^+). Early rocket and satellite measurements demonstrated the predominance of NO^+ and O_2^+ molecular ions at low altitudes (Evans & Oliver, 1972; Hoffman et al., 1969). Because of the similarity of NO^+ and O_2^+ ion masses, they produce an almost identical T_i/m_i ratio value, making them indistinguishable to the ISR radar (Oliver, 1979). Consequently, the exact mixture ratio of those molecular ions is unknown. These molecular ions are assumed to behave as a single
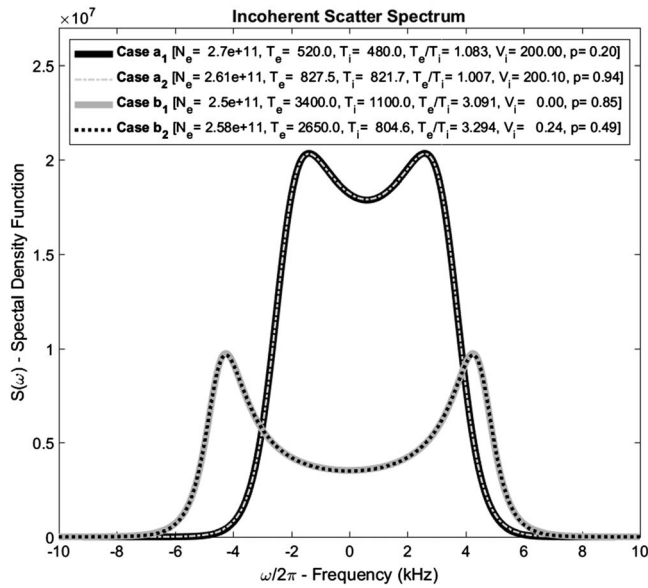
Figure 1. Two different incoherent scatter radar spectrums obtained ambiguously with two alternative sets of plasma parameters. Spectrums Case a_1 (black line) and Case a_2 (dotted gray line) were obtained with different plasma parameters displaced from the center frequency of the radar by ion drift velocity (V_i) of approximately 200 m/s. Alternatively, spectrums Case b_1 (gray line) and Case b_2 (dotted black line) were obtained without V_i but having higher values of T_e/T_i.

molecular specie (M^+) with mass 30.5 amu for a mixture of 25% O_2^+ and 75% NO^+ (Cabrit & Kofman, 1996; Lathuillere et al., 1983). At altitudes varying from ~130 to ~300 km, these molecular plasma constituents migrate to a plasma dominated by lighter atomic oxygen ions (O^+; Kelley & Heelis, 1989).

Due to the TICA problem, different combinations of molecular and atomic oxygen ion species obtain almost the same ion acoustic spectrum (Oliver, 1979). Two different possible estimations of plasma parameters are obtained ambiguously (Lathuillere et al., 1983; Oliver, 1979). In Figure 1 the TICA effect is shown for two different spectrums obtained with two different combinations of parameters. Under certain conditions, this mixture of ion species might be determined unambiguously (Aponte et al., 2007; Lathuillere et al., 1983; Oliver, 1979; Wu et al., 2015). The determination of the conditions required for an unambiguous estimation is the main focus of this study. The ion composition parameter p is defined as the molecular ion fraction, computed as the relative abundance of molecular ions with respect to the total ion concentration (Aponte et al., 2007; Lathuillere et al., 1983; Wu et al., 2015):

$$p = 1 - \frac{n(O^+)}{N_e}, \tag{1}$$

where $n(O^+)$ is the atomic oxygen density. Note that charge neutrality has been assumed (i.e., $N_e = N_i = n(O^+) + n(NO^+) + n(O_2^+)$). Note also that the ion composition range is $0 \leq p \leq 1$.

2.3. Noise Addition Scheme

To consider realistic radar conditions, simulated ISR spectrum signals ($\boldsymbol{m}$) were assumed to be describable as a function ($f(\boldsymbol{x})$) that depends on the plasma parameter vector ($\boldsymbol{x}$) and an additional noise contribution (ε; Vallinkoski, 1988):

$$\boldsymbol{m} = f(\boldsymbol{x}) + \boldsymbol{\varepsilon}, \tag{2}$$

where $f(\boldsymbol{x})$ function represents the theoretical backscatter signal spectra indicated in section 2.1. Both the theoretical signal and the noise contribution are vectors of data points with values distributed at the different frequencies of the spectra.

The ISR signal is a stochastic process due to the random nature of all plasma particles that contribute with backscatter (Lehtinen & Huuskonen, 1996). This stochastic characteristic provides noise-like fluctuations to the received signal. To obtain an accurate statistical estimate of ISSs (or ACFs), the signals from many radar pulses must be integrated, assuming stationary plasma characteristics during integration time (Farley, 1969). Additional noise contributions come from background electromagnetic radiations (i.e., sky noise) and from the internal receiver system noise (i.e., thermal noise; Lehtinen & Huuskonen, 1996). For strong radar SNR, the stochastic nature of the plasma dominates the signal fluctuation (Huuskonen & Lehtinen, 1996). Therefore, signals obtained with a strong backscatter from the ionosphere, due to high electron densities or to the use of high-power transmissions, have a signal-dependent noise (Nikoukar et al., 2008). Alternatively, for weak SNR conditions, the thermal and sky noises are responsible for the error (Sulzer, 1986a). In this work, weak SNR conditions were assumed, obtaining a noise independent from the signal.

The integration of many pulses of the signal provides Gaussian characteristics according to the central limit theorem (Vallinkoski, 1988). Therefore, the noise (ε) was assumed to be an additive Gaussian white noise (AWGN; Lehtinen & Huuskonen, 1996). Consequently, a Gaussian random noise with zero mean was added directly to the ISS spectral signal. The standard deviation (σ) of this random noise was calculated relative to the maximum absolute value of the spectrum amplitude ($\sigma = \delta(\%)/100 \cdot \max(|f(\boldsymbol{x})|)$), as in Aponte et al.

(2007). In our work, the signal fluctuation (δ) is assumed to be given by the root-mean-square error of the ACF estimator obtained by Farley (1969):

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

where SNR is the signal-to-noise ratio of the radar backscattered signal (Lu et al., 2016) at the corresponding altitude range and N is the integration length (i.e., the number of radar signals averaged).

A reduction of the signal fluctuation can be obtained by increasing the integration length or the received signal strength, as indicated in equation (3). Nevertheless, signals must have $SNR \ll 1$ to ensure that noise contributions are independent from the signal (Huuskonen & Lehtinen, 1996). Achievable signal fluctuation values under this weak SNR condition are shown in supporting information Figure S1 for different integration lengths and periods.

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. Consequently, 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 multipulse or alternating code experiments (Vallinkoski, 1988). Therefore, simulated spectra were assumed 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).

2.4. Plasma Parameters Estimation Method

2.4.1. Maximum Likelihood Estimator

The standard ISR plasma estimation process assumes signals with AWGN stochastic noise characteristics and a measurement variance-covariance matrix $\boldsymbol{C_m}$ (Swoboda et al., 2017; Vallinkoski, 1988). With sufficiently small SNR, the matrix $\boldsymbol{C_m}$ is assumed diagonal and only populated by measurement variances (σ_i^2; Erickson, 1998; Vallinkoski, 1988). In this case, the maximum likelihood estimator statistical criterion (Kay, 1993) used to determine the plasma parameters is equivalent to the standard least squares estimator given by (Erickson, 1998; Lathuillere et al., 1983):

$$\widehat{\boldsymbol{x}} = \underset{\boldsymbol{x}}{\text{argmin}}\left\{ (\boldsymbol{m} - f(\boldsymbol{x}))^T \boldsymbol{C_m^{-1}} (\boldsymbol{m} - f(\boldsymbol{x})) \right\} = \underset{\boldsymbol{x}}{\text{argmin}}\left\{ \sum_{i=1}^{M} \frac{(m_i - f_i(\boldsymbol{x}))^2}{\sigma_i^2} \right\}, \tag{4}$$

where m_i, σ_i^2, and $f_i(\boldsymbol{x})$ are the M vector components of the measured signal, measured signal variance, and theoretical model function, respectively.

2.4.2. Reduced Chi-Square Cost Function

The minimization argument of the right-hand side of equation (4) is known as the chi-square cost function (χ^2), and the minimization process is identified as an unconstrained weighted least squares fitting (Erickson, 1998). Unbiased estimates of signals with AWGN characteristics and known measurement variances obtain a cost function value approximately equal to the number of degrees of freedom (DoF) of the estimation problem (Bevington & Robinson, 2003; Taylor, 1997). In this study, the reduced chi-square cost function criteria was used to normalize the resulting values ($\chi_r^2 = \chi^2/DoF$). The χ_r^2 criteria allow the *goodness of the fit* to be statistically determined to discriminate the convergence of the estimated solution (Bevington & Robinson, 2003). The DoF was assumed equal to the input signal vector length minus the number of parameters of the estimation problem (P; i.e., $DoF = M - P$). This assumption is commonly used in standard error analysis studies (Bevington & Robinson, 2003; Taylor, 1997), although it is known that this is an approximation only valid for linear models with linearly independent basis functions (Andrae et al., 2010).

Figure 2 shows an example of the χ_r^2 cost function for different projections of T_e, T_i, and p parameters, assuming that N_e and V_i are known a priori parameters. These images were generated selecting the minimum χ_r^2 value of the nonrepresented parameter, allowing a 2-D representation of a 3-D structure. These simulations were done for different fluctuation levels to represent the impact of the noise addition. Figure 2 shows that in
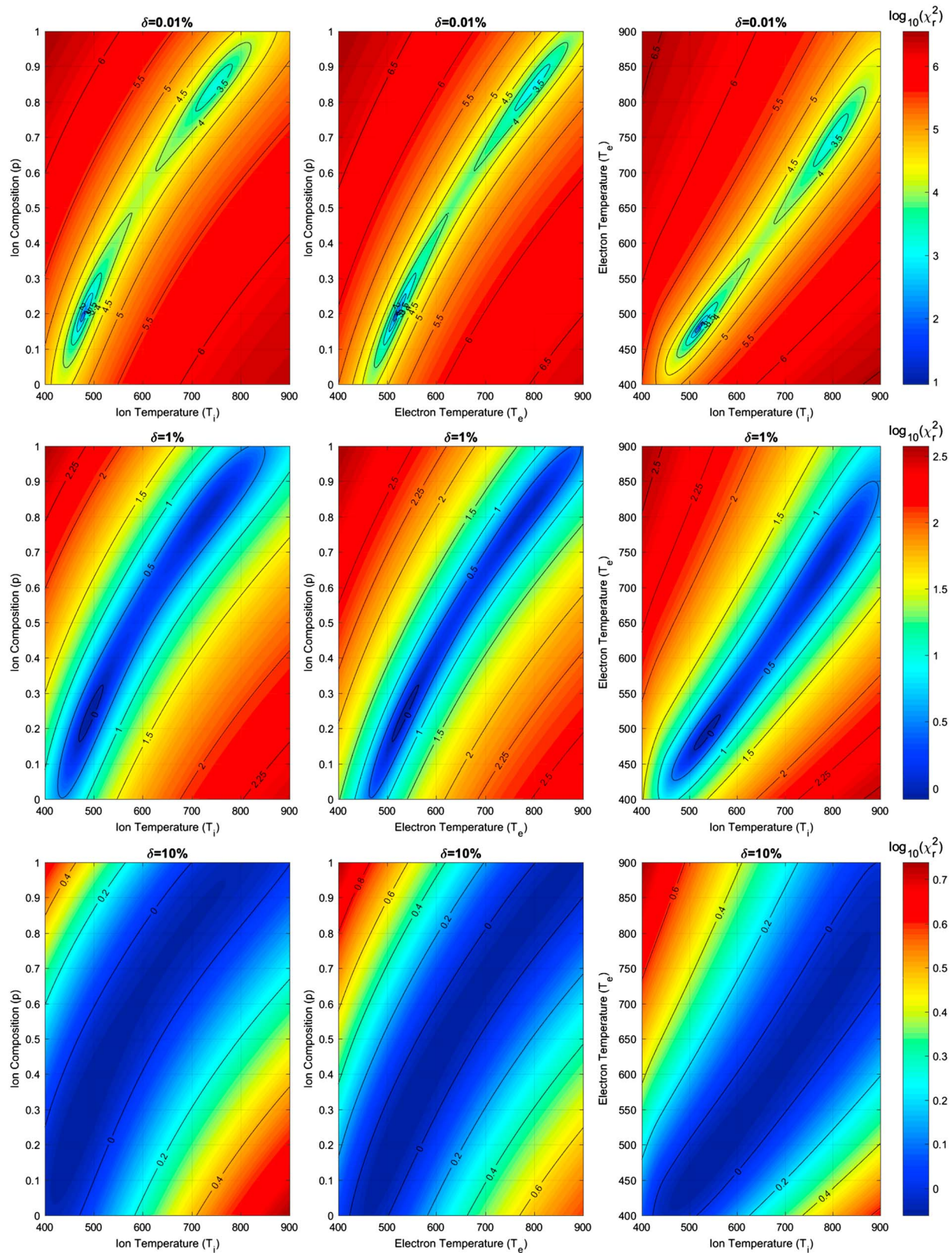
Figure 2. Graphic representation of the χ_r^2 cost function minimum value (in logarithmic scale) for different combinations of plasma parameters at different noise levels (top $\delta = 0.01\%$, middle $\delta = 1\%$, and bottom $\delta = 10\%$). Data have been obtained from the simulation of plasma parameters of spectrum Case a_1 of Figure 1 (T_e=520 °K, T_i=480 °K, and $p = 0.2$), assuming N_e known a priori and consequently not represented.

almost noiseless scenarios ($\delta = 0.01\%$) one of the two minimums of the χ_r^2 cost function has a much smaller value (i.e., the global minimum). This allows the determination of the "correct" set of parameters by selecting the solution with the lowest χ_r^2 value. Furthermore, this figure also shows that it is more difficult to discriminate the "correct" solution by selecting the lowest χ_r^2 value in noisy scenarios ($\delta = 1\%$ and 10%).

The existence of two minimums in Figure 2 verifies the presence of the TICA problem. These two minimums are located at almost opposite ion composition values symmetric with respect to $p = 0.5$, as initially indicated by Lathuillere et al. (1983) and more recently verified by Wu et al. (2015). As the "incorrect" ion composition solution ($p_{\text{incorrect}}$) is an almost symmetric solution of the "correct" solution (p_{correct}), those solutions should be associated as $p_{\text{incorrect}} \approx 1 - p_{\text{correct}}$. Consequently, the "incorrect" solution would be related to the abundance of atomic oxygen ions ($p_{\text{incorrect}} \approx n(\text{O}^+)/N_e$; see equation (1)). Furthermore, these two minimums are obtained because the ISR spectrum depends 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 are also double those of atomic temperatures ($T_{M^+} \approx 2T_{\text{O}^+}$), as indicated by Oliver (1979).

2.4.3. NLLS Optimization Algorithm Used

The minimization of the χ_r^2 cost function is computed using an NLLS optimization algorithm: the Levenberg-Marquardt (L-M) algorithm (Levenberg, 1944; Marquardt, 1963). The L-M has been selected because it 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) methods. The L-M algorithm is a standard line-search technique for NLLS unconstrained problems that combines different methods depending on the distance to the optimum solution (Gavin, 2019). The search step is iteratively updated using a gradient descent method when the solution is far from the optimum or using a Gauss-Newton method near the minimum cost function.

The L-M optimization algorithm suffers from the local minimum determination problem. The determination of the solution is subject to the initial search parameters of the algorithm, the parameter search range, and the algorithm configuration (Gavin, 2019). When several minimums of the cost function exist, NLLS search algorithms select the first minimum found during the search and consequently may not find the "correct" solution (i.e., global minimizer; Wu et al., 2015). To reduce the effects 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 PSO algorithm (Wu et al., 2015), which have not been implemented in our work but will be analyzed in future studies.

In our study, plasma parameters were estimated using an optimization algorithm for range gate ISS analysis, as in Aponte et al. (2007) and Wu et al. (2015). The implementation of the L-M algorithm to minimize the χ_r^2 cost function was based on the library created by Gavin (2019). Resolution problems arise in the calculus of the Jacobian matrix due to differences of orders of magnitude of plasma parameters. To avoid this resolution problem, electron density parameter values were converted to a logarithmic scale ($\log_{10}(N_e)$) in the estimation process, as in Cabrit & Kofman (1996). Alternatively, the search range of the parameters ($\boldsymbol{x}_{\text{search min}} \leq \hat{\boldsymbol{x}} \leq \boldsymbol{x}_{\text{search max}}$) was configured to be equal to the input parameter ranges of the Monte Carlo simulation (explained in section 2.5). 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 °K. The configured tolerances and search stop conditions of the L-M algorithm (Gavin, 2019) were determined by simulations shown in supporting information Text S1 and Figure S2.

2.5. 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 1,000 or 2,000 different true input parameters ($\boldsymbol{x}_{\text{true}}$) uniformly selected between the following parameter ranges: electron density, $10^9 \leq N_e \leq 10^{12}$ m^{-3}, electron temperature, $300 \leq T_e \leq 5000$ °K, ion temperature, $300 \leq T_i \leq 3000$ °K, electron-to-ion temperature ratio, $0.1 \leq T_e/T_i \leq 5$, ion drift velocity, $-250 \leq V_i \leq 250$ m/s, and ion composition, $0 \leq p \leq 1$. 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 ranges of T_e/T_i and V_i used were larger than the typical measurements obtained at standard midlatitude ionospheric conditions shown in (Scherliess et al., 2001; Wand, 1970).

In this work, the ion drift velocity parameter was assumed to be known a priori. As V_i only generates a Doppler shift, it does not affect the determination of other parameters, and it can be estimated independently (Wu et al., 2015). Results of Monte Carlo simulations shown in section 4.3 verify the ion drift parameter estimation independence.

To study the effect of increasing the noise level of ISR signals in the estimation of the TICA problem, Monte Carlo simulations at different signal fluctuation levels were done. First, different ISS signals were created using the uniformly selected true input plasma parameters and the theoretical ISR spectrum model explained in section 2.1. The noise contribution scheme indicated in section 2.3 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. The same set of uniformly selected true input plasma parameters was used in all simulations, allowing a direct comparison of results for different noise fluctuation levels. Finally, the plasma parameters were estimated using the L-M optimization algorithm, minimizing the χ_r^2 cost function, as indicated in section 2.4. To provide sufficient statistical representation, 500 repetitions of the estimation process for each set of input parameters were done with a different random noise added.

2.6. Uniformly Selected Initial Parameters

The solution of the L-M 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 (Bilitza et al., 2017) or the Mass Spectrometer Incoherent Scatter Radar (Picone et al., 2002) models. 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 of the Monte Carlo simulation. These initial parameters were uniformly selected from the full search range of parameters indicated in section 2.4.3 (i.e., $[\boldsymbol{x}_{\text{search min}}, \boldsymbol{x}_{\text{search max}}]$). Identical uniformly selected initial parameters were used in all simulations to allow a direct comparison of results of the different simulations done.

Furthermore, to determine the effects derived from the use of an inaccurate initial guess, Monte Carlo simulations were done with different configurations of initial parameters and at different signal fluctuation levels. Results of these simulations are shown in section 4.1. These simulations were done for 1,000 different true input values of N_e, T_e, T_i, and p parameters. This parameter configuration corresponds to the typical estimation process for ISR radars, when no a priori information is provided.

The initial parameters ($\boldsymbol{x}_{\text{initial}}$) of simulations of section 4.1 were selected randomly using a uniform distribution centered on the true input plasma parameters. The selection range was configured relative to the full search range of parameters as follows:

$$\boldsymbol{x}_{\text{initial}} = \boldsymbol{x}_{\text{true}} + \frac{\beta(\%)}{100} \cdot (\boldsymbol{x}_{\text{search max}} - \boldsymbol{x}_{\text{search min}}) \cdot (\text{rand}(n) - 0.5), \tag{5}$$

where β is the initial parameters range percentage ($\beta(\%) \in [0,100]$) and $\text{rand}(n)$ 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 follows:

$$\boldsymbol{x}_{\text{initial}} \in [\boldsymbol{x}_{\text{true}} - \Delta/2, \boldsymbol{x}_{\text{true}} + \Delta/2], \tag{6}$$

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

2.7. Addition of Plasma Line Information

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. The combinations of parameters studied were the following: (a) no a priori information, (b) N_e parameter known a priori, (c) T_e/T_i and N_e parameters known a priori, and (d) T_e and N_e parameters known a priori. Similar fittings were made by Wu et al. (2015) for different combinations of i-unknown parameters (i = 1, 2, and 3). 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 (Aponte et al., 2007; Waldteufel, 1971; 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 (as shown in Figure 9). The analysis of information with uncertainty in the a priori known parameters was studied separately and results are shown in section 4.4.

3. Statistical Analysis of Ambiguous Results

3.1. Determination of Convergence

Results from Monte Carlo simulations were analyzed statistically to determine the impact of the ambiguity in different noise scenarios. In some cases, the optimization algorithm did not converge to a valid solution. To identify which solutions converged to a minimum of χ_r^2, the goodness of the fit was calculated based on the statistical distribution of the cost function (Bevington & Robinson, 2003; Taylor, 1997).

Estimated parameters ($\widehat{x}$) with χ_r^2 smaller than a maximum cost function value ($\chi_r^2 \lesssim \chi_{r,\max}^2$) were considered to have a valid convergence fit. Due to the random nature of the estimated signals, the χ_r^2 cost function follows a chi-square (f_{χ^2}) statistical distribution (Andrae et al., 2010; Bevington & Robinson, 2003). The $\chi_{r,\max}^2$ value has been selected to have a probability of $P\left(x > \chi_{r,\max}^2; x \sim f_{\chi^2}\right) = 0.00317\%$, which corresponds to the probability of the 4σ criterion of a normally distributed function (i.e., $P\left(x > \mu + 4\sigma; x \sim \mathcal{N}(\mu, \sigma^2)\right) = 0.00317\%$). 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. 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.

The f_{χ^2} statistical distributions and the associated estimation errors of parameters are shown in supporting information Text S2 and Figures S3–S5 for different signal fluctuation values.

3.2. Determination of Correctness

Due to the ambiguity problem, in some cases a local minimum of the cost function was selected instead of the global minimum. A solution is selected as "correct" and "incorrect" depending on the distance between the estimated ($\widehat{x}$) and the true input parameters (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 from the simulation results is equivalent to the selection of different data groups in a clustering analysis (Jain et al., 1999).

3.3. Expectation Maximization Algorithm

In this work the expectation maximization (EM) algorithm was used (Dempster 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). Our implementation is based on the EM algorithm library written by Tsui and Boedigheimer (2006). The convergence of the EM algorithm solution is guaranteed because it increases the
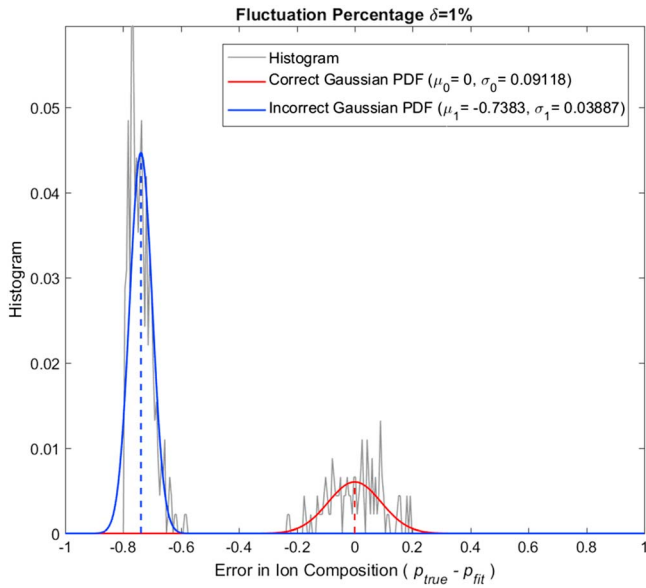
Figure 3. (gray line) Histogram of error of ion composition estimations ($\varepsilon(p)$ $= p_{\text{true}} - \widehat{p}$) obtained from a Monte Carlo simulation with signal fluctuation $\delta = 1\%$. Input plasma parameters of this simulation were those of spectrum Case a_1 of Figure 1, assuming electron density and ion drift parameters known a priori. The red and blue curves show respectively the "correct" and "incorrect" distributions of the Gaussian mixture model calculated using the expectation maximization algorithm. PDF = probability density function.

likelihood function at every iteration (Dempster et al., 1977; Moon, 1996). Nevertheless, the EM algorithm suffers from the local maximum selection problem in cases of likelihood functions with multiple maxima (Moon, 1996).

A total of 500 repetitions of the estimation process was 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 the following:

$$f_{\text{GMM}}\left(x|\alpha, \mu_0, \sigma_0^2, \mu_1, \sigma_1^2\right) = \alpha \cdot \mathcal{N}\left(x|\mu_0, \sigma_0^2\right) + (1-\alpha) \cdot \mathcal{N}\left(x|\mu_1, \sigma_1^2\right), \quad (7)$$

where α is the weight of the mixture distribution ($\alpha \in [0, 1]$) and $\mathcal{N}\left(x|\mu_i, \sigma_i^2\right)$ is the probability density function (PDF) of the Gaussian or Normal distribution where μ_i and σ_i^2 are the mean and variance of the distribution, respectively. Therefore, $\mathcal{N}\left(x|\mu_0, \sigma_0^2\right)$ and $\mathcal{N}\left(x|\mu_1, \sigma_1^2\right)$ are the PDFs of "correct" ($\widehat{x}_{\text{correct}}$) and "incorrect" ($\widehat{x}_{\text{incorrect}}$) results, respectively.

In this work, the EM algorithm was used to analyze the statistical distribution of the ion composition estimation error ($\mathcal{E}_p = p_{\text{true}} - \widehat{p}$, where p_{true} is the true input ion composition and $\widehat{p}$ 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 (Lathuillere et al., 1983; Oliver, 1979; Wu et al., 2015). These two ion composition solutions correspond to the "correct" and "incorrect" Gaussian distributions of the mixture model of equation (7). As the results obtained with a small ion composition error were considered correct solutions ($\widehat{p}_{\text{correct}} \approx p_{\text{true}}$), the "correct" distribution mean is approximately equal to 0 ($\mu_0 \approx 0$). Therefore, to improve the accuracy of the determination of the correctness in noisy scenarios, in this work the mean of the "correct" distribution has been fixed to 0 in the EM algorithm. Alternatively, the "incorrect" distribution has a mean value always different from 0 ($\mu_1 \not\approx 0$). Due to the symmetry of the solutions with respect to 0.5 (Lathuillere et al., 1983; Wu et al., 2015), the "incorrect" distribution mean value should be approximately equal to $\mu_1 \approx 2p_{\text{true}} - 1$. Estimated parameters were clustered as "correct" or "incorrect" based on the distance of each ion composition estimation error to the GMM PDFs. The clustering results of each Monte Carlo simulation can be seen in supporting information Figures S12, S17, S22, S27, S32, S37, S42, S47, S52, and S57.

Figure 3 shows the histogram of ion composition estimation errors ($\mathcal{E}_p$) and the corresponding GMM PDFs calculated by the EM algorithm, obtained from the Monte Carlo estimation of plasma parameters indicated as Case a_1 in Figure 1. Specific configurations and implementation details of the EM algorithm are indicated in supporting information Text S3.

3.4. Probabilities of Convergence and Correctness

Different types of probabilities were calculated to study 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 ($P_{\text{fit valid}}$), defined as the probability of finding a minimum of the χ_r^2 cost function by the L-M optimization algorithm, was calculated as follows:

$$P_{\text{fit valid}} = N_{\text{fit valid}}/N_{\text{total}}, \quad (8)$$

where $N_{\text{fit valid}}$ is the number of valid convergence results from the simulation ($\chi_r^2(\widehat{x}) \leq \chi_{r,\max}^2$) and N_{total} is the total number of parameters simulated. Note that $N_{\text{total}} = N_{\text{MC}} \cdot N_{\text{rep}}$, where N_{MC} is the number of different input parameters of the Monte Carlo simulation (i.e., $N_{\text{MC}} = 1{,}000$ or $2{,}000$ depending on the simulation) and N_{rep} is the number of repetitions of each input parameter of the simulation (i.e., $N_{\text{rep}} = 500$).

Alternatively, the probability of "correct" estimation (P_{correct}), which represents the probability to successfully select the global minimum of the χ_r^2 cost function having already converged to a valid solution, was calculated as follows:

$$P_{\text{correct}} = N_{\text{correct}}/N_{\text{fit valid}}, \tag{9}$$

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

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

$$P_{\text{fit valid\&correct}} = P_{\text{fit valid}} \cdot P_{\text{correct}} = N_{\text{correct}}/N_{\text{total}}. \tag{10}$$

Note that $P_{\text{fit valid \& correct}}$ represents the probability of solving a set of parameters and estimating it without ambiguity and consequently indicates the total probability of unambiguous estimation.

4. Results and Discussion

4.1. Effect of the Initial Parameters Uncertainty

4.1.1. Accurate Initial Guess

Figure 4 shows the results of Monte Carlo simulations with initial parameter range percentages of $\beta = 1\%$, 5%, 10%, 20%, 30%, 40%, 50%, and 100%, demonstrating the effects of increasing the uncertainty on the initial parameters.

The most remarkable result 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., $P_{\text{correct}} \approx 100\%$) were obtained only with signal fluctuations smaller than $\delta < 0.1\%$. This result implies the existence of a threshold on the noise level that can be tolerated to completely solve the TICA problem, even when using a precise ionospheric model as initial guess. To obtain a $P_{\text{correct}} \geq 95.45\%$ (i.e., a 2σ probabilistic criterion), the signal fluctuation threshold was $\delta_{\text{th}(\beta = 1\%)} = 0.54\%$. Alternatively, P_{correct} decayed to less than 50% for signal fluctuations $\delta > 5\%$, indicating that at highly noisy scenarios it was more probable to select the "incorrect" solution. On the other hand, the $P_{\text{fit valid}}$ was almost 100% in all noise scenarios, indicating that the L-M optimization algorithm was always able to converge to a minimum of the cost function.

4.1.2. Increasing the Uncertainty of Initial Parameters

The increase of the uncertainty of the initial parameters (β) produces a reduction of both probabilities $P_{\text{fit valid}}$ and P_{correct} (Figure 4). Nevertheless, the reduction of $P_{\text{fit valid}}$ 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 inverse proportionality of $P_{\text{fit valid}}$ to β occurs because increasing the uncertainty of the initial parameters is equivalent to increasing the distance between the initial and the true input parameters, making it more difficult to find the global minimum. This effect is shown in supporting information Figure S6, where the number of computing iterations is proportional to the uncertainty of the initial parameters (β).

Nevertheless, the decrease of $P_{\text{fit valid}}$ was found only at the low fluctuation regime. A decrease of signal fluctuation is equivalent to a reduction of the estimated signal variance (σ^2) in equation (4), resulting in higher χ_r^2 cost function values. It is assumed that the increase of χ_r^2 generated nonconvergent "incorrect" estimates ($\chi_{r,\text{incorrect}}^2 > \chi_{r,\text{max}}^2$) at the low fluctuation regime. This effect is shown in the example of Figure 2 (top), where the χ_r^2 cost value of the local minimum was larger than $\chi_{r,\text{max}}^2$ for small signal fluctuations ($\delta = 0.01\%$). This effect reduces the number of convergent solutions and increases the number of "correct" estimates when the signal fluctuation level decreases, as shown in Figure 4.
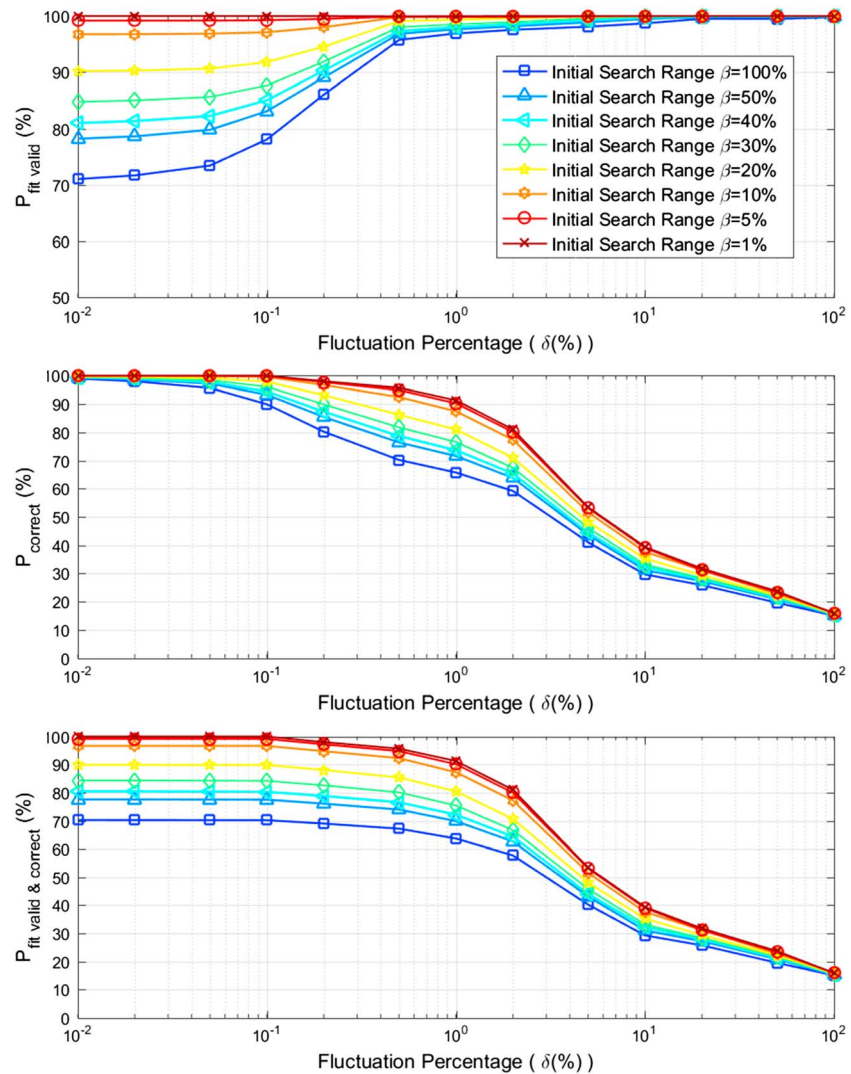
Figure 4. Probability of convergence ($P_{\text{fit valid}}$) and probability of "correct" estimation (P_{correct}) at different signal fluctuation percentages ($\delta[\%]$) obtained by simulations of the estimation of four plasma parameters (N_e, T_e, T_i, and p parameters) without parameters given a priori, for different ranges of initial search parameter selected uniformly around the true input value (β [%]).

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 level of uncertainty of initial parameters. Alternatively, at very small signal fluctuations ($\delta \leq 0.05\%$) an unambiguous estimation ($P_{\text{correct}} \approx 100\%$) is obtained independently of the level of initial guess uncertainty. 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 ($\delta_{\text{th}(\forall\beta)} \approx 0.05\%$ to obtain a $P_{\text{correct}} \geq 95.45\%$).

In summary, the use of ionospheric models to determine initial plasma parameters leads to obtaining erroneous plasma parameters unless signals are almost noiseless. 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.

4.1.3. Proposed Estimation Technique

To obtain results independent of the configuration of initial parameters, the use of a Monte Carlo selection scheme is suggested. This technique consists of executing the optimization algorithm several times with
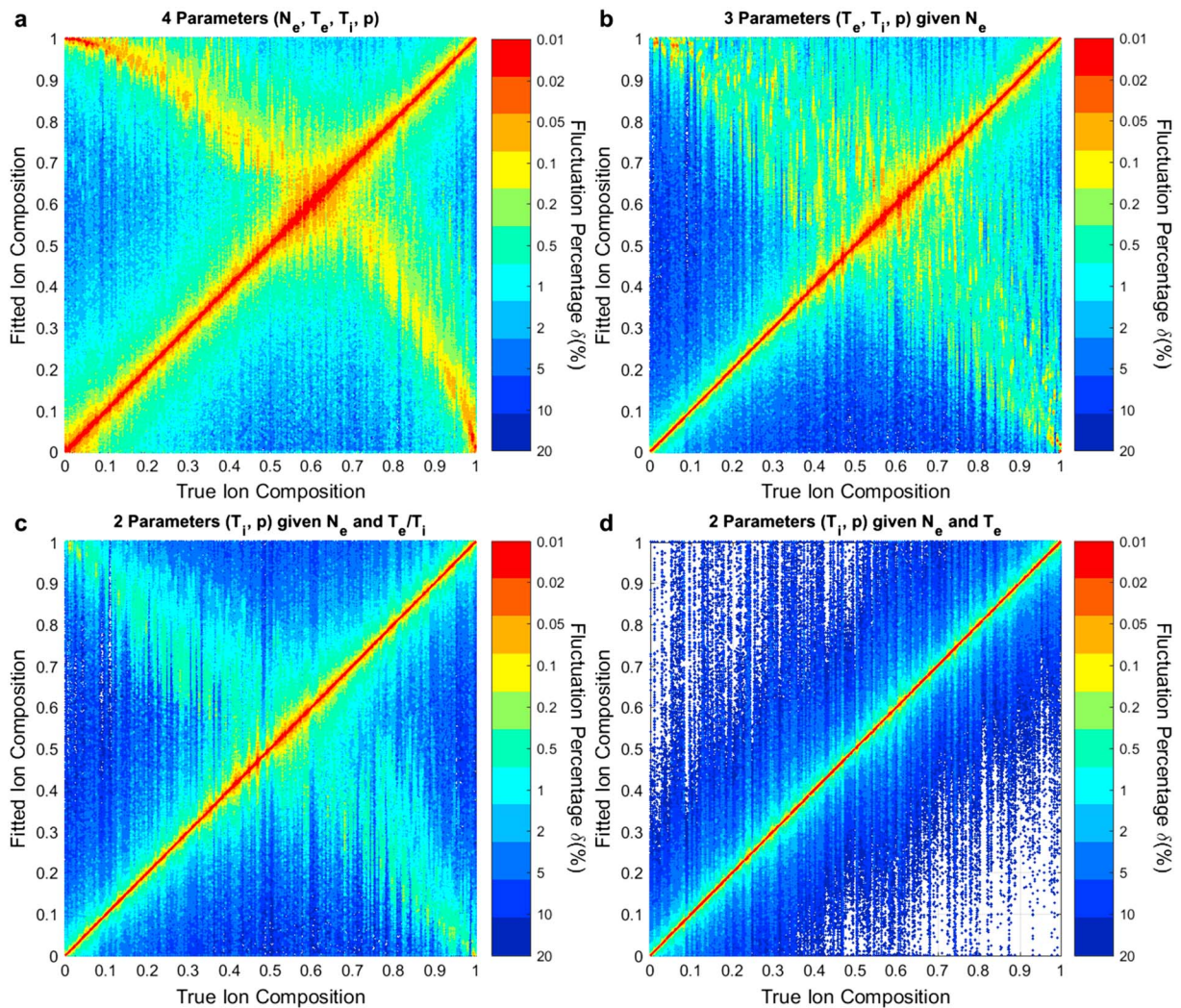
Figure 5. Scatter plot of estimated and input values of ion composition, obtained from the analysis of different combinations of parameters known a priori from the plasma line: (a) without a priori information, (b) given N_e, (c) given N_e and T_e/T_i, and (d) given N_e and T_e. Each color represents results obtained by simulations with a particular signal fluctuation percentage ($\delta[\%]$).

different initial parameters uniformly selected from the global search space of 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.

4.2. Impact of Adding A Priori Information From Plasma Line

4.2.1. Ambiguous Ion Composition Estimate

Monte Carlo simulations were done to analyze the addition of plasma line information without uncertainty. Figure 5 shows the estimated ion composition parameters obtained from simulations of 2,000 different true input plasma parameters with the following combinations of parameters: (a) four unknown parameters (N_e, T_i, T_e, and p); (b) three unknown parameters (T_i, T_e, and p) given a priori N_e; (c) two unknown parameters (T_i and p) given a priori N_e and T_e/T_i; and (d) two unknown parameters (T_i and p) given a priori N_e and T_e. In this figure, different colors represent ion composition values estimated at different signal fluctuations.

In the study case d of Figure 5, 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 noise scenarios. This result implies that this combination of a priori known parameters solved the ambiguity problem, as proposed by Bjørnå and Kirkwood (1988) and Nicolls et al. (2006).
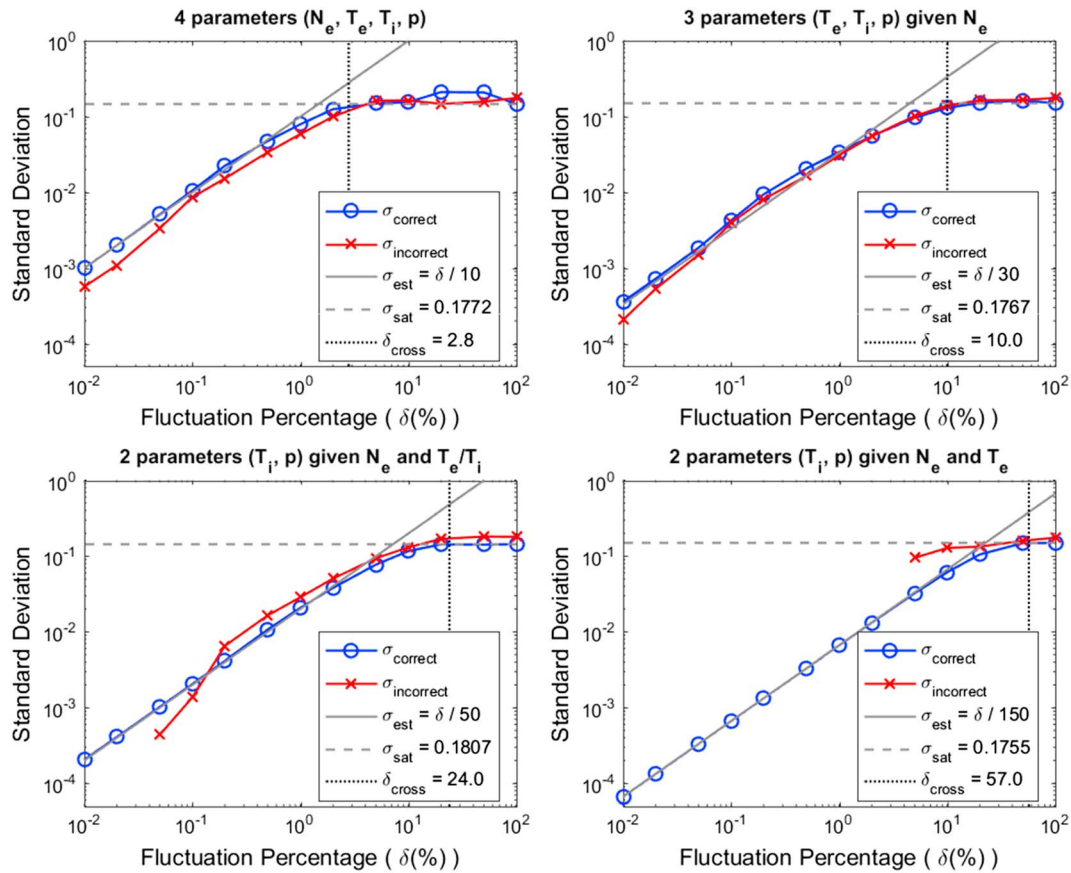
Figure 6. Average values of standard deviation (in logarithmical scale) of "correct" (blue) and "incorrect" (red) statistical distributions obtained by simulations of the estimation of different combinations of known a priori plasma parameters from the plasma line at different signal fluctuation percentages ($\delta[\%]$). 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}).

In the other study cases (a, b, and c), unambiguous estimates were only possible at almost noiseless scenarios ($\delta = 0.01\%$, red colored). At middle signal fluctuations ($0.05\% \leq \delta \leq 1\%$, colors cyan to orange) "incorrect" ion composition values were obtained (i.e., $\hat{p}_{\text{incorrect}} \approx 1 - p_{\text{true}}$). In high fluctuation cases ($\delta \geq 5\%$, blue colored) estimated results were 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 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 $\hat{p}_{\text{incorrect}} = 1 - p_{\text{true}}$. 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). Nevertheless, 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 at these points. This increase of uncertainty near the intersection point was previously indicated in Aponte et al. (2007).

4.2.2. Uncertainty of the Ion Composition Estimate

Figure 6 shows the standard deviation of both "correct" and "incorrect" distributions of the GMM PDFs calculated by the EM algorithm (σ_{correct} and $\sigma_{\text{incorrect}}$, respectively) at different noise levels. These values resemble the average uncertainty of the "correct" and "incorrect" estimated ion compositions. At the low and middle noise regimes (with signal fluctuations smaller than $\delta < 5\%$), both standard deviations increased linearly with the signal fluctuation percentage. Estimated linear regressions of standard deviation (σ_{est})
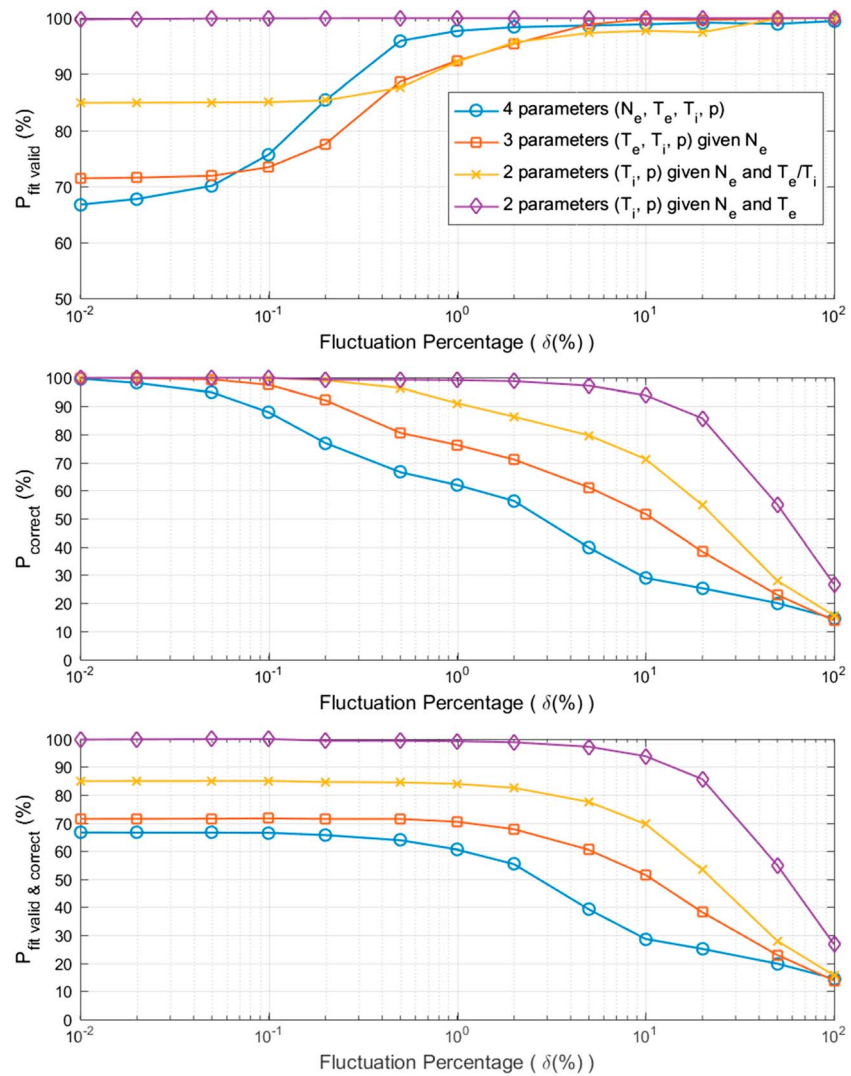
Figure 7. Probability of convergence ($P_{\text{fit valid}}$) and probability of "correct" estimation (P_{correct}; in percentage) at different signal fluctuation percentages ($\delta[\%]$) obtained by simulations of the estimation of different combinations of known a priori plasma parameters from the plasma line: without a priori information (blue circles), given N_e (orange squares), given N_e and T_e/T_i (yellow crosses), and given N_e and T_e (purple rhombus).

calculated for study cases a, b, c, and d were $\delta/10$, $\delta/30$, $\delta/50$, and $\delta/150$, respectively (Figure 6). These results indicate that the uncertainty 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 $\sigma_{\text{sat}} = 0.18$. This saturation value is the maximum standard deviation value that was obtained by the EM algorithm in the ion composition error range ($-1 \leq \mathcal{E}_p \leq 1$). These saturation values were reached at signal fluctuation levels approximately equal to δ_{cross}, where δ_{cross} is the signal fluctuation level at which $P_{\text{correct}} = 50\%$ in Figure 7. Consequently, δ_{cross} represents the maximum signal fluctuation threshold required to obtain a higher number of "correct" and "incorrect" estimates. Supporting information Figures S11, S16, S21, and S26 show the histograms of estimated ion composition of study cases a, b, c, and d. These figures verify that at noise levels higher than δ_{cross} the number of "incorrect" solutions is larger than the number of "correct" solutions.

4.2.3. Probability Results of Plasma Line Information Addition

The probability of convergence of the optimization algorithm, the probability of having a "correct" estimation, and their joint probability, are shown in Figure 7 for different combinations of known a priori

parameters (study cases a, b, c, and d). Each of the different study cases shown in this figure gradually improved P_{correct}, displacing the signal fluctuation threshold to higher values. This indicates that the a priori knowledge of different parameters provides different amounts of information to solve the TICA problem. The signal fluctuation thresholds for obtaining a $P_{\text{correct}} \geq 95.45\%$ were found at $\delta_{\text{th(no info)}} = 0.045\%$, $\delta_{\text{th}(Ne)} = 0.137\%$, $\delta_{\text{th}(Ne\ and\ Te/Ti)} = 0.568\%$, and $\delta_{\text{th}(Ne\ and\ Te)} = 7.93\%$, 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 results 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 $P_{\text{fit valid \& correct}}$.

The addition of a priori information improved the convergence of the optimization algorithm $P_{\text{fit valid}}$. The addition of N_e and T_e information (study case d) obtained a probability of valid convergence $P_{\text{fit valid}} > 99\%$ for all studied noise fluctuations. In this case, the optimization algorithm was always able to find a minimum of the cost function independently of the noise level. Other study cases (a, b, and c) obtained a reduction of $P_{\text{fit valid}}$ for signal fluctuations smaller than $\delta < 0.5\%$. 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 section 4.1.2. In small signal fluctuation cases, "incorrect" estimates were not convergent $(\chi_{r,\text{incorrect}}^2 > \chi_{r,\text{max}}^2)$ and only "correct" solutions were obtained, as shown in the histograms of χ_r^2 and the error of the ion composition estimate shown in supporting information Figures S13, S14, S18, S19, S23, S24, S28, and S29.

The number of computing iterations required was gradually reduced by the addition of a priori information at the different study cases. Supporting information Figure S6 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 use of a large initial parameter uncertainty (i.e., $\beta = 100\%$ in Figure 4) is equivalent to having initial parameters uniformly selected from the entire search range (case a in Figure 7, blue color line).

4.2.4. Relationships Between Estimation Errors of Parameters

Histograms of estimation error of plasma parameters are shown in supporting information Figures S15, S20, S25, and S30 for the different study cases previously indicated. In the study cases a, b, and c, histograms show a nonlinear dependence between the estimation errors of N_e, T_e, and T_i parameters and the estimation error of the ion composition. 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 (knowing a priori N_e and T_e) showed no relationships between the estimation error of T_i and $\mathcal{E}_p$, indicating that there was no ambiguity in this case.

Parameter estimation error percentages $\left(\mathcal{E}_X(\%) = 100 \cdot \left(X_{\text{true}} - \widehat{X}\right)/X_{\text{true}}\right)$ obtained from our Monte Carlo simulations were compared to the theoretical errors calculated with the analytic formulation of Waldteufel (1971). Supporting information Figure S14 shows the theoretical estimation errors obtained by the direct application of equation 10 of Zettergren et al. (2011). Both errors were found similar, indicating that the approximation of Waldteufel (1971) provides accurate error predictions. Nevertheless, electron temperature and electron density errors were much more disperse than those predicted by the analytical formulation.

The values shown in supporting information Figure S14 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 estimation errors obtained larger ion temperature errors. Estimated errors of T_e and T_i parameters 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 $\widehat{T}_{\text{incorrect}}$ $\left(\mathcal{E}_p = -1\right) = 2T_{\text{correct}}$ and $\widehat{T}_{\text{incorrect}}\left(\mathcal{E}_p = 1\right) = T_{\text{correct}}/2$.

4.3. Determination of the Most Relevant A Priori Parameter

4.3.1. Simulations of Single Parameter Addition

To determine which a priori parameter would be the most relevant to solve the ambiguity problem, the amount of information provided by each plasma parameter must be identified. Previous studies (Vallinkoski, 1988; 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 was found between T_i and p parameters, the addition of T_i information completely solved the ambiguity. In previous studies (Vallinkoski, 1988; Vallinkoski & Lehtinen, 1990), the unambiguous estimation was found to be dependent on the signal fluctuation level.

In our study, Monte Carlo simulations of 2,000 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) five unknown parameters (N_e, T_i, T_e, V_i, and p) without a priori information; (ii) four unknown parameters (N_e, T_i, T_e, and p) given a priori V_i; (iii) four unknown parameters (T_i, T_e, V_i, and p) given a priori N_e; (iv) four unknown parameters (N_e, T_i, V_i, and p) given a priori T_e/T_i; (v) four unknown parameters (N_e, T_i, V_i, and p) given a priori T_e; and (vi) four 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 already been analyzed in section 4.2 as study case a. Furthermore, study case iii (assuming N_e is known a priori) is similar to the study case b from section 4.2, but without assuming the V_i parameter is known. It is relevant to review study case iii separately because solutions may depend on the number of parameters.

4.3.2. Probability Results of Single Parameter Addition

Figure 8 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 similar probabilities. This result implies that 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 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 section 4.2 (knowing a priori N_e and V_i, shown in Figure 7) had also similar probabilities. This latter result verifies that parameter V_i does not provide information to solve the ambiguity.

Results of simulations of study case iii (with the a priori knowledge of N_e) had 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 had similar signal fluctuation thresholds to obtain $P_{\text{correct}} \geq 95.45\%$ ($\delta_{\text{th }(N_e)} = 0.14\%$ and $\delta_{\text{th }(T_e/T_i)} = 0.11\%$ for study cases iii and iv, respectively). Furthermore, $P_{\text{fit valid}}$ 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. Alternatively, P_{correct} values of study case c (knowing a priori N_e and T_e/T_i, shown in Figure 7) were much higher than the values obtained 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.

The 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 $\delta_{\text{th}(T_e)} = 1.54\%$ and $\delta_{\text{th}(T_i)} = 3.06\%$. As these signal fluctuations are commonly obtained at 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 $P_{\text{fit valid}(T_e)} > 96\%$ and $P_{\text{fit valid}(T_i)} > 98\%$. Consequently, the a priori knowledge of T_i provided the most relevant information for solving the TICA problem, as indicated by Vallinkoski
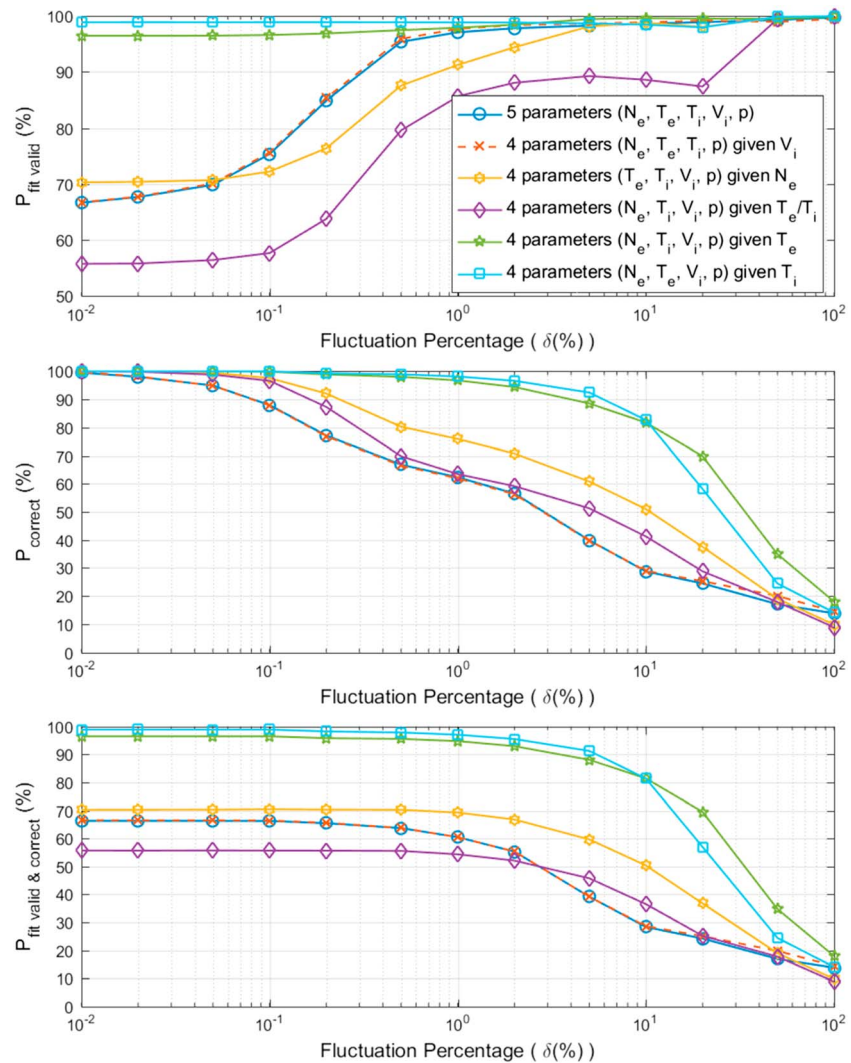
Figure 8. Probability of convergence ($P_{\text{fit valid}}$) and probability of "correct" estimation (P_{correct}; in percentage) at different signal fluctuation percentages ($\delta[\%]$) obtained from the analysis of plasma parameters without a priori information (blue circles), given V_i (orange 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).

(1988) and Vallinkoski and Lehtinen (1990). Nevertheless, the a priori knowledge of T_e obtained similar probabilities to the results of knowing a priori T_i.

Supporting information Figures S9 and S10 show the estimated ion composition values and the standard deviation values obtained for these study cases, respectively. Histograms of plasma parameter estimation error and their relationships with the ion composition estimation error are shown in supporting information Figures S35, S40, S45, S50, S55, and S60. 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).

4.4. Uncertainty of the A Priori Information From the Plasma Line
4.4.1. Uncertainty of Plasma Line 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 and Lehtinen (1990) analyzed the effect of the level of uncertainty of the a priori given parameters. Results from this study demonstrated the
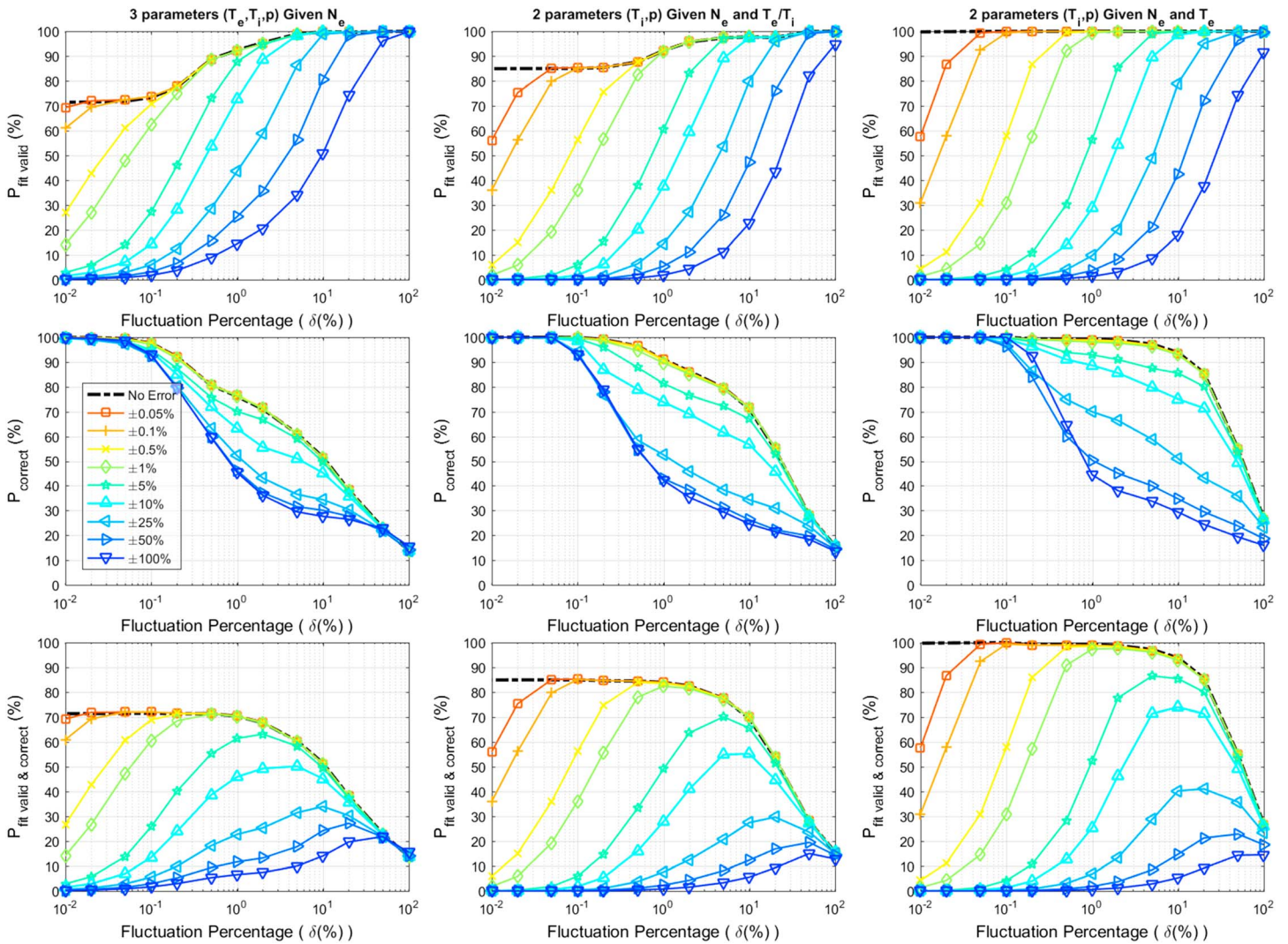
Figure 9. Probabilities obtained from the analysis of plasma parameters with a priori information having different levels of uncertainty. Columns correspond to simulations with different combinations of known a priori plasma parameters from the plasma line: given N_e (left), given N_e and T_e/T_i (middle), and given N_e and T_e (right). Black dotted lines represent simulation results with plasma parameters known a priori without uncertainty.

dependence of the estimation error on the a priori accuracy. Uncertainty levels analyzed in our study were $\epsilon = \pm0.05\%$, $\pm0.1\%$, $\pm0.5\%$, $\pm1\%$, $\pm5\%$, $\pm10\%$, $\pm25\%$, $\pm50\%$, and $\pm100\%$. 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).

Monte Carlo simulations of 1,000 different true input parameters were done to analyze the impact of adding plasma line information with uncertainty. The different combinations of parameters studied were: (1) three unknown parameters (T_i, T_e, and p) given a priori N_e; (2) two unknown parameters (T_i and p) given a priori N_e and T_e/T_i; and (3) two unknown parameters (T_i and p) given a priori N_e and T_e. In these cases, the ion drift velocity has been assumed to be known and independent of other parameter estimates, as shown in section 4.3.

The uncertainty error has been simulated as a random variation of the a priori known parameters. This variability was simulated as a uniform selection in the range of parameters defined by [$x_{\text{true}}(1-|\epsilon|/100)$, $x_{\text{true}}(1+|\epsilon|/100)$], where ϵ is a priori parameter uncertainty percentage and x_{true} is the a priori known true input parameter. Each of the 500 estimation repetitions of the true input parameters was done with a

different random noise 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.

4.4.2. Probability Results of Plasma Line Additions With Uncertainty

Figure 9 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 7) are shown in black dots for visual comparison. 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 $P_{\text{fit valid}}$ values were similar to the results obtained without uncertainty at fluctuation percentage values larger than the uncertainty level ($\delta \geq |\epsilon|$), but at lower noise levels the value of $P_{\text{fit valid}}$ rapidly decayed to zero when reducing the signal fluctuation value. It is assumed that simulations with $\delta \ll |\epsilon|$ did not converge ($P_{\text{fit valid}} = 0$) because the a priori given parameters 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. Therefore, to avoid wrongly estimated results when providing a priori information, verifying the convergence of the solutions by using the method indicated in section 3.1 is strongly suggested. Furthermore, verifying that the uncertainty of the known a priori measurements is higher than or equal to 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. 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 noise levels $\delta \leq 0.05\%$, solutions were "correct" ($P_{\text{correct}} \approx 100\%$) independently of the uncertainty and type of information provided. This latter result implies the existence of a global minimum threshold ($\delta_{\text{th}} = 0.05\%$) to solve the TICA problem, independently of the type of a priori information provided from the plasma line.

4.4.3. Use of In Situ Sensors

Simulations with known a priori information of N_e and T_e parameters (shown in Figures 7 and 9) 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, 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 (Figure 9). Alternatively, the 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 nanosatellite 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.

5. Conclusions

We have 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 NLLS optimization algorithm: the L-M algorithm. The convergence of the estimated solution was determined using the statistical distribution of the χ_r^2 cost function. Results were clustered using the EM algorithm to determine the "correctness" of the result depending on the distance of the ion composition estimation error to the "correct" and "incorrect" statistical probability distribution functions. The probability of convergence of the optimization algorithm ($P_{\text{fit valid}}$) 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. The estimation was conducted at different signal fluctuation levels to determine the

maximum noise threshold required to avoid the TICA problem. The δ_{th} thresholds 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 were performed on the initial parameters (β). 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 and of "correct" estimation 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; 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 ($\delta \ll |\varepsilon|$). The value of $P_{correct}$ was found to be dependent on the uncertainty level. For uncertainties $|\varepsilon| \leq 1\%$, $P_{correct}$ 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. Finally, 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.

Different methods have been described to solve the TICA problem in the ISR literature (Aponte et al., 2007; Bjørnå & Kirkwood, 1988; Cabrit & Kofman, 1996; Evans & Oliver, 1972; Lathuillere et al., 1983; Litvine et al., 1998; Nicolls et al., 2006; Oliver, 1979; Waldteufel, 1971; Wu et al., 2015) but exists a deficiency of published documentation about current practices at ISR facilities. A review and standardization of current ISR estimation methods would provide a significant benefit to the ISR community.

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 cost function to normalize the estimation results (shown in section 2.4.2), determine a maximum cost function value (shown in section 3.1 and supporting information Figure S3), 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*, it is suggested to execute several fittings with different initial parameters uniformly selected from the range of parameters and select the most frequently obtained solution (shown in section 2.6).
3. *To increase the probability of unambiguous estimation when adding a priori information*, it is suggested to determine the value of $P_{correct}$ in Figure 9 depending on the fluctuation level of the radar signal and the uncertainty of the a priori known parameters (shown in section 4.4). When the convergence of the solution is not verified, the probability of unambiguous estimation should be approximated by $P_{fit\ valid\ \&\ correct}$. Furthermore, it is recommended to select $\delta \geq |\varepsilon|$ to increase the probability of convergence of the estimation.

References

Akbari, H., Bhatt, A., La Hoz, C., & Semeter, J. L. (2017). Incoherent scatter plasma lines: Observations and applications. *Space Science Reviews*, *212*(1-2), 249–294. https://doi.org/10.1007/s11214-017-0355-7

Andrae, R., Schulze-Hartung, T., & 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., & González, S. A. (2007). Molecular ion composition measurements in the F1 region at Arecibo. *Journal of Geophysical Research*, *112*, A06322. https://doi.org/10.1029/2006JA012028

Aster, R. C., Borchers, B., & Thurber, C. H. (2013). *Parameter estimation and inverse problems* (2nd ed.). Waltham, MA: Academic Press, ISBN: 9780123850485.

Bevington, P. R., & Robinson, D. K. (2003). *Data reduction and error analysis for the physical sciences* (3rd ed.). New York, NY: McGraw-Hill, ISBN: 0-07-247227-8.

Beynon, W. J. G., & Williams, P. J. S. (1978). Incoherent scatter of radio waves from the ionosphere. *Reports on Progress in Physics*, *41*(6), 909–955. https://doi.org/10.1088/0034-4885/41/6/003

Bilitza, D., Altadill, D., Truhlik, V., Shubin, V., Galkin, I., Reinisch, B., & Huang, X. (2017). International Reference Ionosphere 2016: From ionospheric climate to real-time weather predictions. *Space Weather*, *15*, 418–429. https://doi.org/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*, *94*(A4), 3799–3804. https://doi.org/10.1029/JA094iA04p03799

Bjørnå, N., & Kirkwood, S. (1988). Derivation of ion composition from a combined ion line/plasma line incoherent scatter experiment. *Journal of Geophysical Research*, *93*(A6), 5787–5793. https://doi.org/10.1029/JA093iA06p05787

Blelly, P.-L., Alcaydé, D., & 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*, *115*, A09322. https://doi.org/10.1029/2009JA014876

Cabrit, B., & Kofman, W. (1996). Ionospheric composition measurement by EISCAT using a global fit procedure. *Annales Geophysicae*, *14*(12), 1496–1505. https://doi.org/10.1007/s00585-996-1496-2

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, *39*(1), 1–22. https://doi.org/10.1111/j.2517-6161.1977.tb01600.x

Djuth, F. T., Sulzer, M. P., & 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. https://doi.org/10.1029/94GL01699

Erickson, P. (1998). *Observations of light ions in the midlatitude and equatorial topside ionosphere*. PhD thesis retrieved from author. Ithaca, NY: Cornell University.

Evans, J. V. (1969). Theory and practice of ionosphere study by Thomson scatter radar. 968. *Proceedings of the IEEE*, *57*(4), 496–530. https://doi.org/10.1109/PROC.1969.7005

Evans, J. V., & Oliver, W. L. (1972). The study of E-region ion concentration and composition by incoherent scatter radar. *Radio Science*, *7*(1), 103–112. https://doi.org/10.1029/RS007i001p00103

Farley, D. T. (1969). Incoherent scatter correlation function measurements. *Radio Science*, *4*(10), 935–953. https://doi.org/10.1029/RS004i010p00935

Gavin, H. (2019). The Levenberg Marquardt method for nonlinear least squares curve fitting problem. Retrieved from: http://people.duke.edu/~hpgavin/lm.pdf

Hoffman, J. H., Johnson, C. Y., Holmes, J. C., & Young, J. M. (1969). Daytime midlatitude ion composition measurements. *Journal of Geophysical Research*, *74*(26), 6281–6290. https://doi.org/10.1029/JA074i026p06281

Holt, J. M., Rhoda, D. A., Tetenbaum, D., & Van Eyken, A. P. (1992). Optimal analysis of incoherent scatter radar data. *Radio Science*, *27*(3), 435–447. https://doi.org/10.1029/91RS02922

Huuskonen, A., & Lehtinen, M. S. (1996). The accuracy of incoherent scatter measurements: Error estimates valid for high signal levels. *Journal of Atmospheric and Terrestrial Physics*, *58*(1–4), 453–463. https://doi.org/10.1016/0021-9169(95)00048-8

Hysell, D. L., Chau, J. L., & Huba, J. D. (2009). Topside measurements at Jicamarca during solar minimum. *Annales Geophysicae*, *27*(1), 427–439. https://doi.org/10.5194/angeo-27-427-2009

Hysell, D. L., Rodrigues, F. S., Chau, J. L., & Huba, J. D. (2008). Full profile incoherent scatter analysis at Jicamarca. *Annales Geophysicae*, *26*(1), 59–75. https://doi.org/10.5194/angeo-26-59-2008

Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: A review. *ACM computing surveys (CSUR)*, *31*(3), 264–323. https://doi.org/10.1145/331499.331504

Johnson, S. G. (2012). The Faddeeva package. Open source code distributed under the MIT License, Retrieved from http://ab-initio.mit.edu/Faddeeva

Kay, S. M. (1993). *Fundamentals of statistical signal processing: Estimation theory*. Upper Saddle River, NJ: Prentice Hall, ISBN: 0133457117.

Kelley, M. C., & Heelis, R. A. (1989). *The Earth's ionosphere: Plasma physics and electrodynamics*. New York, NY: Academic Press, num. 43 international geophysics series. ISBN: 0124040128.

Kelly, J. D., & Wickwar, V. B. (1981). Radar measurements of high-latitude ion composition between 140 and 300 km altitude. *Journal of Geophysical Research*, *86*(A9), 7617–7626. https://doi.org/10.1029/JA086iA09p07617

Kennedy, J., & Eberhart, R. C. (1995). Particle swarm optimization. In *Proceedings of IEEE International Conference on Neural Networks IV* (pp. 1942–1948). Piscataway, NJ: IEEE Service Center.

Kudeki, E., & 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. https://doi.org/10.1109/TGRS.2010.2057252

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

Lathuillere, C., Wickwar, V. B., & Kofman, W. (1983). Incoherent scatter measurements of ion-neutral collision frequencies and temperatures in the lower thermosphere of the auroral region. *Journal of Geophysical Research*, *88*(A12), 10,137–10,144. https://doi.org/10.1029/JA088iA12p10137

Lehtinen, M. S., & Huuskonen, A. (1996). General incoherent scatter analysis and GUISDAP. *Journal of Atmospheric and Terrestrial Physics*, *58*(1-4), 435–452. https://doi.org/10.1016/0021-9169(95)00047-X

Lehtinen, M. S., Huuskonen, A., & Pirttilä, J. (1996). First experiences of full-profile analysis with GUISDAP. *Annales Geophysicae*, *14*(12), 1487–1495. https://doi.org/10.1007/s00585-996-1487-3

Levenberg, K. (1944). A method for the solution of certain non-linear problems in least squares. *Quarterly of Applied Mathematics*, *2*(2), 164–168. https://doi.org/10.1090/qam/10666

Litvine, A., Kofman, W., & Cabrit, B. (1998). Ion composition measurements and modelling at altitudes from 140 to 350 km using EISCAT measurements. *Annales Geophysicae*, *16*(10), 1159–1168. https://doi.org/10.1007/s00585-998-1159-6

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

Lu, Z., Yao, M., & Deng, X. (2016). An effective method for incoherent scattering radar's detecting ability evaluation. *Radio Science*, *51*, 852–857. https://doi.org/10.1002/2015RS005827

Marquardt, D. (1963). An algorithm for least-squares estimation of nonlinear parameters. *SIAM Journal on Applied Mathematics*, *11*(2), 431–441. https://doi.org/10.1137/0111030

Milla, M., & 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. https://doi.org/10.1109/TGRS.2010.2057253

Milla, M., Kudeki, E., Reyes, P., & Chau, J. (2013). A multi-beam incoherent scatter radar technique for the estimation of ionospheric electron density and Te/Ti profiles at Jicamarca. *Journal of Atmospheric and Solar - Terrestrial Physics*, *105-106*, 214–229. https://doi.org/10.1016/j.jastp.2013.06.003

Moon, T. K. (1996). The expectation-maximization algorithm. *IEEE Signal Processing Magazine*, *13*(6), 47–60. https://doi.org/10.1109/79.543975

Nicolls, M. J., Sulzer, M. P., Aponte, N., Seal, R., Nikoukar, R., & González, S. A. (2006). High-resolution electron temperature measurements using the plasma line asymmetry. *Geophysical Research Letters*, *33*, L18107. https://doi.org/10.1029/2006GL027222

Nikoukar, R., Kamalabadi, F., Kudeki, E., & Sulzer, M. (2008). An efficient near-optimal approach to incoherent scatter radar parameter estimation. *Radio Science*, *43*, RS5007. https://doi.org/10.1029/2007RS003724

Oliver, W. L. (1979). Incoherent scatter radar studies of the daytime middle thermosphere. *Annales de Geophysique*, *35*, 121–139.

Perrone, L., & Mikhailov, A. V. (2017). Long-term variations of exospheric temperature inferred from foF1 observations: A comparison to ISR Ti trend estimates. *Journal of Geophysical Research: Space Physics*, *122*, 8883–8892. https://doi.org/10.1002/2017JA024193

Perrone, L., & 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*, 8895–8907. https://doi.org/10.1029/2017JA025039

Picone, J. M., Hedin, A. E., Drob, D. P., & Aikin, A. C. (2002). NRLMSISE-00 empirical model of the atmosphere: Statistical comparisons and scientific issues. *Journal of Geophysical Research*, *107*(A12), 1468. https://doi.org/10.1029/2002JA009430

Scherliess, L., Fejer, B. G., Holt, J., Goncharenko, L., Amory-Mazaudier, C., & Buonsanto, M. J. (2001). Radar studies of midlatitude ionospheric plasma drifts. *Journal of Geophysical Research*, *106*(A2), 1771–1783. https://doi.org/10.1029/2000JA000229

Sharma, D. K., Aggarwal, M., & 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. https://doi.org/10.1016/j.asr.2016.11.027

Sulzer, M. P. (1986a). A phase modulation technique for a sevenfold statistical improvement in incoherent scatter data-taking. *Radio Science*, *21*(4), 737–744. https://doi.org/10.1029/RS021i004p00737

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. https://doi.org/10.1029/RS021i006p01033

Swoboda, J., Semeter, J., Zettergren, M., & Erickson, P. J. (2017). Observability of ionospheric space-time structure with ISR: A simulation study. *Radio Science*, *52*, 215–234. https://doi.org/10.1002/2016RS006182

Taylor, J. R. (1997). *An introduction to error analysis: The study of uncertainties in physical measurements* (2nd ed.). Sausalito, University Science Books. ISBN: 0–935702–42-3.

Tsui, P. C., & 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

Vallinkoski, M. (1988). Statistics of incoherent scatter multiparameter fits. *Journal of Atmospheric and Terrestrial Physics*, *50*(9), 839–851. https://doi.org/10.1016/0021-9169(88)90106-7

Vallinkoski, M., & 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*(6–8), 675–685. https://doi.org/10.1016/0021-9169(90)90061-Q

Waldteufel, P. (1971). Combined incoherent-scatter F1-region observations. *Journal of Geophysical Research*, *76*(28), 6995–6999. https://doi.org/10.1029/JA076i028p06995

Wand, R. H. (1970). Electron-to-ion temperature ratio from radar Thomson scatter observations. *Journal of Geophysical Research*, *75*(4), 829–838. https://doi.org/10.1029/JA075i004p00829

Wu, L.-L., Zhou, Q. H., Chen, T.-J., Liang, J. J., & Wu, X. (2015). Application of particle swarm optimization method to incoherent scatter radar measurement of ionosphere parameters. *Journal of Geophysical Research: Space Physics*, *120*, 8096–8110. https://doi.org/10.1002/2014JA020970

Yngvesson, K. O., & Perkins, F. W. (1968). Radar Thomson scatter studies of photoelectrons in the ionosphere and Landau damping. *Journal of Geophysical Research*, *73*(1), 97–110. https://doi.org/10.1029/JA073i001p00097

Zettergren, M., Semeter, J., Heinselman, C., & Diaz, M. (2011). Incoherent scatter radar estimation of F region ionospheric composition during frictional heating events. *Journal of Geophysical Research*, *116*, A01318. https://doi.org/10.1029/2010JA016035

Zhang, S.-R., Holt, J. M., Erickson, P. J., & 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. https://doi.org/10.1029/2017JA024948