Bayesian filtering in incoherent scatter plasma parameter fits

Ilkka I. Virtanen[1], Habtamu W. Tesfaw[1], Lassi Roininen[2], Sari Lasanen[3], and Anita Aikio[1]

[1]Space Physics and Astronomy Research Unit, University of Oulu, Oulu, Finland
[2]School of Engineering Science, Lappeenranta-Lahti University of Technology, Lappeenranta, Finland
[3]Sodankylä Geophysical Observatory, University of Oulu, Sodankylä, Finland

Key Points:

- Full-profile incoherent scatter analysis is implemented by means of Bayesian filtering and correlation priors.
- The technique reaches high time resolutions and enables ion composition fits.
- We have implemented the technique as an additional module to the GUISDAP incoherent scatter analysis tool.

Corresponding author: Ilkka I. Virtanen, ilkka.i.virtanen@oulu.fi

Abstract

Incoherent scatter (IS) radars are invaluable instruments for ionospheric physics, since they observe altitude profiles of electron density (N_e), electron temperature (T_e), ion temperature (T_i) and line-of-sight plasma velocity (V_i) from ground. However, the temperatures can be fitted to the observed IS spectra only when the ion composition is known, and resolutions of the fitted plasma parameters are often insufficient for auroral electron precipitation, which requires high resolutions in both range and time. The problem of unknown ion composition has been addressed by means of the full-profile analysis, which assumes that the plasma parameter profiles are smooth in altitude, or follow some predefined shape. In a similar manner, one could assume smooth time variations, but this option has not been used in IS analysis. We propose a plasma parameter fit technique based on Bayesian filtering, which we have implemented as an additional Bayesian Filtering Module (BAFIM) in the GUISDAP analysis package. BAFIM allows us to control gradients in both time and range directions for each plasma parameter separately. With BAFIM we can fit F_1 region ion composition together with N_e, T_e, T_i and V_i, and we have reached 4 s/900 m time/range steps in four-parameter fits of N_e, T_e, T_i and V_i in E region observations of auroral electron precipitation.

1 Introduction

Incoherent scatter (IS) radars are high-power, large-aperture radars that detect radio wave scattering from thermal fluctuations in the ionospheric plasma. Power spectral density of the scattered signal is a function of number density, temperature, bulk velocity, and ion-neutral collision frequency of a number of ion species and electrons (for example Swartz & Farley, 1979, and references therein). All these parameters cannot be fitted to the spectrum, and a commonly used approximation is the four-parameter fit of N_e, T_e, T_i and V_i. Equal temperatures and bulk velocities are assumed for all ion species, and the ion-neutral collision frequency and ion composition are taken from ionospheric models.

In the F_1 region the four-parameter fit often produces incorrect temperatures (for example Blelly & van Eyken, 2010), because ion composition models are unreliable in the transition region from the E region molecular NO^+ and O_2^+ ions to the F_2 region atomic O^+. Incorrect compositions bias the temperatures, because the IS spectrum is sensitive to the ratio T_i/m_i, where m_i is the mean ion mass. This is known as the "temperature-ion composition ambiguity" (TICA) (Martínez-Ledesma et al., 2019). Several authors have addressed the TICA problem by means of modeling the F_1 region temperature and ion composition profiles (Kelly & Wickwar, 1981; Cabrit & Kofman, 1996; Blelly & van Eyken, 2010; Zettergren et al., 2011; Häggström & Collis, 1990). Also direct estimation of both ion composition and temperature has been reported by Lathuillere, Lejeune, and Kofman (1983), but coarse resolutions were used, since such fits require extremely accurate IS spectra (Martínez-Ledesma et al., 2019).

Even the four-parameter fits are extremely challenging with a few second and a few hundred meter resolutions that are needed in observations of auroral electron precipitation. In high-resolution observations one may replace the fitted N_e with the raw electron density (scaled back-scattered power) N_r. For example Semeter and Kamalabadi (2005), Dahlgren et al. (2011), and Virtanen et al. (2018) used N_r in estimation of primary energy spectra of precipitating electrons. However, N_r equals N_e only if $T_e = T_i$, which may be an unjustified assumption when the precipitation heats the electron gas.

We propose an IS analysis technique that combines Bayesian filtering (for example Särkkä, 2013) in time and correlation priors (Roininen et al., 2011) in range. The combination allows us to extend the idea of full-profile IS analysis (Holt et al., 1992; Lehtinen et al., 1996; Hysell et al., 2008), which assumes smoothness in range, to an assumption of smoothness in both time and range. With this approach we can fit ion compo-

sitions if both ion temperature and composition are assumed to vary smoothly with time and altitude, and we can include temperature fits in high-resolution electron density fits.

In Section 2 we give introduction to IS plasma parameter fits, Bayesian filtering and correlation priors. In Section 3 we explain how the prior models and Bayesian filtering are used in IS analysis and implemented as a 'Bayesian Filtering Module' (BAFIM) in GUISDAP. In Section 4 we demonstrate BAFIM fits of N_e, T_e, T_i, V_i, and ion composition p=$[O^+]/N_e$ in the F_1 region, and high-resolution fits of N_e, T_e, T_i, and V_i in the E region.

2 Theoretical background

Incoherent scatter signal from a small plasma volume is a zero-mean random process with autocorrelation function $R(\tau)$, where τ is time lag. IS radar data are discrete samples of the autocorrelation function at discrete ranges r_i, times t_j, and lags τ_k. Power spectral density of the scattered signal, which is the Fourier transform of the autocorrelation function, is a known function of plasma parameters (for example Swartz & Farley, 1979, and references therein).

Typically, plasma parameters are extracted from the autocorrelation function samples by non-linear least-squares methods with optimization techniques such as Levenberg-Marquardt algorithm. Alternatively, Markov chain Monte Carlo methods can be used for parameter extraction (for example Virtanen et al., 2014), although optimization has remained as academic standard in IS analysis.

2.1 Gated analysis and full profile analysis

IS analysis techniques can be roughly divided into 'gated' and 'full-profile' techniques. In gated analysis one runs the fitting process for each range r_i and time t_j independently from the analysis of neighbouring observational volumes. The EISCAT IS analysis tool GUISDAP (Lehtinen & Huuskonen, 1996) makes gated analysis. In full-profile analysis one fits range profiles of plasma parameters. Main benefit of the full-profile analysis is the possibility to include prior information of plasma parameter altitude profiles.

In its most general form the full-profile analysis performs also deconvolution of lag profiles (Holt et al., 1992; Hysell et al., 2008). A simpler approach is to use phase-coding, for example alternating codes (Lehtinen & Häggström, 1987), and to decode the autocorrelation function samples into high resolution before the plasma parameter fit (Lehtinen et al., 1996). The two-stepped approach can be accomplished with arbitrary transmission modulations if the deconvolution is performed by statistical inversion (Virtanen et al., 2008, 2009). It is technically possible to add prior information already in the lag profile inversion step, but expressing the prior in terms of the actual plasma parameters is difficult in this approach.

2.2 Bayesian filtering and smoothing

Bayesian filtering (for example Särkkä, 2013) is a class of methods for estimating the state of a system from noisy indirect measurements. In IS analysis the state of the system reduces to point estimates of plasma parameter values and their standard deviations, while the indirect measurements are the observed autocorrelation function samples $\mathbf{R}$.

The filtering consists of a sequence of *prediction* and *update* steps. The sequence starts from an initial set of parameters $\mathbf{x}_1^-$ and its covariance matrix $\mathbf{P}_1^-$, which form our prior understanding of the unknown parameters at time t_1. Autocorrelation function sam-

ples $\mathbf{R}_1$ are then used to update the prior model into our best estimates of the parameters and their covariance at time t_1, $\mathbf{x}_1$ and $\mathbf{P}_1$. The update step is accomplished using a *measurement model* M,

$$\mathbf{x}_1 = M\left(\mathbf{x}_1^-, \mathbf{P}_1^-, \mathbf{R}_1\right). \tag{1}$$

The update step is followed by a prediction step, in which $\mathbf{x}_1$ and $\mathbf{P}_1$ are combined with our best understanding of dynamics of the system to create our best prediction of the parameters and their covariance at time t_2, $\mathbf{x}_2^-$ and $\mathbf{P}_2^-$. The prediction step is accomplished using a *dynamic model* D,

$$\mathbf{x}_2^- = D\left(\mathbf{x}_1, \mathbf{P}_1\right). \tag{2}$$

Measurements from time t_2 are then used to update the prediction into the final estimates $\mathbf{x}_2$ and $\mathbf{P}_2$, etc.

The simplest 'dynamic' model is to assume that the parameter values at subsequent time steps are close to each other, which reduces the prediction step into

$$\mathbf{x}_j^- = \mathbf{x}_{j-1}, \tag{3}$$
$$\mathbf{P}_j^- = \mathbf{P}_{j-1} + \mathbf{Q}, \tag{4}$$

where $\mathbf{Q}$ is the *system noise covariance matrix*. The larger values $\mathbf{Q}$ has in its diagonal, the smaller is the correlation between subsequent state estimates and the larger is the *filter gain*.

Bayesian filtering allows one to recursively estimate unknowns using the whole time history of measurements. In Bayesian smoothing the idea is extended to use of also 'future' measurements. Bayesian smoothing reduces variances of the unknown parameters and guarantees that equal amount of information from 'past' and 'future' measurements is included in each estimate of the unknowns. This removes a time shift that may be produced by a low-gain filter.

If the dynamic and measurement models are linear functions, Bayesian smoothing can be implemented as a recursive smoothing step called *Rauch-Tung-Striebel* (RTS) smoother (Rauch, 1963). The smoothing recursion runs backwards in time using equations

$$\mathbf{G}_j = \mathbf{P}_j \mathbf{D}_j^T \left(\mathbf{P}_{j+1}^-\right)^{-1}, \tag{5}$$
$$\mathbf{x}_j^s = \mathbf{x}_j + \mathbf{G}_j \left(\mathbf{x}_{j+1}^s - \mathbf{x}_{j+1}^-\right), \tag{6}$$
$$\mathbf{P}_j^s = \mathbf{P}_j + \mathbf{G}_j \left(\mathbf{P}_{j+1}^s - \mathbf{P}_{j+1}^-\right) \mathbf{G}_j^T, \tag{7}$$

where $\mathbf{D}_j$ is the theory matrix of the linear dynamic model D and the superscript T denotes matrix transpose. $\mathbf{x}_j^s$ and $\mathbf{P}_j^s$ form the Bayesian smoothing solution of the problem.

2.3 Correlation priors

Correlation priors (Roininen et al., 2011) allow one to model mutual covariances of the unknowns of an inverse problem in a well controlled way. Assuming that our prior belief of the unknowns $\mathbf{x}$ is $\mathbf{x}_p$, the prior can be expressed as a linear inverse problem

$$\mathbf{m}_p = \begin{pmatrix} \mathbf{x}_p \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix} = \mathbf{A}_p \mathbf{x} + \boldsymbol{\varepsilon}_p = \begin{pmatrix} \mathbf{A}_{p,0} \\ \mathbf{A}_{p,1} \\ \mathbf{A}_{p,2} \end{pmatrix} \mathbf{x} + \begin{pmatrix} \boldsymbol{\varepsilon}_{p,0} \\ \boldsymbol{\varepsilon}_{p,1} \\ \boldsymbol{\varepsilon}_{p,2} \end{pmatrix}, \tag{8}$$

where $\mathbf{x}_p$ are prior values of the unknown parameters and $\boldsymbol{\varepsilon}_p$ are discrete white noise with variances given in (18), (19), and (20). The theory matrix $\mathbf{A}_p$ is constructed from ze-

roth, first and second order differences $\mathbf{A}_{p,0}$, $\mathbf{A}_{p,1}$, and $\mathbf{A}_{p,2}$, as explained below. Covariance and mean of the prior can be solved from (8) as

$$\boldsymbol{\Sigma}'_p = (\boldsymbol{\Omega}_p)^{-1} = \left(\mathbf{A}_p^T \boldsymbol{\Sigma}_p^{-1} \mathbf{A}_p\right)^{-1}, \tag{9}$$

$$\mathbf{x}'_p = \boldsymbol{\Sigma}'_p \mathbf{A}_p^T \boldsymbol{\Omega}_p \mathbf{m}_p, \tag{10}$$

where $\boldsymbol{\Sigma}_p$ is the error covariance of $\boldsymbol{\varepsilon}_p$, $\boldsymbol{\Omega}_p$ is the precision matrix, $\mathbf{x}'_p$ is the final prior mean, and $\boldsymbol{\Sigma}'_p$ is its error covariance matrix. It is important to notice that the initial profile is smoothed by the correlations and $\mathbf{x}'_p \neq \mathbf{x}_p$. In high-dimensional problems it is important that $\boldsymbol{\Omega}_p$ is a sparse matrix (Norberg et al., 2018).

The zeroth order part of the prior is

$$\mathbf{A}_{p,0} = \mathbf{I}, \tag{11}$$

$$\boldsymbol{\Sigma}_{p,0} = \mathrm{diag}\left(\sigma^2_{p,0,1}, \sigma^2_{p,0,2}, \cdots, \sigma^2_{p,0,N}\right), \tag{12}$$

where the diagonal error covariance matrix $\boldsymbol{\Sigma}_{p,0}$ contains the prior variances of $\mathbf{x}_p$. The first order terms are

$$\mathbf{A}_{p,1} = \begin{pmatrix} 1 & -1 & 0 & \dots & 0 & 0 \\ 0 & 1 & -1 & \dots & 0 & 0 \\ 0 & 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & -1 \end{pmatrix}, \tag{13}$$

$$\boldsymbol{\Sigma}_{p,1} = \mathrm{diag}\left(\sigma^2_{p,1,1}, \sigma^2_{p,1,2}, \dots, \sigma^2_{p,1,N-1}\right), \tag{14}$$

and the second order terms are

$$\mathbf{A}_{p,2} = \begin{pmatrix} 1 & -2 & 1 & 0 & \dots & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 & \dots & 0 & 0 & 0 \\ 0 & 0 & 1 & -2 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 1 & -2 & 1 \end{pmatrix}, \tag{15}$$

$$\boldsymbol{\Sigma}_{p,2} = \mathrm{diag}\left(\sigma^2_{p,2,1}, \sigma^2_{p,2,2}, \dots, \sigma^2_{p,2,N-2}\right). \tag{16}$$

The full prior covariance matrix $\boldsymbol{\Sigma}_p$ is

$$\boldsymbol{\Sigma}_p = \begin{pmatrix} \boldsymbol{\Sigma}_{p,0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{p,1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\Sigma}_{p,2} \end{pmatrix}. \tag{17}$$

Variances of the zeroth, first and second order terms are (Roininen et al., 2011),

$$\sigma^2_{p,0,i} = c_0^{-1} \alpha_i \ell_i / \Delta h_i, \tag{18}$$

$$\sigma^2_{p,1,i} = c_1^{-1} \alpha_i \Delta h_i / \ell_i \quad = c_0/c_1 \sigma^2_{p,0,i} (\Delta h_i/\ell_i)^2, \tag{19}$$

$$\sigma^2_{p,2,i} = c_2^{-1} \alpha_i (\Delta h_i/\ell_i)^3 = c_0/c_2 \sigma^2_{p,0,i} (\Delta h_i/\ell_i)^4, \tag{20}$$

where α_i is *correlation power* in the $i^{(\mathrm{th})}$ range gate, Δh_i is width of the $i^{(\mathrm{th})}$ range gate, and ℓ_i is the corresponding *correlation length*. The correlation lengths define how smooth the profile is, and the correlation power defines width of the prior distribution. The constants c_0, c_1, c_2 define shape of the final covariance structure. For example, $c_0 = 1, c_1 = 1/2, c_2 = 1/8$ produces a Gaussian covariance. The model variances depend on the discretization and correlation length in a way that makes the model essentially grid-independent.

3 BAFIM implementation

We have implemented an IS analysis tool based on Bayesian filtering in time and correlation priors in range as an additional *Bayesian filtering module* (BAFIM) to the

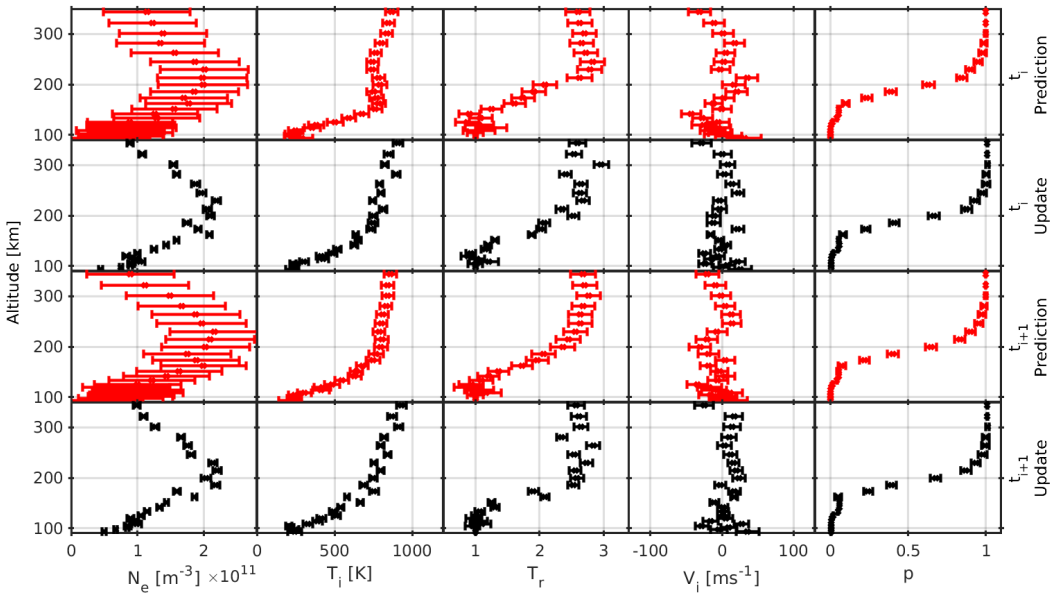

Figure 1. Prediction and update steps of BAFIM. Predicted altitude profiles of N_e, T_i, T_r, V_i, and p at time t_i (first row), updated profiles at t_i (second row), predicted profiles at t_{i+1} (third row), and updated profiles at t_{i+1}.

GUISDAP IS analysis tool (Lehtinen & Huuskonen, 1996). We assume a five parameter fit of electron number density N, ion temperature T, ion-to-electron temperature ratio E, line-of-sight plasma velocity V and ion composition $O = [O^+]/N$, where O^+ is the O^+ ion number density. We use the alternative notation (N=N_e, T=T_i, E=T_r, V=V_i, O=p) in this section to simplify the equations. The vector of plasma parameters at time step j is

$$\mathbf{x}_j = (\mathbf{N}_j, \mathbf{T}_j, \mathbf{E}_j, \mathbf{V}_j, \mathbf{O}_j)^T , \tag{21}$$

where $\mathbf{N}_j$ is the electron density profile in range gates $i = 1, \ldots, M$,

$$\mathbf{N}_j = (N_{1,j}, N_{2,j}, \ldots, N_{M,j}) , \tag{22}$$

and the vectors of the other parameters are defined similarly.

The analysis starts from an initial guess of the plasma parameters at time t_1, $\mathbf{x}_1^-$, and their covariance $\mathbf{P}_1^-$. The parameters $\mathbf{x}_1^-$ are from the International Reference Ionosphere (IRI) model (Bilitza et al., 2017), and $\mathbf{P}_1^-$ is a diagonal matrix with variances equal to the *process noise variances*, defined in (35), in its diagonal. The parameters $\mathbf{x}_1^-$ and their variances $\boldsymbol{\sigma}_1^{2-} = \mathrm{diag}(\mathbf{P}_1^-)$ are used as a prior in a normal GUISDAP fit to measurements $\mathbf{R}_1$. The GUISDAP fit is the update step of the Bayesian filter. The gated GUISDAP analysis does not produce a full error covariance matrix of $\mathbf{x}_1$, but the error covariance matrix $\mathbf{P}_1$ contains mutual correlations of plasma parameters in each range gate.

After the first time step, priors for the following GUISDAP fits are not taken from the IRI model, but the fit results from t_1 are used to predict the parameters and their covariance at t_2. The predicted values $\mathbf{x}_2^-$ and diagonal of $\mathbf{P}_2^-$ are used as prior mean and variance to fit $\mathbf{x}_2$ and $\mathbf{P}_2$ to measurements $\mathbf{R}_2$, $\mathbf{x}_2$ and $\mathbf{P}_2$ are used to calculate the predicted $\mathbf{x}_3^-$ and $\mathbf{P}_3^-$, etc. The analysis steps are illustrated in Figure 1, whose first row shows predicted altitude profiles of N=N_e, T=T_i, E=T_e, V=V_i, and O=p at time t_i. The predictions are used as priors in a GUISDAP fit, which produces the updated profiles

on the second row. The profiles on the second row are used to predict the parameter profiles at time t_{i+1} (third row), the prediction is used as a prior when fitting the parameters at time t_{i+1} (fourth row), etc. Correlations in range are lost and reintroduced in each update and prediction step, correspondingly. This allows us to use the computationally light-weight gated analysis, and the approach is acceptable if the plasma parameters do not change much during a time step.

In the prediction step, a correlation prior is used to create smooth plasma parameter profiles. The measurements $\mathbf{x}_1$ and their covariance $\mathbf{P}_1$ are used as the zeroth order terms in (8),

$$\mathbf{x}_p = \mathbf{x}_1, \tag{23}$$

$$\mathbf{\Sigma}_{p,0} = \mathbf{P}_1. \tag{24}$$

The first and second order differences in (8) are then formed for each plasma parameter separately. Variances of the plasma parameters $\mathbf{x}_1 = (\mathbf{N}_1, \mathbf{T}_1, \mathbf{E}_1, \mathbf{V}_1, \mathbf{O}_1)^T$ are

$$\boldsymbol{\sigma}^2_{p,0} = \left(\boldsymbol{\sigma}^2_{N,0}, \boldsymbol{\sigma}^2_{T,0}, \boldsymbol{\sigma}^2_{E,0}, \boldsymbol{\sigma}^2_{V,0}, \boldsymbol{\sigma}^2_{O,0}\right)^T = \mathrm{diag}\left(\mathbf{P}_1\right). \tag{25}$$

The first order difference matrices (13) for each parameter are identical $M \times M-1$ matrices, $\mathbf{A}_{N,1} = \mathbf{A}_{T,1} = \mathbf{A}_{E,1} = \mathbf{A}_{V,1} = \mathbf{A}_{O,1}$, and the full first order difference matrix is the block diagonal matrix

$$\mathbf{A}_{p,1} = \begin{pmatrix} \mathbf{A}_{N,1} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{T,1} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{A}_{E,1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{A}_{V,1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{A}_{O,1} \end{pmatrix}. \tag{26}$$

Variances of the first order terms are calculated from (19). First order variances for electron density are

$$\sigma^2_{N,1,i} = c_0/c_1 \sigma^2_{N,0,i} (\Delta h_i / \ell_{N,i})^2, \tag{27}$$

and variances of the other parameters are calculated in a similar manner. The first order covariance matrix is the diagonal matrix

$$\mathbf{\Sigma}_{p,1} = \mathrm{diag}\left(\boldsymbol{\sigma}^2_{N,1}, \boldsymbol{\sigma}^2_{T,1}, \boldsymbol{\sigma}^2_{E,1}, \boldsymbol{\sigma}^2_{V,1}, \boldsymbol{\sigma}^2_{O,1}\right). \tag{28}$$

The second order differences and their variances are formed in a similar manner. As a result, we have a matrix equation of the from (8), from which parameter profiles smoothed in range, $\mathbf{x}'$, and their covariance, $\mathbf{\Sigma}'$, can be solved using (9) and (10).

The smoothed parameter profiles $\mathbf{x}'$ are used as the prediction for time step t_2,

$$\mathbf{x}_2^- = \mathbf{x}', \tag{29}$$

and the predicted covariance is the sum of the covariance of $\mathbf{x}'$ and a process noise covariance $\mathbf{Q}$,

$$\mathbf{P}_2^- = \mathbf{\Sigma}' + \mathbf{Q}. \tag{30}$$

The process noise covariance is a diagonal $5M \times 5M$ matrix with a different variance for each plasma parameter (35) in its diagonal,

$$\mathbf{Q} = \mathrm{diag}\left(q_N, \ldots, q_N, q_T, \ldots, q_T, q_E, \ldots, q_E, q_V, \ldots, q_V, q_O, \ldots, q_O\right). \tag{31}$$

The RTS smoother is implemented in BAFIM as a post-processing step. Since only the first $5M$ elements of the vector $\mathbf{m}_p$ are nonzero in (8) and (10), the matrix $\mathbf{D}$ in (5) consists of the first $5M$ columns of the $5M \times (15M - 3)$ matrix

$$\mathbf{D}' = \mathbf{\Sigma}_s \mathbf{A}_p^T \mathbf{\Omega}_p. \tag{32}$$

The RTS smoother is only a linear approximation, but the approximation is reasonable if the time steps are short enough to keep changes in plasma parameters small in between subsequent time steps.

The correlation lengths ℓ_i are proportional to the plasma scale height

$$H_i = \frac{k_B T_i (1 + E_i)/2}{m_i g_i}, \tag{33}$$

calculated from the IRI model parameters. Here k_B is the Boltzmann constant, m_i is the mean ion mass, g_i is the acceleration of gravity, and the subscript i refers to the $i^{(\mathrm{th})}$ range gate. The correlation lengths of N are

$$\ell_{N,i} = s_N^h H_i, \tag{34}$$

where s_N^h is a constant, and the correlation lengths of the other parameters are defined in a similar manner.

In the correlation prior, covariance of the zeroth order terms is the posterior covariance $\mathbf{\Sigma}_{p,0} = \mathbf{P}_1$, and variances of the first and second order terms are proportional to ℓ_i^{-2} and ℓ_i^{-4}, respectively. Thus, at the limit of small correlation lengths ℓ_i, the smoothed profile $\mathbf{x}'$ approaches the fitted profile $\mathbf{x}_1$, and the covariance $\mathbf{\Sigma}'$ approaches $\mathbf{P}_1$. BAFIM can thus be run without the smoothing in range if the correlation lengths ℓ_i are small, i.e. the constants s^h are small.

The process noise variances q_N, q_T, q_E, q_V, q_O are proportional to the time step duration,

$$q_N = \left(s_N^t\right)^2 \Delta t, \tag{35}$$

etc. Each parameter is fitted within an altitude interval $[h_{min}, h_{max}]$, below h_{min} and above h_{max} the parameter is fixed to the IRI model value with a small variance. The heights $h_{min,N}$, $h_{max,N}$, $h_{min,T}$, $h_{max,T}$, ..., the constants s_N^h, s_T^h, s_E^h, s_V^h, s_O^h, and the constants s_N^t, s_T^t, s_E^t, s_V^t, s_O^t are user inputs and may vary from one analysis run to another.

4 Plasma parameter fits with BAFIM

In this section we demonstrate plasma parameter fits with BAFIM in two use cases, ion composition fits in the F_1 region and high-resolution E region analysis during auroral electron precipitation. We use field-aligned observations from the EISCAT Svalbard radar (ESR) and the EISCAT UHF radar. We consider fits of electron density (N_e), ion temperature (T_i), ion-to-electron temperature ratio (T_r), line-of-sight plasma bulk velocity (V_i), and ion composition ($p = [O^+]/N_e$). In the results we show the electron temperature $T_e = T_i \cdot T_r$ instead of T_r. While the assumption of smoothness in range is necessary in the selected demonstrations, we emphasize that BAFIM can be used also without this assumption, for example to improve time resolution of four-parameter fits in low-elevation or bistatic observations. In this section, standard GUISDAP fits and GUISDAP fits with BAFIM are referred to as 'GUISDAP' and 'BAFIM', correspondingly.

Both ESR and UHF data are from experiments that use alternating codes (Lehtinen & Häggström, 1987). The ESR 'ipy' experiment uses a 30-bit code sequence with 30 μs bit length and the data are decoded to 2.25 km resolution. The UHF 'arc1' experiment uses a 64-bit code sequence with 6 μs bit length and the data are decoded to 900 m resolution. In high signal-to-noise conditions GUISDAP may underestimate plasma parameter variances because it neglects correlations between autocorrelation function samples (Huuskonen & Lehtinen, 1996). Both experiments use randomized (Lehtinen et al., 1997) codes to reduce the correlations. If highly correlated data were analysed with BAFIM, smoothing in time and range would be reduced due to the underestimation of errors in the GUISDAP fits.

Table 1. BAFIM settings used in the data analysis. N_e, T_i, T_r, and V_i are fitted at all altitudes above h_{min}. p is not fitted at all in the E region analysis of the UHF data. The constants s^h and s^t are scaling factors that control smoothness of the solutions in range and time, respectively, as explained in Section 3.

	ESR 27 June 2007				UHF 12 December 2012			
	s^h	s^t	h_{min} (km)	h_{max} (km)	s^h	s^t	h_{min} (km)	h_{max} (km)
N_e	0.1	$2.5 \cdot 10^{10}$ m^{-3}s$^{-1/2}$	0	-	1.0	$2.5 \cdot 10^{11}$ m^{-3}s$^{-1/2}$	0	-
T_i	0.3	10 Ks$^{-1/2}$	80	-	0.2	30 Ks$^{-1/2}$	80	-
T_r	0.3	0.05 s$^{-1/2}$	103	-	0.4	0.1 s$^{-1/2}$	103	-
V_i	0.2	2.5 ms$^{-3/2}$	80	-	0.1	5 ms$^{-3/2}$	80	-
p	0.2	0.003 s$^{-1/2}$	150	320	-	-	0	0

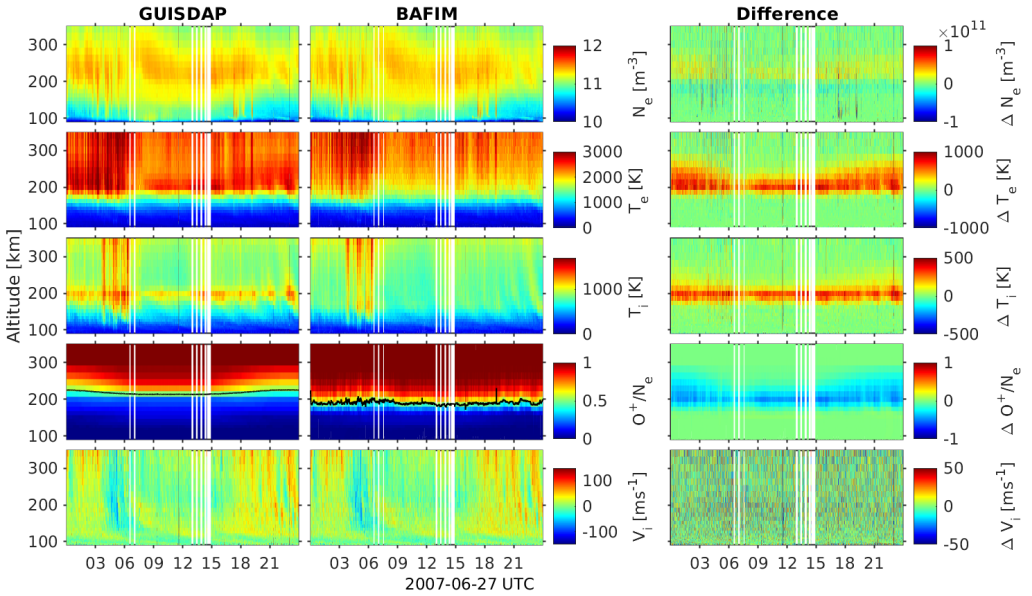

Figure 2. Ion composition analysis. GUISDAP four-parameter fit with 60 s resolution (left), five-parameter BAFIM fit with 6 s time steps (middle), and difference of these two (GUISDAP - BAFIM) (right). In the default GUISDAP fit the ion composition is from the IRI model.

4.1 Ion composition fits

Ion frictional heating occurs when an electric field drives the ionospheric plasma through the neutral atmosphere and the ion gas is heated in collisions with neutral particles. The heating may affect F_1 region ion composition, because reaction rates of some important charge-exchange reactions depend on temperature, and expansion of the neutral atmosphere may increase neutral N_2 concentration in the F region (Kelly & Wickwar, 1981). Deviations from the IRI ion composition may bias F_1 region ion temperature estimates in four-parameter GUISDAP fits of N_e, T_e, T_i and V_i. An example of such an event is shown on the left in Figure 2, where four-parameter GUISDAP fit results with 60 s resolution are shown for 24 hours of ESR data. Ion temperature (third panel on the left) has an artificial local maximum around 200 km altitude, where IRI predicts too much molecular ions (fourth panel).

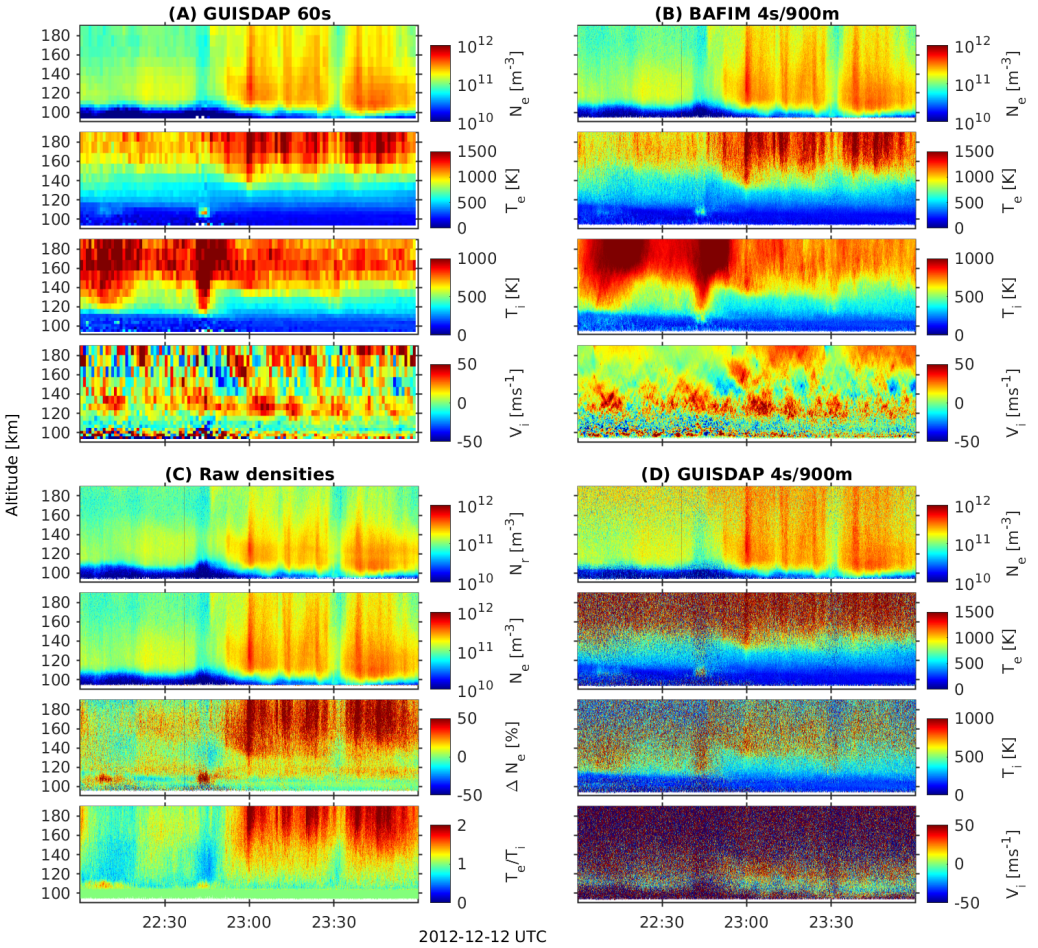

Figure 3. High-resolution E region analysis. (A) GUISDAP fit with 60 s time resolution and range resolution varying from 3 km to 13 km, (B) BAFIM fit with 4 s/900 m steps, (C) N_r, BAFIM-fitted N_e, relative difference $\Delta N_e = (N_e - N_r)/N_r$, and BAFIM T_e/T_i, (D) GUISDAP fit with 4 s/900 m resolutions.

In five-parameter BAFIM fit of the same data (Figure 2, middle panels), also the ion composition p is fitted, and the analysis proceeds with 6 s time steps. Other BAFIM settings are listed in Table 1. The artificial ion temperature maximum, which is visible in the GUISDAP fit, is not produced in the BAFIM fit. The transition altitude, where number density of molecular ions is equal to O^+ density ($p = 50\%$, black lines in the fourth panels), is generally lower than in the IRI model. Difference of the two fit results (GUISDAP - BAFIM) is shown on the right in Figure 2, where one can see how the difference in p affects also T_i, T_e and even N_e profiles. While the artefact around 200 km altitude was removed by BAFIM, the true ion frictional heating events between 4 and 7 UT, as well as the weaker T_i enhancements after 15 UT, are reproduced by BAFIM, demonstrating its ability to maintain true ion temperature maxima. We note that our results are very similar with those of Blelly and van Eyken (2010), who used the same data to demonstrate a full-profile analysis technique based on ion energy equations.

4.2 High-resolution observations of auroral electron precipitation

IS radars can detect impact ionization and electron heating caused by auroral electron precipitation. While existing high-latitude IS radars can typically reach a time resolution of some tens of seconds in the four-parameter fits of N_e, T_i, T_r, and V_i, optical observations show that the precipitation may change substantially in a few seconds and even below (for example Dahlgren et al., 2016). High-resolution E region observations often rely on raw electron densities (for example Semeter & Kamalabadi, 2005; Dahlgren et al., 2011; Virtanen et al., 2018), which are calculated assuming $T_e = T_i$. However, this assumption may not be justified, since the precipitation heats the electron gas.

Figure 3 shows plasma parameter fit results from three different analysis runs of an EISCAT UHF radar measurement on 12 December 2012: (A) a four-parameter GUISDAP fit with 60 s time resolution and range resolution varying from 3 km to 13 km, (B) a BAFIM fit with 4 s/900 m steps, and (D) a GUISDAP fit with 4 s/900 m resolution. BAFIM settings are listed in Table 1. While plasma parameters from the GUISDAP fit are extremely noisy with the 4 s/900 m resolution, the BAFIM fit produces temperatures and velocities that match well with the standard coarse-resolution fit (for example, compare T_i and V_i in panels (A), (B), and (D)).

Importance of the temperature fit is demonstrated in Figure 3 (C), which shows raw electron density N_r, BAFIM-fitted N_e, relative difference $(N_e - N_r)/N_r$, and the temperature ratio T_e/T_i. The raw densities are clear underestimates after 22:50 UT, when electron precipitation heats the electron gas and $T_e > T_i$. This effect was neglected for example in Virtanen et al. (2018), because the high-resolution four-parameter fits were practically impossible.

5 Discussion

BAFIM is the first implementation of Bayesian filtering to IS plasma parameter fits. In this section we discuss some important properties of BAFIM and potential future improvements.

5.1 Resolutions of BAFIM fit results

While the BAFIM analysis proceeds with short steps in range and time, each fit result may contain information from considerably longer intervals because the steps are correlated. One should thus make a clear difference between the effective resolutions – the intervals in range and time which make significant contributions to the fit result – and the step sizes. The effective resolutions are not constants, but they depend on the amount of information gained from each individual measurement. Effective resolutions of each parameter at each altitude may be tuned separately by adjusting the corresponding process noise variances and correlation lengths. Effective resolutions of N_e are typically kept very close to the step sizes, while coarser effective resolutions are accepted for the other parameters to reach an acceptable statistical accuracy.

5.2 Tuning and validating BAFIM

Tuning the process noise variances and correlation lengths of BAFIM may be nontrivial, since the correlations in time allow part of the prior information introduced with the correlation priors to be passed from one time step to another. Any change in process noise variance must thus be compensated with a corresponding change in correlation length to keep the effective smoothing in range unchanged. In addition, changing the process noise and correlation length of one plasma parameter may affect the others due to error correlations.

In this paper, BAFIM was tuned to produce practically uncorrelated electron densities, while correlation lengths and process noise variances of the other parameters were selected in such a way that noise level of the fitted parameters roughly matched with the default GUISDAP fits with 60 s resolution. The only physics-based part of the model are the correlation lengths, which are proportional to the plasma scale heights. Physics-based, automatic ways to tune the filter will be topics of future works. Alternative ways to tune the filter would be to derive theoretical limits for gradients in space and time, or to extract information on the correlation structures from existing measurements. Correlation structures of mesospheric winds have been extracted from meteor radar observations by Vierinen et al. (2019), and a similar work for incoherent scatter radars could be possible.

Validation of BAFIM results, the ion composition fits in particular, is a challenging task due to lack of measurements from other instruments. Observations of F_1 region ion composition are mainly from rockets, and the rocket observation would need to be from vicinity of the radar beam to enable reasonable comparisons. Alternatively, one could analyse simulated radar data corresponding a realistic model ionosphere. Such simulations would be possible for example with the simISR tool (Swoboda et al., 2017).

5.3 Ion composition fits

In the ion composition fits a small process noise variance q_O was used for the ion composition and a relatively large variance q_T was used for the ion temperature, which is equivalent with the assumption that ion temperature varies much more rapidly than ion composition. Only slow variations in composition were allowed, because allowing rapid variations in both ion composition and temperature may lead to unrealistic oscillations due to the temperature-ion composition ambiguity. With the selected tuning BAFIM can follow the relatively slow ion composition variations associated with the large scale convection electric field, but rapid variations caused, for example, by small scale electric fields around auroral arcs are challenging.

Time resolution of the composition fits could be improved if physics-based models were included in the prediction step. One could either model the temperature profiles or include a chemistry model that solves temperature-dependent compositions. The temperature profiles could be modeled, for example, with the techniques of Zettergren et al. (2011) and Blelly and van Eyken (2010), while chemistry modeling could be adopted for example from Richards and Voglozin (2011). Also D region ion composition and temperatures could be observed if a sufficient model, for example the Sodankylä Ion and Neutral Chemistry (SIC) model (Turunen et al., 2016) was used.

5.4 EISCAT_3D

EISCAT_3D (McCrea et al., 2015) is the next-generation geospace radar system currently being built in northern Norway, Sweden, and Finland. The radar will provide an order-of-magnitude improvement in measurement speed, and it will be the first multistatic, multibeam incoherent scatter radar system. EISCAT_3D will be able to conduct volumetric observations, including 3D observations of plasma flows.

If BAFIM-like analysis was applied to field-aligned EISCAT_3D measurements, the order-of-magnitude improvement would mean sub-second time steps in four-parameter fits, and resolutions sufficient for rapidly varying conditions in association with aurora in ion composition fits. The volumetric observations would allow one to implement 3D models of the ionosphere in the prediction step. An EISCAT_3D analysis tool could be designed for the volumetric observations and could make optimal use of the multistatic, multibeam data, following the idea of Virtanen et al. (2014).

6 Conclusions

We have introduced an incoherent scatter analysis technique that allows us to control plasma parameter gradients in both time and space using Bayesian filtering and correlation priors. The technique is implemented as a Bayesian filtering module (BAFIM) in the GUISDAP analysis package. BAFIM allows us to to fit F_1 region ion compositions and transition altitudes, and to include ion and electron temperatures in high resolution plasma parameter fits, in field-aligned incoherent scatter measurements. Improvements provided by the new analysis tool were demonstrated with EISCAT radar data, including fits of F_1 region ion composition and high-resolution E region plasma parameter fits during short-lived auroral precipitation events. The technique could be extended to volumetric, multistatic observations of the EISCAT_3D radar and supplemented with ion chemistry models.

Acknowledgments

This work is supported by the Academy of Finland, application numbers 285474, 326240, and 301542, and the Kvantum institute of the University of Oulu. EISCAT is an international association supported by research organisations in China (CRIRP), Finland (SA), Japan (NIPR and STEL), Norway (NFR), Sweden (VR), and the United Kingdom (NERC). The EISCAT data and the GUISDAP software are available for download from the EISCAT web page (http://www.eiscat.se). BAFIM is available at https://doi.org/10.5281/zenodo.4033904.

References

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*(2), 418-429. doi: 10.1002/2016SW001593

Blelly, P.-L., & van Eyken, A. P. (2010). A new analysis method for determining polar ionosphere and upper atmosphere characteristics from ESR data: Illustration with IPY period. *J. Geophys. Res.*, *115*, A09322. doi: 10.1029/2009JA014876

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

Dahlgren, H., Gustavsson, B., Lanchester, B. S., Ivchenko, N., Brändström, U., Whiter, D. K., ... Marklund, G. (2011). Energy and flux variations across thin auroral arcs. *Ann. Geophys.*, *29*, 1699-1712. doi: 10.5194/angeo-29-1699-2011

Dahlgren, H., Lanchester, B. S., Ivchenko, N., & Whiter, D. K. (2016). Electrodynamics and energy characteristics of aurora at high resolution by optical methods. *J. Geophys. Res. Space Physics*, *121*, 5966-5974. doi: 10.1002/2016JA022446

Häggström, I., & Collis, P. N. (1990). Ion composition changes during F-region density depletions in the presence of electric fields at auroral latitudes. *J. Atmos. Terr. Phys.*, *52*(6-8), 519-529. doi: 10.1016/0021-9169(90)90050-W

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

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

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

Kelly, J. D., & Wickwar, V. B. (1981). Radar measurements of high-latitude ion

composition between 140 and 300 km altitude. *J. Geophys. Res.*, *86*(A9), 7617-7626. doi: 10.1029/JA086iA09p07617

Lathuillere, C., Lejeune, G., & Kofman, W. (1983). Direct measurements of ion composition with EISCAT in high-latitude F_1 region. *Radio Sci.*, *18*(6), 887-893. doi: 10.1029/RS018i006p00887

Lehtinen, M. S., & Häggström, I. (1987). A new modulation principle for incoherent scatter measurements. *Radio Sci.*, *22*, 625-634. doi: 10.1029/RS022i004p00625

Lehtinen, M. S., & Huuskonen, A. (1996). General incoherent scatter analysis and GUISDAP. *J. Atmos. Terr. Phys.*, *58*, 435-452. doi: 10.1016/0021-9169(95)00047-X

Lehtinen, M. S., Huuskonen, A., & Markkanen, M. (1997). Randomization of alternating codes: Improving incoherent scatter measurements by reducing correlations of gated autocorrelation function estimates. *Radio Sci.*, *32*, 2271-2282. doi: 10.1029/97RS02556

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

Martínez-Ledesma, M., Quezada, D., & Andrés, M. (2019). Determination of the Signal Fluctuation Threshold of the Temperature-Ion Composition Ambiguity Problem Using Monte Carlo Simulations. *J. Geophys. Res. Space Physics*, *124*(4), 2897-2919. doi: 10.1029/2018JA026217

McCrea, I., Aikio, A., Alfonsi, L., Belova, E., Buchert, S., Clilverd, M., ... Vierinen, J. (2015). The science case for the EISCAT_3D radar. *Prog. Earth Planet. Sc.*, *2*(1). doi: 10.1186/s40645-015-0051-8

Norberg, J., Vierinen, J., Roininen, L., Orispää, M., Kauristie, K., Rideout, W. C., ... Lehtinen, M. S. (2018). Gaussian Markov Random Field Priors in Ionospheric 3-D Multi-Instrument Tomography. *IEEE Trans. Geosci. Remote Sens.*, 1-13. doi: 10.1109/TGRS.2018.2847026

Rauch, H. R. (1963). Solutions to the linear smoothing problem. *IEEE T. Automat. Contr.*, *8*(4), 371-372.

Richards, P. G., & Voglozin, D. (2011). Reexamination of ionospheric photochemistry. *J. Geophys. Res. Space Physics*, *116*(A8). doi: 10.1029/2011JA016613

Roininen, L., Lehtinen, M. S., Lasanen, S., Orispää, M., & Markkanen, M. (2011). Correlation Priors. *Inverse Problems and Imaging*, *5*(1), 167-184. doi: 10.3934/ipi.2011.5.167

Särkkä. (2013). *Bayesian Filtering and Smoothing*. Cambridge University Press.

Semeter, J., & Kamalabadi, F. (2005). Determination of primary electron spectra from incoherent scatter radar measurements of the auroral E region. *Radio Sci.*, *40*, RS2006. doi: 10.1029/2004RS003042

Swartz, W. E., & Farley, D. T. (1979). A Theory of Incoherent Scattering of Radio Waves by a Plasma 5. The Use of the Nyquist Theorem in General Quasi-Equilibrium Situations. *J. Geophys. Res.*, *84*(A5), 1930-1932. doi: 10.1029/JA084iA05p01930

Swoboda, J., Semeter, J., Zettergren, M., & Erickson, P. J. (2017). Observability of ionospheric space-time structure with ISR: A simulation study. *Radio Sci.*, *52*(2), 215-234. doi: 10.1002/2016RS006182

Turunen, E., Kero, A., Verronen, P. T., Miyoshi, Y., Oyama, S.-I., & Saito, S. (2016). Mesospheric ozone destruction by high-energy electron precipitation assciated with pulsating aurora. *J. Geophys. Res. Atmos.*, *121*, 11852-11861. doi: 10.1002/2016JD025015

Vierinen, J., Chau, J. L., Charuvil, H., Urco, J. M., Clahsen, M., Avsarkisov, V., ... Volz, R. (2019). Observing mesospheric turbulence with specular meteor radars: A novel method for estimating second-order statistics of wind velocity. *Earth and Space Science*, *6*(7), 1171-1195. doi: 10.1029/2019EA000570

Virtanen, I. I., Gustavsson, B., Aikio, A. T., Kero, A., Asamura, K., & Ogawa, Y. (2018). Electron Energy Spectrum and Auroral Power Estimation from Incoherent Scatter Radar Measurements. *J. Geophys. Res. Space Physics*, *123*, 6865-6887. doi: 10.1029/2018JA025636

Virtanen, I. I., Lehtinen, M. S., Nygrén, T., Orispää, M., & Vierinen, J. (2008). Lag profile inversion method for EISCAT data analysis. *Ann. Geophys.*, *26*, 571-581. doi: 10.5194/angeo-26-571-2008

Virtanen, I. I., McKay-Bukowski, D., Vierinen, J., Aikio, A. T., Fallows, R., & Roininen, L. (2014, December). Plasma parameter estimation from multistatic, multibeam incoherent scatter data. *J. Geophys. Res. Space Physics*, *119*(12), 10528-10543. doi: 10.1002/2014JA020540

Virtanen, I. I., Vierinen, J., & Lehtinen, M. S. (2009, July). Phase-coded pulse aperiodic transmitter coding. *Ann. Geophys.*, *27*, 2799-2811. doi: 10.5194/angeo-27-2799-2009

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