



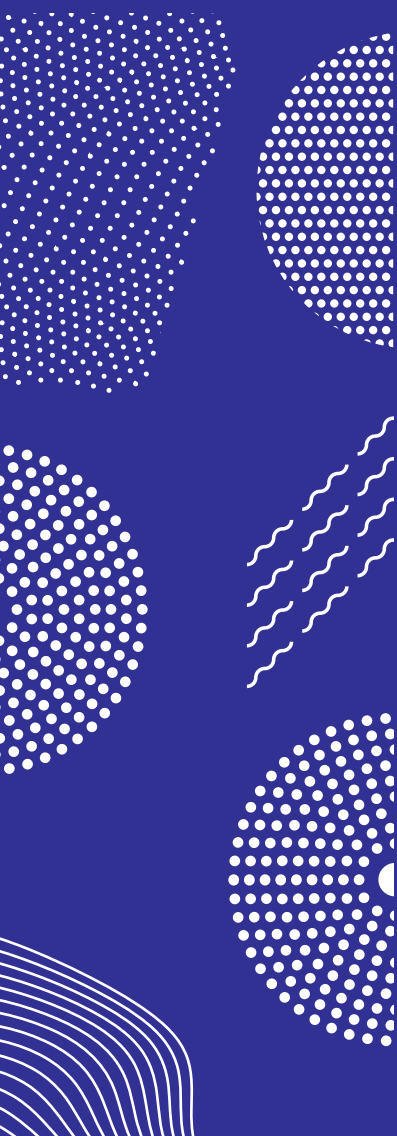
ILMATIETEEN LAITOS
METEOROLOGISKA INSTITUTET
FINNISH METEOROLOGICAL INSTITUTE

173

CONTRIBUTIONS

BAYESIAN APPROACH TO IONOSPHERIC IMAGING WITH GAUSSIAN MARKOV RANDOM FIELD PRIORS

JOHANNES NORBERG



Bayesian approach to ionospheric imaging with Gaussian Markov random field priors

Johannes Norberg

*Doctoral dissertation, to be presented for public discussion with the
permission of the Faculty of Science of the University of Helsinki,
in Auditorium PII, Porthania, Yliopistonkatu 3, Helsinki,
on the 19th of August, 2020 at 12 o'clock.*

Department of Mathematics and Statistics
University of Helsinki
Helsinki, Finland

Supervisors

Prof. Samuli Siltanen, University of Helsinki, Finland

Prof. Markku Lehtinen, Sodankylä Geophysical Observatory,
University of Oulu, Finland

Dr. Olaf Amm, Finnish Meteorological institute, Helsinki, Finland

Dr. Kirsti Kauristie, Finnish Meteorological institute, Helsinki, Finland

Pre-examiners

Prof. Aku Seppänen, University of Eastern Finland

Dr. M Mainul Hoque, Institute for Solar-Terrestrial Physics,
Neustrelitz, Germany

Opponent

Prof. Matthew Angling, Spire Global Ltd UK

Custos

Prof. Samuli Siltanen, University of Helsinki, Finland

Contact information

Department of Mathematics and Statistics
P.O. Box 64 (Gustav Hällströmin katu 2)
FI-00014 University of Helsinki
Finland
URL: <http://mathstat.helsinki.fi/>
Telephone: +358 29 419 11

Finnish Meteorological Institute
P.O. Box 503 (Erik Palménin aukio 1)
FI-00101 Helsinki Finland
URL: <https://ilmatieteenlaitos.fi/>
Telephone: +358 29 539 1000

Copyright © 2020 Johannes Norberg

ISSN: 0782-6117

ISBN: 978-952-336-123-2 (paperback)

ISBN: 978-952-336-124-9 (pdf)

<https://doi.org/10.35614/isbn.9789523361249>



Published by	Finnish Meteorological Institute (Erik Palménin aukio 1), P.O. Box 503 FIN-00101 Helsinki, Finland	Series title, number and report code of publication Contributions, 173, FMI-CONT-173 Date 17.7.2020
Author	Johannes Norberg	ORCID iD 0000-0003-3155-8894
Title	Bayesian approach to ionospheric imaging with Gaussian Markov random field priors	
Abstract		

Ionosphere is the partly ionised layer of Earth's atmosphere caused by solar radiation and particle precipitation. The ionisation can start from 60 km and extend up to 1000 km altitude. Often the interest in ionosphere is in the quantity and distribution of the free electrons. The electron density is related to the ionospheric refractive index and thus sufficiently high densities affect the electromagnetic waves propagating in the ionised medium. This is the reason for HF radio signals being able to reflect from the ionosphere allowing broadcast over the horizon, but also an error source in satellite positioning systems.

The ionospheric electron density can be studied e.g. with specific radars and satellite in situ measurements. These instruments can provide very precise observations, however, typically only in the vicinity of the instrument. To make observations in regional and global scales, due to the volume of the domain and price of the aforementioned instruments, indirect satellite measurements and imaging methods are required.

Mathematically ionospheric imaging suffers from two main complications. First, due to very sparse and limited measurement geometry between satellites and receivers, it is an ill-posed inverse problem. The measurements do not have enough information to reconstruct the electron density and thus additional information is required in some form. Second, to obtain sufficient resolution, the resulting numerical model can become computationally infeasible.

In this thesis, the Bayesian statistical background for the ionospheric imaging is presented. The Bayesian approach provides a natural way to account for different sources of information with corresponding uncertainties and to update the estimated ionospheric state as new information becomes available. Most importantly, the Gaussian Markov Random Field (GMRF) priors are introduced for the application of ionospheric imaging. The GMRF approach makes the Bayesian approach computationally feasible by sparse prior precision matrices.

The Bayesian method is indeed practicable and many of the widely used methods in ionospheric imaging revert back to the Bayesian approach. Unfortunately, the approach cannot escape the inherent lack of information provided by the measurement set-up, and similarly to other approaches, it is highly dependent on the additional subjective information required to solve the problem. It is here shown that the use of GMRF provides a genuine improvement for the task as this subjective information can be understood and described probabilistically in a meaningful and physically interpretative way while keeping the computational costs low.

Publishing unit		
Space Research and Observation Technologies		
Classification (UDC)	Keywords	
519.676, 551.510.413.5	Ionosphere, inverse problems, Bayes, data assimilation, tomography	
ISSN and series title	ISBN	
0782-6117	978-952-336-123-2 (paperback)	
Finnish Meteorological Institute Contributions	978-952-336-124-9 (pdf)	
DOI	Language	Pages
https://doi.org/10.35614/isbn.9789523361249	English	85



Julkaisija	Ilmatieteen laitos (Erik Palménin aukio 1) PL 503, 00101 Helsinki	Julkaisun sarja, numero ja raporttikoodi Contributions, 173, FMI-CONT-173 Päiväys 17.7.2020
Tekijä	Johannes Norberg	ORCID iD 0000-0003-3155-8894
Nimeke	Gaussiset Markovin satunnaiskentät bayesiläisessä ionosfääriin kuvantamisessa	
Tiivistelmä		

Ionosfääri on noin 60–1000 kilometrin korkeudella sijaitseva ilmakehän kerros, jossa kaasuatomien ja -molekyylien elektroneja on päässyt irtomaan auringon säteilyn ja auringosta peräisin olevien nopeiden hiukkasten vaikutuksesta. Näin syntyneillä ioneilla ja vapailla elektroneilla on sähkö- ja magneettikenttien kanssa vuorovaikuttava sähkövaraus. Ionosfäärillä on siksi merkittävä rooli radioliikenteessä. Se voi mahdollistaa horisontin yli tapahtuvat pitkät radiolähetykset heijastamalla lähetetyn sähkömagneettisen signaalin takaisin maata kohti. Toisaalta ionosfääri vaikuttaa myös sen läpäiseviin korkeampitaajuuksiin signaaleihin. Esimerkiksi satelliittipaikannuksessa ionosfääriin vaikutus on parhaassakin tapauksessa otettava huomioon, mutta huonoimmassa se voi estää paikannuksen täysin. Näkyvin ja tunnetuin ionosfääriin liittyvä ilmiö lienee revontulet.

Yksi keskeisistä suureista ionosfääriin tutkimuksessa on vapaiden elektronien määrä kuutiometrin tilavuudessa. Käytännössä elektronitiheyden mittaaminen on mahdollista mm. tutkilla, kuten Norjan, Suomen ja Ruotsin alueilla sijaitsevalla EISCAT-tutkajärjestelmällä, sekä raketti- tai satelliittimittauksilla. Mittaukset voivat olla hyvinkin tarkkoja, mutta tietoa saadaan ainoastaan tutkakeilan suunnassa tai mittalaitteen läheisyydestä. Näillä menetelmillä ionosfääriin tutkiminen laajemmalla alueella on siten vaikeaa ja kallista.

Olemassa olevat paikannussatelliitit ja vastaanotinverkot mahdollistavat ionosfääriin elektronitiheyden mittaamisen alueellisessa, ja jopa globaalissa mittakaavassa, ensisijaisen käyttötarkoituksensa sivutuotteena. Satelliittimittausten ajallinen ja paikallinen kattavuus on hyvä, ja kaiken aikaa kasvava, mutta esimerkiksi tarkkoihin tutkamittauksiin verrattuna yksittäisten mittausten tuottama informaatio on huomattavasti vähäisempää.

Tässä väitöstyössä kehitettiin tietokoneohjelmisto ionosfääriin elektronitiheyden kolmiulotteiseen kuvantamiseen. Menetelmä perustuu matemaattisten käänteisongelmien teoriaan ja muistuttaa lääketieteessä käytettyjä viipalekuvausmenetelmiä. Satelliittimittausten puutteellisesta informaatiosta johtuen työssä on keskitytty etenkin siihen, miten ratkaisun löytymistä voidaan auttaa tilastollisesti esitetyllä fysikaalisella ennakkotiedolla.

Erityisesti työssä sovellettiin gaussisiin Markovin satunnaiskenttiin perustuvaa uutta korrelaatiopriori-menetelmää. Menetelmä vähentää merkittävästi tietokonelaskennassa käytettävän muistin tarvetta, mikä lyhentää laskenta-aikaa ja mahdollistaa korkeamman kuvantamisresoluution.

Julkaisijayksikkö		
Avaruustutkimus ja havaintoteknologiat		
Luokitus (UDK)	Asiasanat	
519.676, 551.510.413.5	Ionosfääri, inversio-ongelmat, Bayes, data-assimilaatio, tomografia	
ISSN ja avainnimeke	ISBN	
0782-6117	978-952-336-123-2 (paperback)	
Finnish Meteorological Institute Contributions	978-952-336-124-9 (pdf)	
DOI	Kieli	Sivumäärä
https://doi.org/10.35614/isbn.9789523361249	Englanti	85

List of publications

This thesis consists of an introductory part and the following original publications:

- I** J. Norberg, L. Roininen, J. Vierinen, O. Amm, D. McKay-Bukowski, and M. S. Lehtinen, Ionospheric tomography in Bayesian framework with Gaussian Markov random field priors, *Radio Sci.*, 50(2): 138–152, 2015. ISSN 1944799X. doi: 10.1002/2014RS005431.
- II** J. Vierinen, J. Norberg, M. S. Lehtinen, O. Amm, L. Roininen, A. Väänänen, P. J. Erickson and D. McKay-Bukowski. Beacon satellite receiver for ionospheric tomography *Radio Sci.*, 49(12):1141-1152, 2014. ISSN 1944799X. doi: 10.1002/2014RS005434.
- III** J. Norberg, I. I. Virtanen, L. Roininen, J. Vierinen, M. Orispää, K. Kauristie, and M. S. Lehtinen, Bayesian statistical ionospheric tomography improved by incorporating ionosonde measurements, *Atmos. Meas. Tech.*, 9(4): 1859–1869, 2016. ISSN 18678548. doi: 10.5194/amt-9-1859-2016.
- IV** J. Norberg, J. Vierinen, L. Roininen, M. Orispää, K. Kauristie, W. C. Rideout, A. J. Coster, and M. S. Lehtinen, Gaussian Markov Random Field Priors in Ionospheric 3-D Multi-Instrument Tomography, *IEEE Trans. Geosci. Remote Sens.*, 1–13, 2018. ISSN 1558-0644. doi: 10.1109/TGRS.2018.2847026.

Author's contribution

Publication I: “Ionospheric tomography in Bayesian framework with Gaussian Markov random field prior”

The paper presents the Gaussian Markov random field priors for two-dimensional ionospheric imaging. It is shown how the matrices providing the prior covariance information are built numerically. A simulation study is made by generating samples from prior distributions, from where tomographic measurements are simulated assuming a low Earth orbit satellite overflight and five ground-based receivers. The sampled ionosphere is then reconstructed with the presented algorithm in lower resolution. The results demonstrate the performance of Bayesian inversion in an unrealistic situation where the prior used to simulate the unknown electron density is known exactly and in a more realistic case, where a more rough prior is used. All the numerical simulations and main parts of the writing were carried out by the author.

Publication II: “Beacon satellite receiver for ionospheric tomography”

The paper demonstrates a dual frequency receiver for ground-based measurements of 150 and 400 MHz signals transmitted from low Earth orbit beacon satellites. The paper is a companion paper for Publication I and uses the method presented there for reconstructing the two-dimensional ionospheric electron density from measurements obtained with four new ground-based receivers from a COSMOS 2463 satellite overflight. The results are compared with results from an existing tomographic system in the same region. The deployment of most of the receivers, numerical analysis and writing for the parts concerning tomography were carried out by the author.

Publication III: “Bayesian statistical ionospheric tomography improved by incorporating ionosonde measurements”

The paper presents a study where the developed tomography method is used with different prior models. The compared prior models include the International Reference Ionosphere 2007 model, extrapolation of an European Incoherent Scatter Scientific Association's (EISCAT) ionosonde measurements and a prior assuming a zero electron density. The two-

dimensional results are validated with EISCAT's ultra-high frequency incoherent scatter radar measurements. The numerical analysis and main parts of the writing were carried out by the author. The EISCAT measurements were designed and carried out by the author, while I. I. Virtanen provided reanalysis for the radar measurements.

Publication IV: “Gaussian Markov Random Field Priors in Ionospheric 3-D Multi-Instrument Tomography”

This is the main paper of this thesis. It generalises the method presented in the earlier articles for multi-instrumental three-dimensional ionospheric imaging. The multi-instrument set-up consists of dense ground-based networks of radio receivers for GPS satellite signals, a low Earth orbit satellite receiver network, ionosonde, satellite occultation as well as satellite in situ measurements. The results are validated with EISCAT's ultra- and very-high frequency incoherent scatter radars. The numerical benefits from the sparsity of Gaussian Markov random field priors, inclusion of time propagation and differential bias correction of GPS satellite data are discussed. All the numerical simulations and main parts of the writing were carried out by the author. A. Coster and W. C. Rideout provided the differential bias correction for the GPS data.

Acknowledgements

I would like to pay my special regards to Markku Lehtinen for the original scientific ideas and for taking me in his group already at a very early stage of my studies.

I wish to express my deepest gratitude to the late Olaf Amm who first took me in FMI as his PhD student and got all this started. After Olaf, Kirsti Kauristie took me under her wing, and I would like to thank her for providing me with extraordinary time and space to carry out my endeavours, as well as for the kind support at every turn along the way.

I am thankful to my custos and supervisor at University of Helsinki, Samuli Siltanen, who took me in as a pig in a poke and has provided important support ever since.

I would like to thank Aku Seppänen and Mainul Hoque for the time and effort they put in the pre-examination of my thesis. The supportive feedback from respected professionals encouraged me a lot at the final stages of my work.

My wish was to have one of the true experts in the ionospheric imaging community to act as my opponent. Hence, I am very thankful to Matthew Angling for accepting the invitation.

This study has been carried out in Finnish Meteorological Institute, Helsinki, and Sodankylä Geophysical Observatory, University of Oulu. I wish to thank both organisations. I am especially grateful to Ari-Matti Harri, Jouni Pulliainen and Esa Turunen for their exemplary leadership that has provided an active, science-focused and inspirational working environment.

I cannot thank my colleagues enough. Without Antti Kero, Sebastian Käki, Sari Lasanen, Mikko Orispää, Pentti Posio, Tero Raita, Tomi Teppo, Thomas Ulich, Juha Vierinen and Ilkka Virtanen I would not have been able to complete this research. I am particularly grateful to Lassi Roininen, whose contribution to my study has been essential. In addition to his scientific findings, Lassi's interactive way of working and networks have helped me greatly.

I would like to thank all the collaborators: Anita Aikio, Anthea Coster, Michael Fletcher, Maxime Grandin, Esa Kallio, Thomas Leyser, Derek McKay, Minna Palmroth, William Rideout, Mike Rietveld, Tomas Tallkvist, Heikki Vanhamäki and Dan Whiter. I am looking forward to future collaborative projects.

I would like to acknowledge Baylie Damtie, Björn Gustavsson, Heikki Haario, Marko Laine, Jussi Markkanen, Markku Markkanen, Petteri Piironen, Jouni Susiluoto and Simo

Särkkä for the valuable discussions, advice and help they have provided.

I would like to thank Mwaba Hiltunen, Tomi Karppinen, Ulpu Leijala, Kimmo Rautainen and Miia Salminen for their invaluable peer support.

I am also indebted to Pilvi Ahonen, Riitta Aikio, Dan Anderson, Mikael Frisk, Minna Huuskonen, Sari Jokiniemi, Marina Kurten, Juha Lemmetyinen, Harry Lonka, Kari Mäenpää, Teija Manninen, Marita Mökkönen, Arto Oksanen, Noora Partamies, Kaisa Rynänen, Mikko Syrjäso, Matias Takala, Juho Vehviläinen and Riika Ylitalo for the help they have given me.

I am lucky to have so many good friends, old and new, whom I would like to thank, that the list would be too long and I would still forget someone. I hope that you know who you are.

My biggest gratitude goes first to my beloved godchildren: Lilja, Olavi, Siiri and Tatu, and finally, to my dear family: my grandmother Elma, sisters Taru and Anna, father Kauko, my mother Annukka and little niece Li.

Helsinki, Finland, July 2020
Johannes Norberg

Contents

List of publications	9
Author's contribution	10
1 Introduction	17
2 Tomography	22
2.1 Radon transform	22
2.2 Filtered backprojection algorithm (FBP)	23
2.3 Incomplete data	23
2.4 Discrete model	24
3 Inverse problem	27
3.1 General model	27
3.2 Linear model	27
3.2.1 Discretisation	28
3.3 Ill-posed problem	28
3.4 Classical regularisation methods	29
3.4.1 Least squares solution (LS)	29
3.4.2 Minimum norm solution (MN)	30
3.4.3 Truncated singular value decomposition (TSVD)	30
3.4.4 Tikhonov regularisation	31
3.4.5 Generalised Tikhonov regularisation	31
3.5 Iterative solutions for linear system	32
3.5.1 Kaczmarz method	32
3.5.2 Algebraic reconstruction technique (ART)	33
3.5.3 Multiplicative algebraic reconstruction technique (MART)	34
3.5.4 Simultaneous iterative reconstruction technique (SIRT)	34
3.5.5 Simultaneous algebraic reconstruction technique (SART)	34

4	Bayesian statistical approach	36
4.1	Introduction to Bayesian inference	36
4.2	Gaussian priors for linear inverse problems	38
4.2.1	Continuous Gaussian random field prior	39
4.2.2	Discrete multivariate Gaussian prior	40
4.2.3	Model space solution	41
4.3	Gaussian Markov random field (GMRF) priors	41
4.3.1	Correlation priors	42
5	Spatiotemporal evolution	44
5.1	Recursive linear estimation	44
5.2	Kalman filtering	45
5.3	Kalman smoothing	46
5.4	Ensemble Kalman filter (EnKF)	46
6	Ionospheric measurements	49
6.1	Electromagnetic wave propagation	49
6.1.1	Ionospheric refractive index	50
6.1.2	Group refractive index	51
6.1.3	Tropospheric refractive index	52
6.2	Radio measurements of satellite transmissions	52
6.2.1	Refractive indices for VHF and UHF signals	52
6.2.2	Wave propagation of VHF and UHF signals	53
6.2.3	Observables	54
6.2.4	Carrier phase leveling	59
6.2.5	LEO beacon satellite measurement model	60
6.2.6	GNSS satellite measurement model	61
6.3	Ionosonde measurements	62
6.4	Incoherent scatter radar measurements	64
6.5	Langmuir probe in situ measurements	65
7	Development of methodology in ionospheric imaging	66
7.1	TomoScand	69
8	Discussion and conclusions	73
	Publications	87

Chapter 1

Introduction

The ionosphere is a shell of ionisation surrounding the Earth. The ionisation is controlled by solar radiation, particle precipitation, and interactions with the electrically neutral atmosphere. For ionospheric imaging the key plasma parameter is the electron density i.e. the number of free electrons divided by unit volume, often given in scaled units of $\frac{10^{11}}{\text{m}^3}$. The atmospheric electron density is typically horizontally stratified and depends on factors including latitude, season, local time and solar activity. Figure 1.1 presents four vertical incoherent scatter radar measurement profiles of typical daytime ionospheric electron density over Tromsø, Norway. Generally, the electron density maximum takes place around an altitude of 300 km at the so-called F region. Below, around an altitude of 100 km is the E region. In local daytime the E region can be seen as a small enhancement of electron density below the much higher density in the F region. However, at high latitudes during auroral particle precipitation events, especially the E region can have very rapid changes with peak electron densities exceeding that of the F region. The D region takes place at altitudes between 60 and 90 km. The conditions in the D region are strongly coupled with neutral atmospheric processes and the region often has relatively small electron density. The ionosphere extends to around an altitude of 1000 km where it transforms into a plasmasphere with substantially lower electron content. The ionospheric electron density is also often described as *total electron content* (TEC) i.e. the integrated electron density between two locations. Vertical TEC (VTEC) is the TEC integrated along a vertical column. TEC and VTEC are usually given in TEC units $\left(1 \text{ TECU} = \frac{10^{16}}{\text{m}^2}\right)$.

Ionospheric electron density can be observed e.g. with incoherent scatter radars, ionosondes, satellite in situ measurements and remote measurements of the global navigation satellite system (GNSS) and low Earth orbit (LEO) satellite beacons. A two-dimensional simplification of different ionospheric electron density measurements is given in Figure 1.2.

In ionospheric imaging the aim is to reconstruct the two- or three-dimensional electron density from available measurements. The ground-based measurements of GNSS satellite beacon signals is typically the most important data component. The use of the terms

imaging, *tomography* and *data assimilation* is somewhat mixed in the ionospheric literature. The term *imaging* is usually used as a general term to cover the different reconstruction methods. In an optimal case of *tomography*, the unknown would be reconstructed mostly from the available measurements. When operating regionally, especially in two-dimensional cases, the situation in ionospheric imaging is similar to conventional tomographic problems such as medical X-ray tomography, and thus many of the same techniques have been used. On the other hand, in Global three- and four-dimensional situations, the measurements can be extremely sparse and even relatively large areas can be left without any measurements. In these situations some strict background models are required and combined optimally with the available observations. A more illustrative and the most commonly used term in this case is *data assimilation*. Data assimilation and its nomenclature originates mostly from the field of numerical weather prediction.

Even in the best situation, due to limitations in the measurement geometry, the ionospheric imaging problem can be considered a limited angle tomography problem with sparse measurements. This rules out the the generally widely used tomographic algorithms that are based on backprojection. Mathematically the tomographic imaging of ionosphere is an ill-posed inverse problem. In practice this means that the measurements do not contain enough information of the unknown electron densities to give a unique and realistic solution.

Most of the early approaches to ionospheric imaging were based on iterative reconstruction techniques that were developed independently within the fields of image processing and linear algebra. The starting point is an initial guess about the unknown, which is then modified iteratively to correspond with the measurements. The downside is that with incomplete data the result is very dependent on the initial value.

Another approach is provided by so-called classical regularisation methods. With classical regularisation the original problem is modified to as a similar well-posed problem as possible. The problem here is that the interpretation of classical regularisation methods is mostly mathematical: the reason for numerical instability is examined and adjusted. In severely ill-posed problems, as in ionospheric imaging, it can be difficult to interpret the regularisation physically. On the other hand, there can be a lot of physical information available that is difficult to represent accurately with these methods.

In the division used here, the last family of ionospheric imaging methods is provided by the Bayesian approach. In the Bayesian approach, a prior distribution is used to control the set of possible solutions. The prior distribution can often be understood as a probabilistic description of the uncertainty related to the physical quantity of interest. Even though the information in prior distribution and hence the whole approach can be considered subjective, there often exists indisputable physical information that can be used in the construction of the prior. Also, as in ill-posed problems some additional information is required in any case, it is beneficial to know how the information limits the possible solutions. Most of the data assimilation methods used in ionospheric imaging are Bayesian, where physical background models are used in the determination of the prior distribution.

The problems with the Bayesian approach are mostly computational. The numerical computations with proper probability distributions require operations with covariance matrices. Especially in the three-dimensional case the covariance matrices can become excessively large for computation. Hence, one way of seeing the differences within the Bayesian approaches is how the formation and computation of covariance matrices is handled.

In this work Gaussian Markov random field (GMRF) priors are introduced for Bayesian ionospheric imaging. GMRF is a Gaussian random field, but instead of mean and covariance, it is more conveniently defined with its mean and inverse covariance i.e the precision matrix. Following Roininen et al. (2011, 2013), with a suitable parametrisation, the precision matrix of a GMRF can give close approximations for known covariance functions and due to Markov property, the precision matrices are sparse matrices. This reduces the computational costs significantly, making the direct inversion possible for relatively large three-dimensional cases. The use of GMRF then allows the usage of proper prior distributions with physical interpretation, while keeping the computational burden similar to the classical regularisation methods.

The structure of this dissertation summary is the following. Chapter 2 introduces the mathematical background of tomography and the commonly used backprojection methods. In Chapter 3, the measurement model, the resulting linear inverse problem, the classical regularisation method solutions and the iterative solution techniques are presented. The Bayesian approach, and most importantly, the GMRF priors are introduced in Chapter 4. In the original publications, the modelling of the temporal dynamics is discussed only shortly in Publication IV. The generalisation of the method for the spatiotemporal situation is somewhat straightforward, but in a broader context so central that the most used recursive filtering algorithms are presented in Chapter 5. The different ionospheric measurements are then exhibited in Chapter 6. In Chapter 7, a review on the usage and development of the aforementioned imaging methods within the ionospheric research, as well as a description of the numerical method developed within this work is given. Finally, discussion and conclusions are provided in Chapter 8.

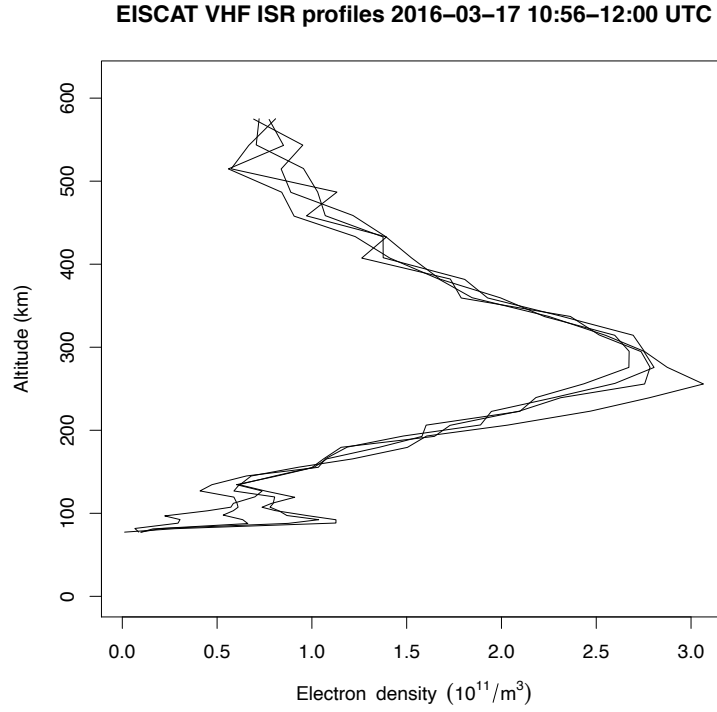


Figure 1.1: Four measurement profiles from European incoherent scatter scientific association’s very-high frequency incoherent scatter radar in Tromsø, Norway. The profiles depict the typical vertical structure of daytime ionosphere, with the F-region maximum just below 300 km and the local E-region maximum around 100 km. Local time (UTC + 1 h).

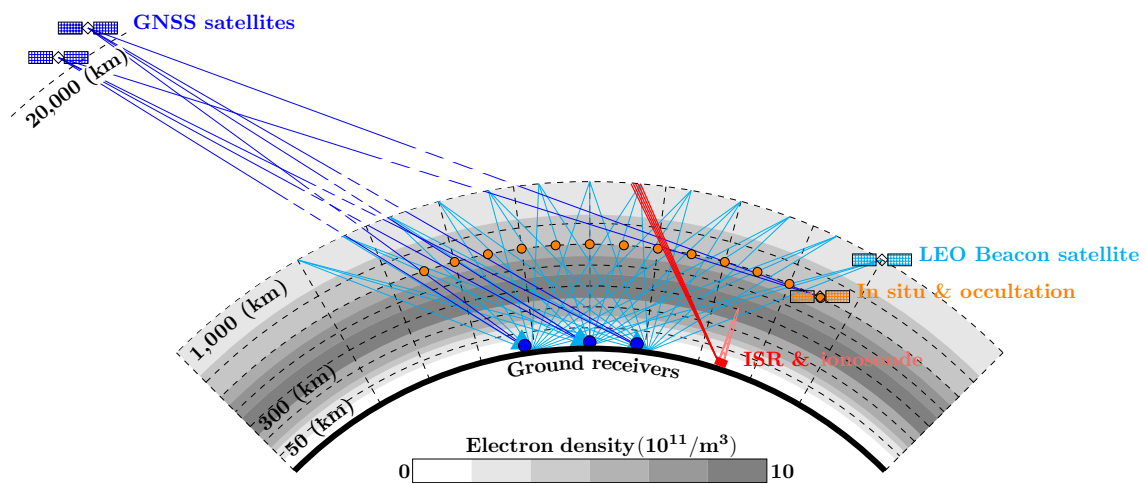


Figure 1.2: Two-dimensional simplification of measurements used in ionospheric imaging.

Chapter 2

Tomography

Tomography refers to cross-sectional imaging of an object from measurements provided by some penetrating waves. Tomographic methods are used in various fields from medicine to geophysics. Good overviews on tomography are provided by Kak and Slaney (1988); Natterer and Wübbeling (2001); Hsieh (2009).

Arguably the simplest and most common type of tomographic set-up is the parallel beam tomography. As the name suggests, several beams are transmitted in parallel on a two-dimensional plane. The beams propagate through the domain and are measured on the opposite side on a plane receiver. In X-ray tomography the measurement would be the attenuation of an X-ray signal during its pass. The set-up of transmitter and receiver planes is then circled around the object, to provide measurements from all directions. Besides the parallel beam tomography, there are various different scanning geometries, the fan beam and cone beam scans being probably the best known alternatives of regular scans.

2.1 Radon transform

Mathematically the situation in parallel beam tomography can be written by describing an unknown image as a function $f : \Psi \rightarrow \mathbb{R}$ on a physical domain $\Psi \in \mathbb{R}^2$. A measurement along a signal path of an angle perpendicular to θ and distance s from the origin can then be written generally as a *Radon transform*

$$\mathcal{R}f(\theta, s) = \int_{L(\theta, s)} f(z) dz = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(z_1, z_2) \delta(z_1 \cos \theta + z_2 \sin \theta - s) dz_1 dz_2, \quad (2.1)$$

where $z = (z_1, z_2) \in \Psi$ and δ is a delta function defining the signal path $L(\theta, s)$ as a line in the image domain. With a fixed θ

$$P_\theta(s) := \mathcal{R}f(\theta, s), \quad (2.2)$$

where $P_\theta(s)$ is the projection corresponding to parallel beam measurements made in a direction perpendicular to θ .

2.2 Filtered backprojection algorithm (FBP)

The main task in tomography is to reconstruct the unknown image $f(z_1, z_2)$ from the measured projections P_θ . The most straightforward approach is to backproject each measurement over the image domain along the corresponding signal path. When all backprojections are summed, the internal structures will accumulate in the reconstruction. However, the simple backprojection typically produces blurred results. Intuitively the blurring effect can be understood if one assumes a local minimum point with a value of zero. If the Radon transform is non-zero for any of the intersecting lines, in practice the reconstructed value of that point will always be greater than zero.

The blurring can be avoided with the *filtered backprojection* (FBP) algorithm. The FBP is based on the Fourier slice theorem, which states that when a two-dimensional image is projected to a plane with an angle θ , the one-dimensional Fourier transform of that projection corresponds to a radial slice of the two-dimensional Fourier transform of the same angle θ . Hence, the two-dimensional frequency domain of the unknown image can be built slice by slice with the Fourier transformed tomographic projections.

The use of inverse Fourier transform requires an interpolation to a rectangular grid in the frequency domain, or preferably, a change of variables between polar and rectangular coordinates. The change of variables introduces a Jacobian that can be interpreted as a high-pass filter. This is the filter part of FBP and in the frequency domain it is a multiplication operation. The filtering highlights edges and reduces the blurring in the final image.

Especially in most medical applications, where the conditions are well controlled and extensive measurements can be performed, the FBP is the first choice algorithm for its accuracy and relative ease of implementation (Kak and Slaney, 1988). However, the problems with FBP arise especially in the situations where the information provided by the measurements is limited and the Radon transform (2.1) is known only partially with respect to parameter θ or s or both. Including regularising additional information is difficult, hence the approach is severely affected by the incompleteness of data. It is also assumed in FBP that the measurements are precise and thus the measurement errors cannot be modelled explicitly.

2.3 Incomplete data

A common type of incompleteness in tomographic data is referred to as a *limited angle* tomography. As the name suggests, in limited angle tomography the θ angles are available only from a subset of the optimal sphere/half sphere. Examples of such cases are e.g. dental imaging (Hyvönen et al., 2010) and most geophysical tomographic problems such as borehole tomography (Justice et al., 1989).

Another type of incompleteness is the sparseness of the data. *Sparse tomography* can sometimes refer to sparseness of available measurement angles and hence overlap with the

definition of limited angle tomography above. However, it can also express the availability of different possible s and hence the resolution of each projection. The source points s of measurement paths can also be limited in range. With $q \in \mathbb{R}^+$ as a limiting constant, the situation $|s| > q$ is called an *exterior problem* and $|s| < q$ an *interior problem* (Natterer and Wübbeling, 2001).

The aforementioned limitations can be caused e.g. by inherent physical obstructions and inconveniences or medical or economical incentives. As a typical example, in medical tomography the patients exposition to harmful X-rays needs to be kept down, hence the radiation dose is reduced by using sparser angular resolution for measurement directions.

Another limiting factor can be the dynamics of the system. Even if the scanning geometry could provide sufficient accuracy with respect to measurement angles and amount and distribution of measurement paths, the unknown object can experience temporal changes in a shorter time scale than it takes to perform all the measurements (Hahn, 2015).

In satellite tomography, the angles in satellite-to-ground measurements are naturally limited (Figure 1.2). Additionally, for instance Brekke (1997) reports multifold change in electron density within 20 s, whereas GNSS data is typically integrated at least for some minutes for sufficient spatial coverage. Hence, ionospheric tomography can be considered a sparse limited angle tomography problem with relatively high temporal dynamics.

2.4 Discrete model

In practice the measurements (2.1) are made from a finite number of points and angles. Here R projections are assumed from a half sphere as

$$\boldsymbol{\theta} = \left(0, \frac{\pi}{R}, 2\frac{\pi}{R}, \dots, (R-1)\frac{\pi}{R}\right)^T = (\theta_1, \dots, \theta_R)^T \in \mathbb{R}^R.$$

For each angle θ , the $P_\theta(s)$ in Equation (2.2) is projected on S points

$$\mathbf{s} = \left(\left(1 - \frac{S+1}{2}\right)\Delta s, \dots, \left(S - \frac{S-1}{2}\right)\Delta s\right)^T = (s_1, \dots, s_S) \in \mathbb{R}^S,$$

where Δs is the lateral offset between two adjacent projection points.

Altogether, this discretisation will provide measurements

$$\mathbf{m} = (P_{\theta_1}(s_1), \dots, P_{\theta_1}(s_S), \dots, P_{\theta_R}(s_S))^T = (m_1, \dots, m_j, \dots, m_M)^T \in \mathbb{R}^M,$$

where $M = RS$. Similarly, the corresponding lines L are denoted as

$$\mathbf{L} = (L(\theta_1, s_1), \dots, L(\theta_1, s_S), \dots, L(\theta_R, s_S))^T = (L_1, \dots, L_j, \dots, L_M)^T \in \mathbb{R}^M,$$

For notational clarity, the one-index representation of the right-hand side will be used for all above measurement variables in the sequel.

For numerical modelling, also the unknown function needs to be discretised and understood as an array of unknown values. Typically the function is evaluated on a cartesian grid on the domain Ψ . As the unknown function is typically an image, it is easier to understand and visualise as a matrix, but for the algebraic and notational convenience it is collapsed to a vector and reindexed to one-index representation

$$\mathbf{f} := f(\mathbf{z}) = (f(z_{1,1}), \dots, f(z_{n,1}), \dots, f(z_{n,n}))^T \in \mathbb{R}^N,$$

where an $n \times n = N$ discretisation is made at points $\mathbf{z} = (z_{1,1}, \dots, z_{n,1}, \dots, z_{n,n})^T \in \mathbb{R}^{N \times 2}$.

Each measurement m_j can then be modelled as an integral in Equation (2.1) and approximated as a Riemann sum

$$m_j = \int_{L_j} f(z) dz \approx \sum_{i=1}^N a_{ji} f_i, \quad (2.3)$$

where $a_{ji} \in \mathbb{R}$ gives the intersection length between the path L_j and pixel i . In matrix form the measurements can then be written as

$$\mathbf{m} \approx \mathbf{A} \mathbf{f}, \quad (2.4)$$

where $\mathbf{A} \in \mathbb{R}^{M \times N}$ is a *theory matrix*, where row j is a vector $\mathbf{a}_j = (a_{j1}, \dots, a_{ji}, \dots, a_{jN}) \in \mathbb{R}^N$.

Generally, the extension to three-dimensional tomography is often carried out by reducing the problem to several two-dimensional problems and reconstructed layer by layer. An alternative approach is to move the two-dimensional scan along the axis of symmetry during the scan. This will result in a three-dimensional helical scan. In cone beam tomography the setup is similar to fan beam, but whereas the fan beam is considered two-dimensional and the corresponding measurement one-dimensional, here the transmitted signal is a three-dimensional cone, received on a plane as a two-dimensional measurement. The Radon transform (2.1) and its inverse apply directly to parallel beam geometry, but alternative formulations for different scans are available and provided e.g. by Natterer and Wübbeling (2001). In ionospheric tomography, in a case where one satellite overflight is measured over a chain of receivers, the problem can be modelled as two-dimensional tomography. As the satellites have different orbits, in a general case where all possible measurements from several satellites are utilised, the measurements take place irregularly in a volume and the problem needs to be modelled in three dimensions.

Three-dimensional discrete model

In a three-dimensional case where the tomographic analysis is carried out directly in $\Psi \in \mathbb{R}^3$, the dimension of the unknown increases to

$$\mathbf{f} := f(\mathbf{z}) = (f(z_{1,1,1}), \dots, f(z_{n,1,1}), \dots, f(z_{n,n,1}), \dots, f(z_{n,n,n}))^T \in \mathbb{R}^N,$$

where now

$$\mathbf{z} = (z_{1,1,1}, \dots, z_{n,1,1}, \dots, z_{n,n,1}, \dots, z_{n,n,n})^T \in \mathbb{R}^{N \times 3} \quad (2.5)$$

and $n \times n \times n = N$. The Equations (2.3) and (2.4) remain the same with the corresponding change in dimensions.

As each measurement is here assumed an integral over a line, one measurement intersects only with a small portion of voxels. It is then notable that as the index i in the discrete models run through all unknown voxels, most of the intersection lengths a_i are typically zero and thus the matrix \mathbf{A} is a so-called sparse matrix (see Section 4.3).

Chapter 3

Inverse problem

This section presents the general measurement model used in ionospheric imaging. As will be shown later in Chapter 6, most of the measurements used in the ionosphere imaging are fairly straightforward to linearise. Hence, here the focus is on the linear case. Particular attention is paid to the mathematical interpretation of issues caused by incomplete data and the means to overcome them. The main references for this chapter are Kaipio and Somersalo (2005), Calvetti and Somersalo (2007), Mueller and Siltanen (2012).

3.1 General model

A general forward model with measurement error is here given as

$$\mathbf{m} = \mathcal{A}(f, \boldsymbol{\varepsilon}), \quad (3.1)$$

where $f : \mathbb{R}^d \rightarrow \mathbb{R}$, \mathcal{A} is a possibly nonlinear observation operator applied to function f and $\mathbf{m} \in \mathbb{R}^M$ the corresponding measurement vector. All physical measurements suffer from some degree of measurement errors. The error can be related to instrumentation, measurement conditions, or natural variability in the measured phenomena etc. Here a general measurement error $\boldsymbol{\varepsilon}$ is included in the model.

3.2 Linear model

If the observation operator is linear and the measurement error additive, the model can be written as

$$\mathbf{m} = Af + \boldsymbol{\varepsilon}, \quad (3.2)$$

where $f : \mathbb{R}^d \rightarrow \mathbb{R}$, A is a linear observation operator applied to function f and the measurement error vector $\boldsymbol{\varepsilon} \in \mathbb{R}^M$ is now additive. Here a zero-mean Gaussian measurement error $\boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}})$ is assumed. The measurement error and the function f are also assumed statistically independent $\boldsymbol{\varepsilon} \perp f$.

3.2.1 Discretisation

For numerical computations, a discrete model is required. Here a model

$$\mathbf{m} = \mathbf{A}\mathbf{f} + \boldsymbol{\varepsilon}, \quad (3.3)$$

is assumed, where similarly to Equation (2.4), the vector $\mathbf{f} \in \mathbb{R}^N$ is a discrete approximation of f on lattice $\mathbf{z} \in \mathbb{R}^{N \times d}$ and $\mathbf{A} \in \mathbb{R}^{M \times N}$ a linear transformation matrix that is a discrete approximation of A .

The discrete numerical model is always inaccurate compared to real-world measurements and can induce errors. For discussion on modelling errors see Kaipio and Somersalo (2007).

3.3 Ill-posed problem

Modelling of physical phenomena often results in mathematical problems where the unknown quantities of interest are measured indirectly. The actual measured property is not the primary interest, but physically and mathematically related to it. In the model equations of the previous section, the interest is not in the measurement \mathbf{m} , but in the unknown \mathbf{f} . The task is, given the measurements and the measurement model, solve the unknown \mathbf{f} , or, in presence of a noise model, estimate \mathbf{f} . The task is generally known as an *inverse problem*. To understand the origins of the difficulties and inaccuracies that arise with the inverse problems, the concept of a well-posed problem is first recalled.

Following Hadamard, a *well-posed* problem satisfies the following properties:

1. The solution exists.
2. The solution is unique.
3. The solution changes continuously with respect to the data.

If one or several of these properties is violated, the problem is referred to as *ill-posed*. In the linear situation given in Equation (3.3), the first condition is fulfilled if and only if $\mathbf{m} \in \text{Range}(\mathbf{A})$. It can be violated by the approximative nature of matrix \mathbf{A} and the noise model. The second condition is fulfilled if and only if $\text{Ker}(\mathbf{A}) = \{0\}$, which depends on the geometry of the measurements. To see how and when the third condition is violated the *singular value decomposition* (SVD) becomes an essential tool.

SVD of matrix \mathbf{A} can be written as

$$\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^T, \quad (3.4)$$

where $\mathbf{U} \in \mathbb{R}^{M \times M}$ and $\mathbf{V} \in \mathbb{R}^{N \times N}$ are orthogonal matrices and

$$\mathbf{D} \in \mathbb{R}^{M \times N} = \text{diag}(d_1, \dots, d_{\min(M,N)}) \quad (3.5)$$

is a non-negative diagonal matrix, where $d_1 \geq d_2 \geq \dots \geq d_{\min(M,N)}$ are singular values. Generally \mathbf{D} is not a square matrix and even if $N = M$ it can be singular or near singular. This can be concluded from the *condition number*

$$\text{Cond}(\mathbf{A}) = \frac{d_1}{d_{\min(M,N)}}. \quad (3.6)$$

For the third condition of Hadamard, it is required that the condition number of \mathbf{A} is not excessively large (Kaipio and Somersalo, 2005). A matrix with a large condition number is called *ill-conditioned*.

In exact equations of linear algebra one would be concerned with only the first and the second of Hadamard's conditions. The concept of ill-conditioning rises with real world measurements or computational approaches that are contaminated with errors. An ill-conditioned problem is sensitive to errors as even small measurements errors can get amplified to have unrealistically large effects on the numerical solution.

To obtain a unique and stable solution for an inverse problem, some manouvres are required to overcome the ill-posedness. In the following sections, first the classical direct regularisation methods are presented, then the most commonly used iterative reconstruction techniques are described, before going to the Bayesian approach for inverse problems.

3.4 Classical regularisation methods

Approaches to solving the ill-posed problem are often referred to as regularisation methods, stabilisation or prior information. Usually, the procedure can be seen as not solving the original ill-posed problem, but a very similar one that is well-posed.

In the following, the most commonly used regularisation approaches are presented. SVD (3.4) is used here to display the ill-conditioning effect, as well as to demonstrate how the differences between the methods can be reduced to selection of a suitable diagonal matrix to replace the nonexistent inverse of the singular value matrix \mathbf{D} (3.5).

3.4.1 Least squares solution (LS)

For an overdetermined linear inverse problem of the form given in Equation (3.3), often the first attempt to obtain a solution is made with the *least squares* (LS) method

$$\mathbf{f}_{\text{LS}} = \arg \min_{\mathbf{f} \in \mathbb{R}^N} \|\mathbf{m} - \mathbf{A}\mathbf{f}\|^2 = \mathbf{A}^\dagger \mathbf{m} \quad (3.7)$$

where

$$\mathbf{A}^\dagger = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T = \mathbf{V} \mathbf{D}^\dagger \mathbf{U}^T \quad (3.8)$$

and

$$\mathbf{D}^\dagger = (\mathbf{D}^T \mathbf{D})^{-1} \mathbf{D}^T = \text{diag}(1/d_1, \dots, 1/d_N) \in \mathbb{R}^{N \times M}. \quad (3.9)$$

In many occasions, in linear inverse problems, the matrix \mathbf{A} can have so strong linear dependencies that, despite $M > N$, the problem is effectively underdetermined and ill-conditioned. Often this can be observed from rapidly decreasing singular values. The least squares method is unable to provide a reasonable solution for severely ill-conditioned problems.

3.4.2 Minimum norm solution (MN)

For an underdetermined system the least squares method fails as it cannot select a unique value satisfying the minimisation criteria of Equation (3.7). As the name suggest, the solution with *minimum norm* (MN) is selected from the subspace of all existing least squares solutions as

$$\mathbf{f}_{\text{MN}} = \arg \min_{\mathbf{f} \in \mathbf{f}_{\text{LS}}} \|\mathbf{f}\| = \mathbf{A}^+ \mathbf{m} \quad (3.10)$$

where $\mathbf{A}^+ = \mathbf{V}\mathbf{D}^+\mathbf{U}^T$ and

$$\mathbf{D}^+ = \text{diag}(1/d_1, \dots, 1/d_p, 0, \dots, 0) \in \mathbb{R}^{N \times M} \quad (3.11)$$

and $p = \max\{i \mid 1 \leq i \leq M, d_i > 0\}$. However, differences of magnitude between the non-zero singular values can also make the MN solutions numerically unstable.

3.4.3 Truncated singular value decomposition (TSVD)

The *truncated singular value decomposition* (TSVD) solution can be obtained as a MN solution for a system where all the singular values of \mathbf{A} that are less than a selected threshold α are set to zero.

$$\mathbf{f}_{\text{TSVD}} = \arg \min_{\mathbf{f} \in \mathbf{f}_{\text{LS}_\alpha^+}} \|\mathbf{f}\| = \mathbf{A}_\alpha^+ \mathbf{m} \quad (3.12)$$

where

$$\mathbf{f}_{\text{LS}_\alpha^+} = \arg \min_{\mathbf{f} \in \mathbb{R}^N} \|\mathbf{m} - \mathbf{A}_\alpha^+ \mathbf{f}\|^2, \quad \mathbf{A}_\alpha^+ = \mathbf{V}\mathbf{D}_\alpha^+\mathbf{U}^T \quad (3.13)$$

and

$$\mathbf{D}_\alpha^+ = \text{diag}(1/d_1, \dots, 1/d_{p_\alpha}, 0, \dots, 0) \in \mathbb{R}^{N \times M} \quad (3.14)$$

with $p_\alpha = \max\{i \mid 1 \leq i \leq \min(N, M), d_i > \alpha\}$.

With LS and MN methods a unique solution can be found, but the solutions can remain ill-conditioned. TSVD stabilises the problem by replacing the $(\min(N, M) - p_\alpha)$ smallest singular values with zeros. Consequently the method ignores the corresponding singular vectors and typically simplifies the structure of the solution. However, there is no unambiguous criterion for selecting an optimal α .

3.4.4 Tikhonov regularisation

Tikhonov regularisation is also known as *Phillips* or *Tikhonov-Phillips regularisation* and *Ridge regression* (Tikhonov and Arsenin, 1977; Phillips, 1962; Hoerl and Kennard, 1970). The method concerns both the residuals and the L^2 norm of the solution. The Tikhonov regularised solution is the minimiser

$$\mathbf{f}_T = \arg \min_{\mathbf{f} \in \mathbb{R}^N} \{ \|\mathbf{m} - \mathbf{A}\mathbf{f}\|^2 + \alpha \|\mathbf{f}\|^2 \} = \mathbf{A}_\alpha^\dagger \mathbf{m}, \quad (3.15)$$

where

$$\mathbf{A}_\alpha^\dagger = (\mathbf{A}^T \mathbf{A} + \alpha \mathbf{I})^{-1} \mathbf{A}^T = \mathbf{V} \mathbf{D}_\alpha^\dagger \mathbf{U}^T, \quad (3.16)$$

$$\mathbf{D}_\alpha^\dagger = \text{diag} \left(\frac{d_1}{d_1^2 + \alpha}, \dots, \frac{d_{\min(M,N)}}{d_{\min(M,N)}^2 + \alpha} \right) \in \mathbb{R}^{N \times M}. \quad (3.17)$$

Here α is a *regularisation parameter* that controls the balance between the residuals and the norm of the solution.

From the diagonal matrix $\mathbf{D}_\alpha^\dagger$ it is somewhat intuitive to see how Tikhonov regularisation reverts to situations of LS and MN and how α controls the ill-conditioning. Similarly to TSVD, Tikhonov regularisation can provide solutions to ill-conditioned situations where LS and MN methods fail.

The optimal selection of regularisation parameter α is again an ambiguous task, however, different selection criteria are available such as Morozov's discrepancy principle and the L-curve method (Kaipio and Somersalo, 2005; Mueller and Siltanen, 2012).

3.4.5 Generalised Tikhonov regularisation

The Tikhonov regularised solution can be generalised to situation where additional constraints are set for the solution

$$\begin{aligned} \mathbf{f}_T &= \arg \min_{\mathbf{f} \in \mathbb{R}^N} \{ \|\mathbf{A}\mathbf{f} - \mathbf{m}\|^2 + \alpha \|\mathbf{L}(\mathbf{f} - \bar{\mathbf{f}})\|^2 \} \\ &= (\mathbf{A}^T \mathbf{A} + \alpha \mathbf{L}^T \mathbf{L})^{-1} (\mathbf{A}^T \mathbf{m} + \alpha \mathbf{L}^T \mathbf{L} \bar{\mathbf{f}}) \end{aligned} \quad (3.18)$$

In Equation (3.18) the norm at the right-hand side restricts the solution close to vector $\bar{\mathbf{f}} \in \mathbb{R}^N$. Often a difference matrix is selected as $\mathbf{L} \in \mathbb{R}^{N_L \times N}$ to require smoothness for the solution.

The generalised Tikhonov regularisation can also be seen as a solution for a system where the following linear constraints are added to the original equation (3.3). The so-called stacked form is given as

$$\begin{bmatrix} \mathbf{m} \\ \sqrt{\alpha} \mathbf{L}^T \bar{\mathbf{f}} \end{bmatrix} = \begin{bmatrix} \mathbf{A} \\ \sqrt{\alpha} \mathbf{L} \end{bmatrix} \mathbf{f}_T. \quad (3.19)$$

3.5 Iterative solutions for linear system

The most widely used iterative algorithms in tomography are the algebraic reconstruction technique and the EM algorithm. In ionospheric imaging the algebraic reconstruction technique and its derivatives have been used much more frequently and are therefore presented in this chapter. For the EM algorithm see e.g. Natterer and Wübbeling (2001).

Despite the fact that in the field of image reconstruction and tomography, the following iterative techniques have been developed specifically to handle incomplete error contaminated data, they still are general solvers for exact linear systems. Hence, given the error-contaminated measurements \mathbf{m} and the matrix \mathbf{A} in Equation (3.3), these methods actually solve a system

$$\mathbf{m} = \mathbf{A}\mathbf{f}_\varepsilon. \quad (3.20)$$

However, to simplify the notations, the subindex ε has been omitted in the remainder of this chapter.

As will be stated in the sections below, the iterative techniques also provide some regularisation for the problem. In practise, the measurements predicted with iterative solutions will never equal the actual measurements with errors, and thus a stopping criterion is needed for the iteration. With incomplete data the methods are then referred to as *truncated iterative methods* (Kaipio and Somersalo, 2005), as the selection of the stopping criterion can be seen as a part of the regularisation scheme.

In the following, the notion of *iteration* refers to the update of $\mathbf{f}^{(k)}$ to obtain a new improved approximation $\mathbf{f}^{(k+1)}$. One iteration can consist of other repetitive operations. All approaches require an initial starting value for the unknown, $\mathbf{f}^{(0)}$. With incomplete data the solution can be highly dependent on the initial value.

3.5.1 Kaczmarz method

The Kaczmarz method (Kaczmarz, 1937) is a general method for iterative approximative solutions for a system of linear equations, such as Equation 3.20. Besides the original article, an intuitive illustrated description of the method is provided in Kak and Slaney (1988). Another mathematically rigorous treatment is provided by Kaipio and Somersalo (2005).

The intuition of the convergence in the approach is that each measurement i.e. single rows

$$m_j = \mathbf{a}_j^T \mathbf{f}, \quad 1 \leq j \leq M$$

define a hyperplane of dimension \mathbb{R}^{N-1} . The algorithm starts with an initial guess $\mathbf{f}^{(0)}$. The next iteration $\mathbf{f}^{(k+1)}$ is obtained by projecting the current solution $\mathbf{f}^{(k)}$ on the corresponding hyperplane. The projection for the $(k+1)^{\text{th}}$ iteration can be written as

$$\mathbf{f}^{(k+1)} = \mathbf{f}^{(k)} + \mathbf{a}_j(\mathbf{a}_j^T \mathbf{a}_j)^{-1}(\mathbf{a}_j^T \mathbf{f}^{(k)} - m_j). \quad (3.21)$$

Often a relaxation parameter $0 < \lambda < 2$ is included to control the size of the correction performed at each iteration

$$\mathbf{f}^{(k+1)} = \mathbf{f}^{(k)} + \lambda \mathbf{a}_j (\mathbf{a}_j^T \mathbf{a}_j)^{-1} (\mathbf{a}_j^T \mathbf{f}^{(k)} - m_j). \quad (3.22)$$

For the first M iterations $j = k$, but often more iterations are required for convergence and the procedure is looped over all measurement equations several times, hence $1 \leq k \leq \kappa M$, where $\kappa \in \mathbb{N}$ and $j = k(\text{mod } M) + 1$.

Another way to understand this algorithm is to see it as similar to backprojection. In Equations (3.21) and (3.22) the difference between the actual measurement m_j and the simulation of the same measurement from the current iteration $\mathbf{a}_j^T \mathbf{f}^{(k)}$ is taken. A backprojection of the difference is then added to corresponding pixels along the ray path.

If a unique solution exists for the linear system, the iterative solution of the Kaczmarz method will converge to it (Tanabe, 1971). In an overdetermined situation $M > N$, if measurement noise is present, the linear system does not have a unique solution as the hyperplanes will not have a unique intersection and the solution will not converge to one point, but will drift between the intersections (Kak and Slaney, 1988). In an underdetermined system $N > M$, where there is again no unique solution available, the algorithm will endogenously provide regularisation as it will converge to the point $\hat{\mathbf{f}}$ of possible solutions that minimises $\|\hat{\mathbf{f}} - \mathbf{f}^{(0)}\|$ i.e. the distance between that point and the given initial value (Tanabe, 1971; Kak and Slaney, 1988).

The Kaczmarz method is primarily an algorithm for solving a linear system, however it is straightforward to include some regularising prior information in it. As said above, the initial value for the unknown already provides one regularisation scheme. In many applications the unknown cannot physically have negative values, if the projection nevertheless produces negative values, the values can be detected and set to zero within the algorithm.

3.5.2 Algebraic reconstruction technique (ART)

The *algebraic reconstruction technique* (ART) was presented in the field of image reconstruction (Gordon et al., 1970). The method is the Kaczmarz method, however some specific features are sometimes included in it.

In the original article Gordon et al. (1970), as well as Kak and Slaney (1988), the weights a_{ij} are not intersection lengths, but they are simply given a value 1 or 0 depending on whether the pixel center is within the signal path with width Δs or not. This has been done to ease the computation as the in/out decision is faster than computing the precise intersection lengths. However, this shortcut is known to often give rise to so-called salt and pepper noise (Kak and Slaney, 1988). Another feature often included in ART is the non-negativity constraint.

3.5.3 Multiplicative algebraic reconstruction technique (MART)

Whereas ART converges to the least squares solution of the linear system, the *multiplicative algebraic reconstruction technique* (MART) (Gordon et al., 1970) is a modification of ART that converges to the maximum entropy solution (Censor, 1983; Raymund et al., 1990). As the name suggests, instead of additional corrections, the unknowns along each raypath are scaled by multiplying as

$$\mathbf{f}_i^{(k+1)} = \left(\frac{m_j}{\mathbf{a}_j^T \mathbf{f}^{(k)}} \right)^{\lambda_k a_{ji}} \mathbf{f}_i^{(k)}, \quad i = 1, \dots, N \quad (3.23)$$

The update formula is written for a single unknown element as the exponent includes the intersection length between the j^{th} raypath and that specific i^{th} unknown element. The relaxation parameter fulfills $0 \leq \lambda_k \leq 1$ and the initial value for the unknown is given as $\mathbf{f}^{(0)} = e^{-1} \mathbf{1}$ (Censor, 1983).

3.5.4 Simultaneous iterative reconstruction technique (SIRT)

The update caused by single measurement j in Equation (3.21) can be written as a correction required for the unknown

$$\Delta \mathbf{f}^{(k+1),j} = \mathbf{f}^{(k+1)} - \mathbf{f}^{(k)} = \mathbf{a}_j (\mathbf{a}_j^T \mathbf{a}_j)^{-1} (\mathbf{a}_j^T \mathbf{f}^{(k)} - m_j), \quad 1 \leq j \leq M. \quad (3.24)$$

The *simultaneous iterative reconstruction technique* (SIRT) is a modification of ART where the correction (3.24) is computed from each measurement without updating \mathbf{f} in between. Only after the corrections are computed for every measurement $j = 1, \dots, M$, the new iteration is obtained as

$$\mathbf{f}_i^{(k+1)} = \mathbf{f}_i^{(k)} + \frac{1}{M_i} \sum_j \Delta \mathbf{f}_i^{(k+1),j}, \quad i = 1, \dots, N,$$

where M_i is the number of measurements intersecting the corresponding unknown. The above formula is written for a single unknown element as the number M_i varies for different i . The convergence of SIRT is slower than in ART, but the quality of the reconstructed image can often be better (Kak and Slaney, 1988).

3.5.5 Simultaneous algebraic reconstruction technique (SART)

The *simultaneous algebraic reconstruction technique* (SART) (Andersen and Kak, 1984) combines some features of ART and SIRT methods. An important idea in SART is that the reconstruction can be improved with a more accurate modelling of the projections in the forward model. Hence, instead of the pixel approximation, the representation of

the unknown is generalised to a finite set of weighted base functions. In SART specifically bilinear elements are utilised as base functions. The iteration is then carried out non-sequentially with resemblance to SIRT, but in steps of individual projections. In one iteration, the corrections obtained from all measurements in one view angle are combined and used simultaneously in the update. Finally, when the corrections are applied to unknown elements along the ray paths in the projection, a Hamming window function is used to emphasise the corrections made at the middle of the ray and to damp the beginning and the end of the ray.

Chapter 4

Bayesian statistical approach

4.1 Introduction to Bayesian inference

In Bayesian statistical inference, all the variables and parameters are modelled as random variables. The randomness describes the lack of information concerning the realisations of the variables. The conclusions are based on probabilistic statements that are compiled with Bayes' formula

$$p(\mathbf{f}|\mathbf{m}) = \frac{p(\mathbf{m}|\mathbf{f})p(\mathbf{f})}{p(\mathbf{m})}, \quad (4.1)$$

where $p(\mathbf{f}|\mathbf{m})$ is the *posterior probability distribution* and $p(\mathbf{f})$ the *prior probability distribution* of \mathbf{f} , and $p(\mathbf{m}|\mathbf{f})$ is the *sampling distribution* of \mathbf{m} , but can also be seen as the *likelihood function* of \mathbf{f} given \mathbf{m} . For a fixed \mathbf{m} , the marginal distribution $p(\mathbf{m})$ is a constant and independent of \mathbf{f} , however, it can be difficult to derive from a complicated joint distribution, hence the following *unnormalised posterior distribution* is often considered instead

$$p(\mathbf{f}|\mathbf{m}) \propto p(\mathbf{m}|\mathbf{f})p(\mathbf{f}). \quad (4.2)$$

The prior distribution indicates the most likely state and the related uncertainty of the unknown parameter \mathbf{f} before the observations \mathbf{m} are made. The posterior probability distribution is obtained by updating the prior distribution with the likelihood function that connects unknown parameters with the information provided by the observations. The posterior distribution is the solution that combines all the available information on \mathbf{f} .

As high-dimensional posterior distributions can be difficult to visualise, the distribution is usually characterised with some point and spread estimates. One of the mostly used point estimates is the *maximum a posteriori* (MAP) estimate

$$\mathbf{f}_{\text{MAP}} = \arg \max_{\mathbf{f} \in \mathbb{R}^N} p(\mathbf{f}|\mathbf{m}). \quad (4.3)$$

If the maximiser for the estimator (4.3) exists, it is possible that it is not unique. Another point estimate is the *conditional mean* (CM), which is defined as

$$\mathbf{f}_{\text{CM}} = \mathbb{E}\{\mathbf{f}|\mathbf{m}\} = \int_{\mathbb{R}^N} \mathbf{f} p(\mathbf{f}|\mathbf{m}) d\mathbf{f}. \quad (4.4)$$

Conditional covariance is an estimator for the spread of the posterior distribution. It is defined as

$$\text{cov}(\mathbf{f}|\mathbf{m}) = \int_{\mathbb{R}^N} (\mathbf{f} - \mathbf{f}_{\text{CM}})(\mathbf{f} - \mathbf{f}_{\text{CM}})^{\text{T}} p(\mathbf{f}|\mathbf{m}) d\mathbf{f} \in \mathbb{R}^{N \times N}, \quad (4.5)$$

provided that the integral converges. The spread of the posterior distribution describes the remaining uncertainty of the unknown parameter. A typical illustration for the spread is to calculate probability intervals from the posterior covariance estimator.

If the true state of the unknown parameter \mathbf{f} is given a non-zero prior probability, as the sample size increases, the posterior distribution is asymptotically independent of the prior distribution and the maximum a posteriori estimate converges to the well-known *maximum likelihood* estimate

$$\mathbf{f}_{\text{ML}} = \arg \max_{\mathbf{f} \in \mathbb{R}^N} p(\mathbf{m}|\mathbf{f}). \quad (4.6)$$

Then again, if the measurements provide only little information on the parameter of interest, the posterior is dominated by the prior.

A connection to ill-posed inverse problems can be seen in situations where the maximum likelihood estimate is not identifiable, but when a prior distribution is included, a proper posterior distribution can be obtained. Especially in highly ill-posed problems, the selection of the prior can then be the most critical phase in the inference and should be done based on expert knowledge on the studied quantity. One of the advantages of the Bayesian approach for inverse problems is that the required stabilisation can be given in a very interpretative manner in terms of physical quantities and related uncertainties.

Before considering the specific linear forward model presented in Section 3.2 the Gaussian model is considered for the general variables \mathbf{f} and \mathbf{m} . By assuming that \mathbf{f} and \mathbf{m} have a joint multivariate Gaussian distribution

$$\begin{bmatrix} \mathbf{f} \\ \mathbf{m} \end{bmatrix} \sim \mathcal{N} \left(\begin{bmatrix} \bar{\mathbf{f}} \\ \bar{\mathbf{m}} \end{bmatrix}, \begin{bmatrix} \Sigma_{\mathbf{f}} & \Sigma_{\mathbf{f}\mathbf{m}} \\ \Sigma_{\mathbf{m}\mathbf{f}} & \Sigma_{\mathbf{m}} \end{bmatrix} \right), \quad (4.7)$$

with Gaussian identities the conditional distribution for \mathbf{f} given \mathbf{m} can be written as

$$p(\mathbf{f}|\mathbf{m}) \propto \exp \left(-\frac{1}{2} (\mathbf{f} - \bar{\mathbf{f}}^{(1)}) \left(\Sigma_{\mathbf{f}}^{(1)} \right)^{-1} (\mathbf{f} - \bar{\mathbf{f}}^{(1)})^{\text{T}} \right). \quad (4.8)$$

In the Gaussian system the CM and MAP estimators are the same. With Gaussian identities the MAP and posterior covariance estimators can be written as

$$\bar{\mathbf{f}}^{(1)} = \mathbf{f}_{\text{MAP}} = \mathbf{f}_{\text{CM}} = \bar{\mathbf{f}} + \boldsymbol{\Sigma}_{\mathbf{f}\mathbf{m}}\boldsymbol{\Sigma}_{\mathbf{m}}^{-1}(\mathbf{m} - \bar{\mathbf{m}}) \quad (4.9)$$

and

$$\boldsymbol{\Sigma}_{\mathbf{f}}^{(1)} = \text{cov}(\mathbf{f}|\mathbf{m}) = \boldsymbol{\Sigma}_{\mathbf{f}} - \boldsymbol{\Sigma}_{\mathbf{f}\mathbf{m}}\boldsymbol{\Sigma}_{\mathbf{m}}^{-1}\boldsymbol{\Sigma}_{\mathbf{m}\mathbf{f}}. \quad (4.10)$$

4.2 Gaussian priors for linear inverse problems

For the likelihood function, the nature of measurements and central limit theorem often justifies the use of Gaussian normal distribution. Then again, the assumption of Gaussian prior distribution is not always the most realistic choice. A downside is that the distribution cannot be easily truncated to consist only of non-negative values. The benefit of the Gaussian prior is that it is a conjugate prior for the Gaussian likelihood, resulting in a Gaussian posterior distribution with the closed form estimators (4.9); (4.10). Hence often, as is the case here, a Gaussian prior distribution is assumed.

The discretisation of the linear forward model (3.2) was briefly discussed in Section 3.2.1. However, the discretisation can be performed at different phases of the solution. Following Tarantola (1987), when modelling an inverse problem, one should first consider whether it is easier to imagine the forward problem acting on a sequence of parameters or on a field. In many geophysical applications, such as ionospheric imaging, it is indeed a natural way to conceptualise the unknown and the prior distribution as a continuous field. In this section Gaussian random fields (Tarantola, 1987; Christakos, 2005; Rasmussen and Williams, 2006) will be utilised.

Definition 4.1. Given a probability space (Ω, \mathcal{F}, P) , a real-valued d -dimensional *spatial random field* (RF) $f(z) := f(z, \omega)$, $z \in \mathbb{R}^d, \omega \in \Omega$ is a family of random variables $\{f(z_1), f(z_2), \dots\}$ at points z_1, z_2, \dots , where each random variable is real valued and defined on (Ω, \mathcal{F}, P) .

A spatial RF is a generalisation of a stochastic process. Where a stochastic process is often seen as indexed by points in time, a spatial RF is indexed by d -dimensional Euclidean space, where d is typically two or three.

Definition 4.2. A multivariate *Gaussian random field* (GRF) is a spatial random field where any finite number of random variables have a joint Gaussian distribution.

GRF is then completely specified by its mean and covariance functions and denoted here as

$$f(z) \sim \mathcal{GRF}(\bar{f}(z), K(z, z')), \quad (4.11)$$

where

$$\begin{aligned}\bar{f}(z) &= \mathbb{E}[f(z)] \\ K(z, z') &= \mathbb{E}[(f(z) - \bar{f}(z))(f(z') - \bar{f}(z'))^T].\end{aligned}\tag{4.12}$$

The GRF (4.11) is used here as the prior for obtaining all of the following posterior estimators in this chapter. For a finite set of points $\mathbf{z} \in \mathbb{R}^{N \times d}$, as denoted in Equation (2.5), a GRF is simply a multivariate Gaussian normal distribution

$$\mathbf{f} \sim \mathcal{N}(\bar{\mathbf{f}}, \mathbf{\Sigma}_{\mathbf{f}}),\tag{4.13}$$

where $\bar{\mathbf{f}} \in \mathbb{R}^N$ and $\mathbf{\Sigma}_{\mathbf{f}} = K(\mathbf{z}, \mathbf{z}) \in \mathbb{R}^{N \times N}$.

4.2.1 Continuous Gaussian random field prior

GRFs are closed under linear operations. Hence, when considering a linear problem (3.2) with GRF prior (4.11), the covariance and cross-covariances can be written as

$$\begin{aligned}\mathbb{E}[(\mathbf{m} - \bar{\mathbf{m}})(\mathbf{m} - \bar{\mathbf{m}})^T] &= \mathbb{E}[(A_z f(z) + \boldsymbol{\varepsilon} - A_z \bar{f}(z))(A_{z'} f(z') + \boldsymbol{\varepsilon} - A_{z'} \bar{f}(z'))^T] \\ &= \mathbb{E}[(A_z(f(z) - \bar{f}(z)))(A_{z'}(f(z') - \bar{f}(z')))^T] + \mathbb{E}[\boldsymbol{\varepsilon} \boldsymbol{\varepsilon}^T] \\ &= A_z K(z, z') A_{z'}^T + \mathbf{\Sigma}_{\boldsymbol{\varepsilon}} = \mathbf{K}_{A_f A_f} + \mathbf{\Sigma}_{\boldsymbol{\varepsilon}},\end{aligned}\tag{4.14}$$

$$\begin{aligned}\mathbb{E}[(\mathbf{m} - \bar{\mathbf{m}})(f(z') - \bar{f}(z'))^T] &= \mathbb{E}[(A_z f(z) - A_z \bar{f}(z))(f(z') - \bar{f}(z'))^T] = A_z K(z, z') \\ \mathbb{E}[(f(z) - \bar{f}(z))(\mathbf{m} - \bar{\mathbf{m}})^T] &= \mathbb{E}[(f(z) - \bar{f}(z))(A_{z'} f(z') - A_{z'} \bar{f}(z'))^T] = K(z, z') A_{z'}^T,\end{aligned}\tag{4.15}$$

where $\bar{\mathbf{m}} = A_z \bar{f}(z)$, $f \perp \boldsymbol{\varepsilon}$ and the subscript i in operator A_i indicates the parameter in the covariance function that it acts upon. The transpose is defined as $(A_z K(z, z'))^T = K(z, z') A_{z'}^T$.

The interest can now be in arbitrary locations $\mathbf{z} \in \mathbb{R}^{N^* \times d}$, which does not need to be the full lattice and thus $N^* \leq N$. Then, given the measurements \mathbf{m} and a GRF prior (4.11) with mean and covariance as above (4.12), the covariance is a matrix

$$K(\mathbf{z}, \mathbf{z}) = \mathbf{K}_{ff} \in \mathbb{R}^{N^* \times N^*}$$

and the covariance matrices between the measurements and the unknowns

$$A_z K(z, \mathbf{z}) = \mathbf{K}_{A_f f} = (\mathbf{K}_{f A_f})^T = (K(\mathbf{z}, z) A_z^T)^T \in \mathbb{R}^{M \times N^*}.$$

The posterior distribution for the unknown field at locations \mathbf{z} is then given with Equation (4.8), where

$$\bar{\mathbf{f}}^{(1)} = \bar{f}(\mathbf{z}) + \mathbf{K}_{f A_f} (\mathbf{K}_{A_f A_f} + \mathbf{\Sigma}_{\boldsymbol{\varepsilon}})^{-1} (\mathbf{m} - \bar{\mathbf{m}}) \in \mathbb{R}^{N^*}\tag{4.16}$$

$$\Sigma_f^{(1)} = \mathbf{K}_{ff} - \mathbf{K}_{fAf} (\mathbf{K}_{AfAf} + \Sigma_\epsilon)^{-1} \mathbf{K}_{Aff} \in \mathbb{R}^{N^* \times N^*}. \quad (4.17)$$

If the linear transformations with A_z in Equations (4.14) and (4.15) can be solved analytically without discretising f , the discretisation of the covariance kernel leading to a covariance matrix for the complete discretised domain can be avoided in the MAP estimator. In literature this approach is often referred to as *Kriging* or *Gaussian process* (Rasmussen and Williams, 2006). Especially in these approaches a parameterised covariance kernel is chosen and the parameters are estimated from the data. Gaussian processes with linear operations are discussed by Särkkä (2011) and Minkwitz et al. (2015).

4.2.2 Discrete multivariate Gaussian prior

Here the Bayesian approach is applied to a discretised linear system of Equation (3.3) on a lattice $\mathbf{z} \in \mathbb{R}^{N \times d}$, such as given in Equation (2.5). Hence, the multivariate normal distribution prior given in Equation (4.13) is used. The covariances and cross-covariances between the variables are then

$$\begin{aligned} \mathbb{E}[(\mathbf{m} - \bar{\mathbf{m}})(\mathbf{m} - \bar{\mathbf{m}})^T] &= \mathbb{E}[(\mathbf{A}\mathbf{f} + \epsilon - \mathbf{A}\bar{\mathbf{f}})(\mathbf{A}\mathbf{f} + \epsilon - \mathbf{A}\bar{\mathbf{f}})^T] \\ &= \mathbb{E}[(\mathbf{A}\mathbf{f} - \mathbf{A}\bar{\mathbf{f}})(\mathbf{A}\mathbf{f} - \mathbf{A}\bar{\mathbf{f}})^T] + \mathbb{E}[\epsilon\epsilon^T] \\ &= \mathbf{A}\Sigma_f\mathbf{A}^T + \Sigma_\epsilon \in \mathbb{R}^{M \times M} \end{aligned} \quad (4.18)$$

and

$$\begin{aligned} \mathbb{E}[(\mathbf{m} - \bar{\mathbf{m}})(\mathbf{f} - \bar{\mathbf{f}})^T] &= \mathbb{E}[(\mathbf{A}\mathbf{f} - \mathbf{A}\bar{\mathbf{f}})(\mathbf{f} - \bar{\mathbf{f}})^T] = \mathbf{A}\Sigma_f \in \mathbb{R}^{M \times N} \\ \mathbb{E}[(\mathbf{f} - \bar{\mathbf{f}})(\mathbf{m} - \bar{\mathbf{m}})^T] &= \mathbb{E}[(\mathbf{f} - \bar{\mathbf{f}})(\mathbf{A}\mathbf{f} - \mathbf{A}\bar{\mathbf{f}})^T] = \Sigma_f\mathbf{A}^T \in \mathbb{R}^{N \times M}, \end{aligned} \quad (4.19)$$

where $\bar{\mathbf{m}} = \mathbf{A}\bar{\mathbf{f}}$. The posterior distribution is then again of the form given in Equation (4.8), with

$$\bar{\mathbf{f}}^{(1)} = \bar{\mathbf{f}} + \Sigma_f\mathbf{A}^T (\mathbf{A}\Sigma_f\mathbf{A}^T + \Sigma_\epsilon)^{-1} (\mathbf{m} - \bar{\mathbf{m}}) \in \mathbb{R}^N \quad (4.20)$$

$$\Sigma_f^{(1)} = \Sigma_f - \Sigma_f\mathbf{A}^T (\mathbf{A}\Sigma_f\mathbf{A}^T + \Sigma_\epsilon)^{-1} \mathbf{A}\Sigma_f \in \mathbb{R}^{N \times N}. \quad (4.21)$$

Although it is now necessary to form the $N \times N$ prior covariance matrix, the matrix inversions in estimators (4.20) and (4.21) take place in the $M \times M$ measurement space. Hence, if $M \leq N$, the numerical computation is generally easier than in the following model space solution. However, with large N , the prior covariance matrix can become excessively large even for numerical storage.

4.2.3 Model space solution

For the Gaussian linear case, the MAP estimator (4.9) and posterior covariance (4.10) can be derived with well-known Gaussian identities from their joint distribution (4.7). By deriving the posterior distribution directly from the Bayes' formula (4.2) for the linear model (3.3), the quadratic form can also be arranged to provide the following equivalent forms for the estimators

$$\bar{\mathbf{f}}^{(1)} = \boldsymbol{\Sigma}_{\mathbf{f}}^{(1)} \left(\mathbf{A}^T \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{-1} \mathbf{m} + \boldsymbol{\Sigma}_{\mathbf{f}}^{-1} \bar{\mathbf{f}} \right) \in \mathbb{R}^N \quad (4.22)$$

$$\boldsymbol{\Sigma}_{\mathbf{f}}^{(1)} = \left(\mathbf{A}^T \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{-1} \mathbf{A} + \boldsymbol{\Sigma}_{\mathbf{f}}^{-1} \right)^{-1} \in \mathbb{R}^{N \times N}. \quad (4.23)$$

The above estimators can also be derived from Equations (4.20) and (4.21) with the matrix inversion lemma also known as Woodbury matrix identity or Sherman-Morrison-Woodbury formula (Golub and Van Loan, 2013). Since it is now necessary to invert two $N \times N$ matrices, this would generally be the preferred approach only in situations where $N \ll M$. However, if the prior information can be given directly as inverse covariance the situation can change, as it is demonstrated in the following section.

4.3 Gaussian Markov random field (GMRF) priors

The sparsity of a matrix signifies the large proportion of strict zeros in matrix elements. *Sparse linear system* then refers to a linear system that is so large and sparse that it is beneficial to rethink the standard factorising methods of two-dimensional arrays (Golub and Van Loan, 2013). For a *sparse matrix* it requires significantly less memory to index the non-zero matrix elements as vectors and to use operations designed for such systems.

When solving the posterior estimates given in previous sections, the measurement error covariance $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}$ is typically assumed as a diagonal matrix. If the theory matrix \mathbf{A} is sparse, as is the case here (2.4), the main concern is the prior covariance matrix $\boldsymbol{\Sigma}_{\mathbf{f}}$. A proper prior distribution with a non-diagonal covariance structure results in a dense $N \times N$ covariance matrix.

A *Gaussian Markov random field* (GMRF) is a multivariate Gaussian distribution Satisfying the Markov property. In GMRF, the Markov property indicates that an element conditioned with its neighbouring elements is independent of the rest of the elements in the field. The independence between elements is equivalent to the precision between the elements being zero. Typically GMRFs are used in a situations where the neighbourhood does not include the complete field, hence the precision matrix is characteristically a sparse matrix. A comprehensive introduction to GMRF is provided by Rue and Held (2005).

Definition 4.3. Neighbourhood N_i to \mathbf{f}_i is the set $\{\mathbf{f}_j, j \in N_i \mid \|\mathbf{z}_i - \mathbf{z}_j\| \leq r, j \neq i\}$, where radius $r > 0$.

Definition 4.4. A random vector $\mathbf{f} \in \mathbb{R}^N$ is called GMRF with respect to neighbourhood N_i with mean $\bar{\mathbf{f}}$ and precision matrix $\mathbf{Q}_f > 0$ if and only if its density has the form

$$\mathbf{f} \sim (2\pi)^{-n/2} |\mathbf{Q}_f|^{1/2} \exp \left(-\frac{1}{2} (\mathbf{f} - \bar{\mathbf{f}})^T \mathbf{Q}_f (\mathbf{f} - \bar{\mathbf{f}}) \right)$$

and

$$[\mathbf{Q}_f]_{i,j} \neq 0 \iff j \in N_i \quad \forall \quad i \neq j.$$

With GMRF prior, the model space estimators in Equations (4.22) and (4.23) can then be written by replacing $\mathbf{Q}_f = \Sigma_f^{-1} \in \mathbb{R}^{N \times N}$, resulting in

$$\bar{\mathbf{f}}^{(1)} = \Sigma_f^{(1)} (\mathbf{A}^T \Sigma_\epsilon^{-1} \mathbf{m} + \mathbf{Q}_f \bar{\mathbf{f}}) \in \mathbb{R}^N \quad (4.24)$$

$$\Sigma_f^{(1)} = (\mathbf{A}^T \Sigma_\epsilon^{-1} \mathbf{A} + \mathbf{Q}_f)^{-1} \in \mathbb{R}^{N \times N}. \quad (4.25)$$

It is not required that the precision matrix \mathbf{Q}_f is invertible. A GMRF with a symmetric positive semidefinite precision matrix is called *intrinsic* GMRF, which results in improper prior density for \mathbf{f} . A useful intrinsic GMRF, and an improper prior distribution, can be constructed by selecting $\mathbf{Q}_f = \beta \mathbf{L}^T \mathbf{L}$, where \mathbf{L} is a difference matrix and β a weight parameter. Such a difference prior promotes smoothness, but is invariant to an addition of a constant. If no boundary conditions are added, $\text{Ker}(\mathbf{L}) \neq \{0\}$. Then again, if $\text{Ker}(\mathbf{A}) \cap \text{Ker}(\mathbf{L}) = \{0\}$, the resulting Gaussian posterior distribution defines a proper probability density, with mean and covariance given in Equations (4.24) and (4.25) above. Further, by assuming $\Sigma_\epsilon = \sigma^2 \mathbf{I}$ and denoting $\alpha := \sigma^2 \beta$, the resulting MAP estimator (4.24) is the generalised Tikhonov regularised solution given in Equation (3.18).

4.3.1 Correlation priors

The strength of a proper Gaussian prior is in the covariance where the provided information is easily interpretable in a probabilistic and physical sense. The downside is that it results in a full covariance matrix, making the storage and computation problematic when the number of unknowns is large.

GMRF priors are convenient when the ill-posed problem can be stabilised with moderate smoothing. In these cases the use of an intrinsic GMRF is straightforward, the interpretation in a mechanical sense is clear and the sparsity of a difference matrix \mathbf{L} allows computations for much higher dimensional problems than working with a full prior covariance matrix Σ_f . Even so, if the problem is severely ill-conditioned and requires stronger stabilisation, the implementation of more strict constraints and boundary conditions in the precision matrix can get complicated, the effect of the possibly overlapping constraints unpredictable, and the physical interpretation difficult.

If sufficient boundary conditions are included into a difference matrix \mathbf{L} , the precision $\mathbf{L}^T \mathbf{L} = \mathbf{Q}_f$ becomes invertible and $(\mathbf{L}^T \mathbf{L})^{-1} = \mathbf{Q}_f^{-1} = \mathbf{\Sigma}_f$. However, for efficient computation it would be profitable to work with the sparse \mathbf{L} matrix while knowing the covariance structure in $\mathbf{\Sigma}_f$ without solving it.

So-called *correlation priors* were introduced in Roininen et al. (2011, 2013). Similarly to the GRF case, the starting point for building a correlation prior is the selection of a continuous prior covariance function

$$K(z, z') = \text{Cov}(z, z', \alpha, \ell, \mathbf{c}), \quad (4.26)$$

where the covariance function between points z and z' is parametrised with variance scaling parameter α , correlation length parameters ℓ and shape parameters \mathbf{c} . In the aforementioned articles it is shown that certain classes of covariance functions can be represented as solutions for systems of stochastic partial differential equations and that these systems can be approximated discretely with combinations of difference matrices. The matrices are formed with differences weighted with α, ℓ and \mathbf{c} parameters inherited from the original covariance function, and with a discretisation length parameter h . The solution for the discrete system is a multivariate normal distribution with covariance matrix $(\mathbf{L}_C^T \mathbf{L}_C)^{-1}$, where \mathbf{L}_C contains the required weighted difference matrices in a stacked form. The resulting covariance is discretisation independent, which means that the obtained discrete covariance converges to continuous covariance at the discretisation limit. The approach then provides a scheme to write the posterior estimators (4.24) and (4.25) for a known prior covariance function (4.26) with a precision matrix $\mathbf{Q}_f = \mathbf{L}_C^T \mathbf{L}_C$.

Chapter 5

Spatiotemporal evolution

The solutions for the linear inverse problem presented in Chapters 3 and 4 have considered individual snapshots, where the unknown is assumed to be static in time and the measurements observed all at once. In this chapter, the state of the system is solved for sequential time steps. The first intuition would be to use any of the presented methods sequentially for different states. Another approach would be to solve the problem for one state and then use the obtained solution, depending on the method, as an initial guess or a prior for the next. Hence, the Bayesian approach provides a very natural way for dealing with temporally dynamic systems. For a more comprehensive treatment of Bayesian filtering and smoothing see e.g. Särkkä (2013).

5.1 Recursive linear estimation

The measurement model in Equation (3.3) is written for the time step l as

$$\mathbf{m}^{(l)} = \mathbf{A}^{(l)} \mathbf{f} + \boldsymbol{\varepsilon}^{(l)}, \quad (5.1)$$

where $\boldsymbol{\varepsilon}^{(l)} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{(l)})$ and $l = 0, 1, 2, \dots$. It is assumed that the posterior estimates $\bar{\mathbf{f}}^{(l-1)}$ and $\boldsymbol{\Sigma}_{\mathbf{f}}^{(l-1)}$ for time $l-1$ are available. The posterior distribution can then be used as a prior for the estimates of the following state l , and the discrete measurement space estimators (4.20) and (4.21) are

$$\bar{\mathbf{f}}^{(l)} = \bar{\mathbf{f}}^{(l-1)} + \boldsymbol{\Sigma}_{\mathbf{f}}^{(l-1)} (\mathbf{A}^{(l)})^T \left(\mathbf{A}^{(l)} \boldsymbol{\Sigma}_{\mathbf{f}}^{(l-1)} (\mathbf{A}^{(l)})^T + \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{(l)} \right)^{-1} \left(\mathbf{m}^{(l)} - \mathbf{A}^{(l)} \bar{\mathbf{f}}^{(l-1)} \right) \quad (5.2)$$

$$\boldsymbol{\Sigma}_{\mathbf{f}}^{(l)} = \boldsymbol{\Sigma}_{\mathbf{f}}^{(l-1)} - \boldsymbol{\Sigma}_{\mathbf{f}}^{(l-1)} (\mathbf{A}^{(l)})^T \left(\mathbf{A}^{(l)} \boldsymbol{\Sigma}_{\mathbf{f}}^{(l-1)} (\mathbf{A}^{(l)})^T + \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{(l)} \right)^{-1} \mathbf{A}^{(l)} \boldsymbol{\Sigma}_{\mathbf{f}}^{(l-1)}. \quad (5.3)$$

Often temporary variables, innovation \mathbf{S} and Kalman gain \mathbf{G} , are used to write the above in the following steps

$$\begin{aligned}
\mathbf{S}^{(l)} &= \mathbf{A}^{(l)} \boldsymbol{\Sigma}_{\mathbf{f}}^{(l-1)} (\mathbf{A}^{(l)})^T + \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{(l)} \\
\mathbf{G}^{(l)} &= \boldsymbol{\Sigma}_{\mathbf{f}}^{(l-1)} (\mathbf{A}^{(l)})^T (\mathbf{S}^{(l)})^{-1} \\
\bar{\mathbf{f}}^{(l)} &= \bar{\mathbf{f}}^{(l-1)} + \mathbf{G}^{(l)} \left(\mathbf{m}^{(l)} - \mathbf{A}^{(l)} \bar{\mathbf{f}}^{(l-1)} \right) \\
\boldsymbol{\Sigma}_{\mathbf{f}}^{(l)} &= \boldsymbol{\Sigma}_{\mathbf{f}}^{(l-1)} - \mathbf{G}^{(l)} \mathbf{S}^{(l)} (\mathbf{G}^{(l)})^T.
\end{aligned} \tag{5.4}$$

The recursive algorithm allows online updating when new information becomes available. However, the algorithm is based on an assumption that the unknown \mathbf{f} is constant and each update is accumulating information from the same state. Hence, the l^{th} solution could be obtained also by using all measurements at once.

5.2 Kalman filtering

The more general *Bayesian filtering* is restricted here to a linear Gaussian case, when the algorithm is known more famously as *Kalman filter* (Kalman, 1960). Here the unknown $\mathbf{f}^{(l)}$ is assumed to evolve in time with states $l = 1, 2, \dots$ and that its dynamics can be modelled with a probabilistic state space model. In practise, in comparison with recursive linear filtering above, this means mainly the addition of *linear dynamic model*

$$\mathbf{f}^{(l)} = \mathbf{H}^{(l-1)} \mathbf{f}^{(l-1)} + \mathbf{e}^{(l-1)}, \tag{5.5}$$

where the stochastic dynamics are modelled with a transition matrix $\mathbf{H}^{(l-1)}$ acting on the previous state of the system and with process noise $\mathbf{e}^{(l-1)} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{e}}^{(l-1)})$. The measurement model of Equation (5.1) is then written for each state as

$$\mathbf{m}^{(l)} = \mathbf{A}^{(l)} \mathbf{f}^{(l)} + \boldsymbol{\varepsilon}^{(l)}. \tag{5.6}$$

Hence, intuitively in Bayesian inference for a dynamical system, the best guess for the present state is not given by the previous posterior distribution, but by their mappings with the transition matrix, resulting in the predictive distribution

$$\mathbf{f}^{(l)} | \mathbf{m}^{(1:l-1)} \sim \mathcal{N}(\hat{\mathbf{f}}^{(l)}, \hat{\boldsymbol{\Sigma}}_{\mathbf{f}}^{(l)}), \tag{5.7}$$

where the mean and covariance are defined below in Equation (5.8).

Now the solution for state l is the posterior distribution where the predictive distribution is used as the prior and the likelihood constructed from the current measurement model and measurements. Traditionally, the estimators are again separated in steps with the new variables:

Prediction step

$$\begin{aligned}\hat{\mathbf{f}}^{(l)} &= \mathbf{H}^{(l-1)} \bar{\mathbf{f}}^{(l-1)} \\ \hat{\Sigma}_{\mathbf{f}}^{(l)} &= \mathbf{H}^{(l-1)} \Sigma_{\mathbf{f}}^{(l-1)} (\mathbf{H}^{(l-1)})^T + \Sigma_{\varepsilon}^{(l-1)}\end{aligned}\tag{5.8}$$

Update step

$$\begin{aligned}\mathbf{v}^{(l)} &= \mathbf{m}^{(l)} - \mathbf{A}^{(l)} \hat{\mathbf{f}}^{(l)} \\ \mathbf{S}^{(l)} &= \mathbf{A}^{(l)} \hat{\Sigma}_{\mathbf{f}}^{(l)} (\mathbf{A}^{(l)})^T + \Sigma_{\varepsilon}^{(l)} \\ \mathbf{G}^{(l)} &= \hat{\Sigma}_{\mathbf{f}}^{(l)} (\mathbf{A}^{(l)})^T (\mathbf{S}^{(l)})^{-1} \\ \bar{\mathbf{f}}^{(l)} &= \hat{\mathbf{f}}^{(l)} + \mathbf{G}^{(l)} \mathbf{v}^{(l)} \\ \Sigma_{\mathbf{f}}^{(l)} &= \hat{\Sigma}_{\mathbf{f}}^{(l)} - \mathbf{G}^{(l)} \mathbf{S}^{(l)} (\mathbf{G}^{(l)})^T.\end{aligned}\tag{5.9}$$

5.3 Kalman smoothing

In Kalman filtering the earlier and current measurements are used to compute the best possible estimate for the current state of the system. However, when applications are not run online, a complete data set $\mathbf{m}^{(l)}$ for each $l = 1, \dots, T$ might be available. If the interest is not in the last state, but in the whole process, with Bayesian smoothing it is possible to take into account also the future states of the system, while evaluating state l .

The Gaussian linear version of the *Bayesian smoother* is known also as *Rauch-Tung-Striebel smoother* and *Kalman smoother*. In its standard form the smoother algorithm is divided in forward and backward passes. In forward pass, the data is filtered with steps (5.8) and (5.9) and the results are saved. The filtering results are then used in the *backward pass* with steps

$$\begin{aligned}\mathbf{C}^{(l)} &= \Sigma_{\mathbf{f}}^{(l)} (\mathbf{H}^{(l)})^T (\hat{\Sigma}_{\mathbf{f}}^{(l+1)})^{-1} \\ \tilde{\mathbf{f}}^{(l)} &= \bar{\mathbf{f}}^{(l)} + \mathbf{C}^{(l)} (\tilde{\mathbf{f}}^{(l+1)} - \hat{\mathbf{f}}^{(l+1)}) \\ \tilde{\Sigma}_{\mathbf{f}}^{(l)} &= \Sigma_{\mathbf{f}}^{(l)} + \mathbf{C}^{(l)} (\tilde{\Sigma}_{\mathbf{f}}^{(l+1)} - \hat{\Sigma}_{\mathbf{f}}^{(l+1)}) (\mathbf{C}^{(l)})^T,\end{aligned}\tag{5.10}$$

where $\hat{\mathbf{f}}^{(l+1)}$, $\hat{\Sigma}_{\mathbf{f}}^{(l+1)}$, $\bar{\mathbf{f}}^{(l)}$ and $\Sigma_{\mathbf{f}}^{(l)}$ are the predicted and filtered solutions from the forward pass. The backward pass is started from the state T , with $\tilde{\mathbf{f}}^{(T)} = \bar{\mathbf{f}}^{(T)}$ and $\tilde{\Sigma}_{\mathbf{f}}^{(T)} = \Sigma_{\mathbf{f}}^{(T)}$.

5.4 Ensemble Kalman filter (EnKF)

When increasing the spatial resolution in a three-dimensional model the number of unknown variables N increases rapidly and the $N \times N$ dimensional covariance matrices become

infeasible to handle in the above filtering and smoothing algorithms. In *ensemble Kalman filter* (EnKF) (Evensen, 1994, 2003, 2009) the maintenance of large covariance matrices is eased by not solving the updated posterior covariance in Equation (5.9) directly. Instead, samples from each posterior distribution are simulated and the covariance information is carried within the sample. The EnKF belongs to a wider category of particle filters and it was developed mainly for nonlinear problems. Here the main idea is presented in a linear setting.

First an initial prior *ensemble*

$$\mathbf{F}^{(0)} = [\mathbf{f}_1^{(0)}, \dots, \mathbf{f}_{N_{\text{ens}}}^{(0)}] \in \mathbb{R}^{N \times N_{\text{ens}}} \quad (5.11)$$

is generated, where the number of ensemble members $N_{\text{ens}} \ll N$.

At step l an *ensemble of observations*

$$\mathbf{M}^{(l)} = [\mathbf{m}_1^{(l)}, \dots, \mathbf{m}_{N_{\text{ens}}}^{(l)}] \in \mathbb{R}^{M \times N_{\text{ens}}}, \quad (5.12)$$

is simulated as $\mathbf{m}_i^{(l)} = \mathbf{m}^{(l)} + \boldsymbol{\varepsilon}_i$, where $\boldsymbol{\varepsilon}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}})$. The *predicted ensemble* is then obtained as

EnKF prediction step

$$\hat{\mathbf{F}}^{(l)} = \mathbf{H}^{(l-1)} \mathbf{F}^{(l-1)} \in \mathbb{R}^{N \times N_{\text{ens}}}. \quad (5.13)$$

The predicted ensemble mean is then

$$\hat{\mathbf{f}}_{\text{ens}}^{(l)} = \frac{1}{N_{\text{ens}}} \hat{\mathbf{F}}^{(l)} \mathbf{1}_{N_{\text{ens}} \times 1} \in \mathbb{R}^N \quad (5.14)$$

and the corresponding sample covariance

$$\begin{aligned} \hat{\boldsymbol{\Sigma}}_{\mathbf{f}_{\text{ens}}}^{(l)} &= \frac{1}{N_{\text{ens}} - 1} \left(\hat{\mathbf{F}}^{(l)} - \hat{\mathbf{f}}_{\text{ens}}^{(l)} \mathbf{1}_{1 \times N_{\text{ens}}} \right) \left(\hat{\mathbf{F}}^{(l)} - \hat{\mathbf{f}}_{\text{ens}}^{(l)} \mathbf{1}_{1 \times N_{\text{ens}}} \right)^{\text{T}} \\ &= \boldsymbol{\Gamma}_{\mathbf{f}_{\text{ens}}}^{(l)} (\boldsymbol{\Gamma}_{\mathbf{f}_{\text{ens}}}^{(l)})^{\text{T}}, \end{aligned} \quad (5.15)$$

where $\boldsymbol{\Gamma}_{\mathbf{f}_{\text{ens}}}^{(l)} \in \mathbb{R}^{N \times N_{\text{ens}}}$. The *posterior ensemble* matrix $\mathbf{F}^{(l)}$ is then obtained with EnKF update step.

EnKF update step

$$\begin{aligned} \mathbf{V}^{(l)} &= \mathbf{M}^{(l)} - \mathbf{A}^{(l)} \hat{\mathbf{F}}^{(l)} \\ \mathbf{S}^{(l)} &= \left(\mathbf{A}^{(l)} \boldsymbol{\Gamma}_{\mathbf{f}_{\text{ens}}}^{(l)} \right) \left((\boldsymbol{\Gamma}_{\mathbf{f}_{\text{ens}}}^{(l)})^{\text{T}} (\mathbf{A}^{(l)})^{\text{T}} \right) + \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{(l)} \\ \mathbf{G}^{(l)} &= \boldsymbol{\Gamma}_{\mathbf{f}_{\text{ens}}}^{(l)} \left((\boldsymbol{\Gamma}_{\mathbf{f}_{\text{ens}}}^{(l)})^{\text{T}} (\mathbf{A}^{(l)})^{\text{T}} \right) (\mathbf{S}^{(l)})^{-1} \\ \mathbf{F}^{(l)} &= \hat{\mathbf{F}}^{(l)} + \mathbf{G}^{(l)} \mathbf{V}^{(l)}, \end{aligned} \quad (5.16)$$

wherein the order of operations can be selected such that no $N \times N$ matrices are formed at any stage. The ensemble carries correct error statistics for deriving the ensemble mean and covariance, as with increasing ensemble size, the linear EnKF solution converges exactly to the Kalman filter solution (Evensen, 2009).

Chapter 6

Ionospheric measurements

In this chapter the physical background of the mostly used measurements in ionospheric imaging are presented. The chapter begins with an introduction to general electromagnetic wave propagation as it provides the basis for the modelling of radio measurements of satellite-transmitted signals and ionosonde measurements. As the ground-based measurements of *very-high frequency* (VHF, 30–300 MHz) and *ultra-high frequency* (UHF, 300–3000 MHz) satellite signals are the most important data component and the understanding of the related measurement errors and biases is essential, the background of these measurements will be examined in more detail. Ionosonde, incoherent scatter radar and satellite in situ measurements provide precise information on ionospheric electron density; however, as the spatial coverage of these measurements is local-scale, they can be used only as additional or validation data and are reviewed here more briefly.

6.1 Electromagnetic wave propagation

An electromagnetic wave propagating along axis $z \in \mathbb{R}^+$ in temporally and spatially homogeneous medium can be represented as

$$\psi(t, z) = E_0 \cos(\omega t - kz) = E_0 \cos\left(\omega t - \frac{\omega}{c}nz\right), \quad (6.1)$$

where E_0 is the peak amplitude, t is the time, ω is the angular frequency, $k = \omega/v$ is the wavenumber and c and v are the velocities of an electromagnetic wave in vacuum and in the medium correspondingly. Finally the refractive index n describes the velocity of the electromagnetic wave in the medium. It is defined as

$$n = \frac{c}{v}. \quad (6.2)$$

6.1.1 Ionospheric refractive index

The complex *refractive index* of magnetised plasma containing free electrons is given by the Appleton-Lassen or the Appleton-Hartree formula (Budden, 1961)

$$n^2 = 1 - \frac{X}{1 - iZ + \frac{\frac{1}{2}Y^2 \sin^2 \theta}{(1-X-iZ)} \pm \sqrt{\frac{\frac{1}{4}Y^4 \sin^4 \theta}{(1-X-iZ)^2} - Y^2 \cos^2 \theta}}, \quad (6.3)$$

where $i = \sqrt{-1}$ is the imaginary number, $X = \frac{\omega_p^2}{\omega^2}$, $Y = \frac{\omega_H}{\omega}$, $Z = \frac{\nu}{\omega}$, ω_p is the angular plasma frequency, ν is the electron collision frequency, θ is the angle between wave normal and inclination of magnetic field and ω_H is the electron gyrofrequency.

Plasma frequency

The angular *plasma frequency* can be written as

$$\omega_p = \sqrt{\frac{N_e e^2}{\epsilon_0 m}}, \quad (6.4)$$

where N_e is the electron density, e is the electron charge, m is the electron mass and ϵ_0 is the permittivity of free space. Useful conversions between electron density and (temporal) plasma frequency $f_p = \frac{\omega_p}{2\pi}$ can then be obtained by inserting the natural constants to Equation (6.4) and solving

$$\begin{aligned} f_p &\approx 8.98 \times \sqrt{N_e} \quad (\text{Hz}) \quad \text{and} \\ N_e &\approx 0.012 \times f_p^2 \quad (1/\text{m}^3). \end{aligned} \quad (6.5)$$

Collision frequency

The complex part in the Appleton-Lassen formula (6.3) is related to absorption. The absorption results when charged particles that oscillate along with the electromagnetic wave collide with other, mainly neutral particles. The collisions then decrease the energy of the radiation. At mid latitudes, the effective collision frequency is less than 10^4 Hz (Fehmers, 1996). The maximum collision frequency measured in an example case in Tromsø 1991 is around 10^6 Hz at an altitude of 100 km, from where it decreases rapidly as the altitude increases (Brekke, 1997). As can be seen in the Appleton-Lassen formula (6.3), part Z containing collisions decreases with increasing signal frequency. When the frequency of the propagating electromagnetic wave is greater than about 1 MHz the electron collisions can be neglected and Z can be approximated with zero (Budden, 1961; Davies, 1965).

Gyrofrequency and magnetic field

Gyrofrequency can be written as

$$\omega_H = \frac{B_0 |e|}{m}, \quad (6.6)$$

where B_0 is the magnetic field strength. It is the angular frequency of a charged particle, here electron, circling around a uniform magnetic field in a direction perpendicular to the field. In a case of magnetised plasma, the plusminus sign in Appleton-Lassen formula (6.3) indicates how the propagating wave is split in two modes. The mode with the “+” sign is called *ordinary* and the mode with the “−” sign *extraordinary* component.

A general typical value for gyrofrequency is around 1 MHz (Budden, 1961) and 1.5 MHz (Parkinson et al., 1996). Also Y in Appleton-Lassen formula (6.3), that contains the gyrofrequency, decreases with increasing signal frequency.

6.1.2 Group refractive index

As can be seen in Appleton-Lassen formula (6.3), the ionospheric refraction depends on the frequency of the propagating signal, i.e. the ionosphere is a *dispersive* medium. When considering a modulated signal the velocities of the signal carrier phase and modulation envelope will differ due to the dispersion. Following Davies (1965, 1990), to demonstrate this effect, two electromagnetic waves propagating along one-dimensional axis $z \in \mathbb{R}^+$ are considered

$$\begin{aligned} \psi_1(t, z) &= E_0 \cos(\omega t - kz) \\ \psi_2(t, z) &= E_0 \cos((\omega + \Delta\omega)t - (k + \Delta k)z), \end{aligned} \quad (6.7)$$

where $\Delta\omega$ and Δk are the small differences in angular frequency and wavenumber between signals ψ_1 and ψ_2 . An amplitude-modulated signal can be created by summation and then writing with trigonometric identities as

$$\begin{aligned} \psi_1(t, z) + \psi_2(t, z) &= 2E_0 \cos\left(\frac{\Delta\omega}{2}t - \frac{\Delta k}{2}z\right) \cos\left[\left(\omega + \frac{\Delta\omega}{2}\right)t - \left(k + \frac{\Delta k}{2}\right)z\right] \\ &\approx 2E_0 \cos\left(\frac{\Delta\omega}{2}t - \frac{\Delta k}{2}z\right) \cos(\omega t - kz), \end{aligned} \quad (6.8)$$

where the first cosine term represents the modulation envelope and the second the carrier phase. The modulation envelope then propagates with *group velocity*

$$v_g = \lim_{\Delta k \rightarrow 0} \frac{\Delta\omega}{\Delta k} = \frac{d\omega}{dk}. \quad (6.9)$$

The *group refractive index* can then be given with the definition in Equation (6.2) as

$$n_g = \frac{c}{v_g} = c \frac{dk}{d\omega} = \frac{d}{d\omega}(ck) = \frac{d}{d\omega}\left(c \frac{\omega}{v}\right) = \frac{d}{d\omega}\left(\frac{c}{v}\omega\right) = \frac{d}{d\omega}(n(\omega)\omega) = n(\omega) + \omega \frac{dn(\omega)}{d\omega}. \quad (6.10)$$

6.1.3 Tropospheric refractive index

While the electromagnetic wave approaches Earth's surface, in neutral atmosphere, below the ionosphere, the number of free electrons decreases to zero and the ionospheric refractive index approaches one. However, due to dry gases and water vapour, the refractive index in the troposphere differs from the free space. The accumulated phase difference caused by the tropospheric refraction can be significant, but typically less than that of the ionosphere. The changes in the total contribution of the troposphere are also within $\pm 10\%$ even in longer time periods, whereas the ionosphere can have large rapid changes (Klobuchar, 1996). More detailed studies on tropospheric parameters are provided by Bernhardt et al. (2000); Rüeger (2002). According to Wells et al. (1986) the troposphere is nondispersive for frequencies below 30 GHz. Kaplan and Hegarty (2006) state that the limit is 15 GHz. Below these limits, the tropospheric refractive index is then independent of the frequency, and the group and phase velocities are equal. Hence, it is enough to denote the *tropospheric refractive index* here as

$$n_{\text{tr}} = 1 + \Delta n_{\text{tr}}, \quad (6.11)$$

that is the refractive index of vacuum perturbed with the tropospheric contribution Δn_{tr} .

6.2 Radio measurements of satellite transmissions

As of today, there exist several global navigation satellite systems (GNSS), such as GPS, GLONASS, GALILEO and BEIDOU that can be used for ionospheric observations. The different GNSSs operate in UHF frequencies ranging from GALILEO's lowest frequency of 1176.45 MHz to the highest GLONASS frequency of 1605.375 MHz. In ionospheric studies the most used satellite system has been GPS with the main frequencies of 1575.42 (L1) and 1227.60 MHz (L2). The satellite orbit altitudes used by different GNSS are for GLONASS 19,100 km, for GPS 20,180 km, for BeiDou 21,528 km and for GALILEO 23,222 km (Kaplan and Hegarty, 2006). Besides GNSS systems, low Earth orbit (LEO) beacon satellites have also been used frequently in atmospheric studies. LEO satellite beacons operate typically with dual or tri-band VHF and UHF frequencies of 150, 400 and 1067 MHz. LEO refers to orbital altitudes less than 1,500 km (Bernhardt et al., 2000; Yamamoto, 2008; Vierinen et al., 2014).

6.2.1 Refractive indices for VHF and UHF signals

The principle of ionospheric observations with radio measurements of satellite beacon signals is based on the connection between the frequency-dependent refractive index and electron density, given with the Appleton-Lassen formula (6.3). However, in its original form given above, the connection is nonlinear and complex. For an efficient measurement model a linear equation without imaginary part is sought for.

As the UHF and VHF frequencies are much greater than 1 MHz, the electron collisions can be neglected and Z can be approximated with zero (Budden, 1961; Davies, 1965). This removes the imaginary part from the Appleton-Lassen formula (6.3). A gyrofrequency of 1.5 MHz results in $Y = 0.01$ at 150 MHz and decreasing with increasing signal frequency. Hence, Y in the Equation (6.3) can also be approximated with zero.

When electron collisions and gyrofrequencies are both omitted, the Appleton-Lassen formula (6.3) simplifies to

$$n^2 = 1 - X = 1 - \frac{\omega_p^2}{\omega^2}. \quad (6.12)$$

The relation is indeed more simple, however, still nonlinear. Therefore, as $\omega \gg \omega_p$, the refractive index can be approximated with first order Taylor polynomial at $\frac{\omega_p}{\omega} = 0$ with

$$n \approx 1 - \frac{1}{2} \left(\frac{\omega_p}{\omega} \right)^2. \quad (6.13)$$

Inserting $\omega_p = \sqrt{\frac{N_e e^2}{\epsilon_0 m_e}}$ (rad/s) then results in

$$n \approx 1 - \frac{N_e e^2}{2\epsilon_0 m_e \omega^2}. \quad (6.14)$$

The simplifying and linearising assumptions in Approximation (6.14) give rise to an error less than 1% with 150 MHz frequency, decreasing with higher frequency (Fehmers, 1996).

With Equation (6.10), the group refractive index for VHF and UHF signals can then be derived as

$$n_g = n(\omega) + \omega \frac{dn(\omega)}{d\omega} \approx 1 - \frac{N_e e^2}{2\epsilon_0 m_e \omega^2} + 2 \frac{N_e e^2}{2\epsilon_0 m_e \omega^2} = 1 + \frac{N_e e^2}{2\epsilon_0 m_e \omega^2}. \quad (6.15)$$

6.2.2 Wave propagation of VHF and UHF signals

Spatially inhomogeneous medium wave propagation in Equation (6.1) can be written as

$$\psi(t, L) = E_0 \cos \left(\omega t - \frac{\omega}{c} \int_0^L n(z) dz \right), \quad (6.16)$$

where, for the sake of convention, the integral is defined from receiver at $z = 0$ to satellite at distance $z = L$.

It is here enough to concentrate on the integral part inside the cosine function (6.16) that is the effect of the medium for the propagating signal. For the carrier phase it is typically given in radians

$$\phi(L) = \frac{\omega}{c} \int_0^L n(z) dz \quad (6.17)$$

and for the modulation envelope in metres

$$\rho(L) = \int_0^L n_g(z) dz. \quad (6.18)$$

When taking into account the nondispersive tropospheric contribution (6.11), the phase (6.17) can be written for the different intervals as

$$\phi(L) = \frac{\omega}{c} \left(\int_0^{L_0} n_{tr}(z) dz + \int_{L_0}^L n(z) dz \right), \quad (6.19)$$

where the first integral is defined along the signal path from ground receiver to distance L_0 where the ionospheric refraction becomes significant, and the second integral from this altitude to the upper boundary of the ionosphere L . Inserting the refractive index (6.14) results in radians as

$$\begin{aligned} \phi(L) &= \frac{\omega}{c} \left[\int_0^{L_0} (1 + \Delta n_{tr}) dz + \int_{L_0}^L \left(1 - \frac{N_e(z) e^2}{2\epsilon_0 m_e \omega^2} \right) dz \right] \\ &= \frac{\omega}{c} \left[L + \int_0^{L_0} \Delta n_{tr} dz - \int_{L_0}^L \frac{N_e(z) e^2}{2\epsilon_0 m_e \omega^2} dz \right] \\ &= \frac{\omega}{c} L + \frac{\omega}{c} T(L) - \frac{\alpha}{c\omega} TEC(L). \end{aligned} \quad (6.20)$$

Similarly for the group delay, combining Equations (6.18), (6.11) and (6.15) results in metres as

$$\rho(L) = L + T(L) + \frac{\alpha}{\omega^2} TEC(L). \quad (6.21)$$

In both Equations (6.20) and (6.21), L is the range between the transmitter and receiver, $T(L) = \int_0^{L_0} \Delta n_{tr} dz$ is the tropospheric contribution, $\alpha = \frac{e^2}{2\epsilon_0 m_e}$ a combination of constants and

$$TEC(L) = \int_{L_0}^L N_e(z) dz \quad (6.22)$$

is the slant *total electron content* (TEC). The last term, which includes TEC, is positive when considering phase velocity and negative when group velocity is considered. In GNSS literature temporal frequency is often used instead of angular. The coefficient in the ionospheric part is then $\frac{\alpha}{\omega} = \frac{\alpha}{4\pi^2 f^2} \approx \frac{40.3}{f^2}$.

6.2.3 Observables

The two main measurement types, made from satellite beacon signals are the *pseudorange* observable based on the group delay given in Equation (6.21) and the *carrier phase* observable based on the phase advancement given in Equation (6.20). For other ionospheric

effects on satellite signals, such as Doppler shift and Faraday rotation and their use as measurements see e.g. Klobuchar (1985, 1996).

In traditional use of satellite positioning, the main interest in an individual measurement is in the range between a user with an unknown position and a satellite with a known position. In such range measurement the ionosphere is a source of error. In ionospheric measurements the location of the receiver and hence the range to the satellite is known and the primary interest is in the unknown TEC.

The satellite orbital altitudes used by different GNSSs are around 20,000 km. In Equation (6.21), the contribution of a relatively high TEC of 100 TECU in pseudorange extends from 15 m at 1605.375 MHz to 1,800 m at 150 MHz. The total contribution of troposphere to pseudorange in satellite beacon frequencies is approximately between 2.4–25 m (Kaplan and Hegarty, 2006). Hence, for the navigation, with the aid of atmospheric models the range estimation can be carried out with sufficient accuracy even from a single measurement. However, when the interest is in the ionosphere, it is evident from the numbers above that the situation is much worse.

Equations (6.20) and (6.21) describe an ideal measurement taking into account only the physical composition of the atmosphere. Unfortunately, in real life, the measurements also suffer from several other nuisances. Below, the most significant errors and biases for the TEC measurements are included in the models of both observables. The error sources omitted here include antenna-phase center variations, earth tides, ocean loading, and for phase measurement, the phase windup effect (Håkansson et al., 2017). Taking the additional errors and biases into account consolidates the intuition why individual measurements as such are useless for most of the applications considering TEC.

To overcome this problem, at least partially, measurements with two different frequencies ω_1 and ω_2 are used. As part of the errors and biases are dispersive and part nondispersive, i.e. coherent, the coherent errors can then be canceled by combining the measurements, leaving only the frequency-dependent part.

Pseudorange observable

The transmitted GNSS satellite signals are modulated with different pseudorandom noise codes depending on the satellite system and the frequency at issue. When received, the signal is aligned with a reference signal and the modulated signals are compared. Due to the pseudorandomness the maximum correlation between the received and the replica code is achieved only when the signals are aligned perfectly. The amount the replica code needs to be shifted for maximising the correlation provides the travelling time of the signal. When multiplied with velocity c , the range between the receiver and the satellites is obtained (Wells et al., 1986; Kaplan and Hegarty, 2006). In GNSS positioning terminology the measurement of group delay is known as pseudorange, as it includes several biases. By including the most significant bias terms in Equation (6.21), the *pseudorange* can be

written as

$$\begin{aligned}\rho_\omega(L, t) = & c(\tau_{rec} + \tau_{sat} + b_{\rho, rec, \omega, code} + b_{\rho, sat, \omega, code}) + M_{\rho, sat, rec, \omega, code}(t) \\ & + L + T(L) + \frac{\alpha}{\omega^2} TEC(L) + \varepsilon_{\rho, sat, rec, \omega, code}(t),\end{aligned}\quad (6.23)$$

where the new parameters τ are the receiver and satellite clock errors, parameters b are the receiver and satellite hardware biases, M is the multipath error and ε is the measurement error due to thermal noise etc. The dependencies of different parameters to specific observations are given with subindex variables $\omega, \rho, rec, sat, code$ referring in corresponding order to frequency, observable type, receiver and satellite names and the measured code. For example, $b_{\rho, PRN02, L1, C/A}$ is the satellite bias for pseudorange measurements that uses GPS satellite PRN02 and L1 frequency with coarse/acquisition code modulation.

A rule of thumb for the measurement precision is 1% of the period between two code epochs (Wells et al., 1986). For GPS codes, this results in a precision of 1 ns for the P-code and 10 ns for the C/A-code. Converted to TEC measurements the P-code precisions are then 1.9 TECU for L1 and 1.1 TECU for L2. For C/A-code the precision in L1 is 19 TECU.

Differential group delay

When pseudorange measurements with two different angular frequencies ω_1 and ω_2 are available, the coherent part consisting of the range (with possible errors), clock errors and tropospheric error is canceled out in subtraction, resulting in

$$\begin{aligned}\Delta\rho(L, t) = & \rho_{\omega_2}(L, t) - \rho_{\omega_1}(L, t) \\ = & \alpha \left(\frac{1}{\omega_2^2} - \frac{1}{\omega_1^2} \right) TEC(L) \\ & + cb_{\rho, rec, \omega_2} - cb_{\rho, rec, \omega_1} + cb_{\rho, sat, \omega_2} - cb_{\rho, sat, \omega_1} \\ & + M_{\omega_2, \rho, 2}(t) - M_{\omega_1, \rho, 1}(t) + \varepsilon_{\omega_2, \rho}(t) - \varepsilon_{\omega_1, \rho}(t).\end{aligned}\quad (6.24)$$

The TEC can be then solved as

$$\begin{aligned}
\frac{1}{\alpha} \left(\frac{\omega_1^2 \omega_2^2}{\omega_1^2 - \omega_2^2} \right) \Delta \rho(L, t) = & TEC(L) + \frac{1}{\alpha} \left(\frac{\omega_1^2 \omega_2^2}{\omega_1^2 - \omega_2^2} \right) (cb_{\rho, rec, \omega_2} - cb_{\rho, rec, \omega_1}) \\
& + \frac{1}{\alpha} \left(\frac{\omega_1^2 \omega_2^2}{\omega_1^2 - \omega_2^2} \right) (cb_{\rho, sat, \omega_2} - cb_{\rho, sat, \omega_1}) \\
& + \frac{1}{\alpha} \left(\frac{\omega_1^2 \omega_2^2}{\omega_1^2 - \omega_2^2} \right) (M_{\omega_2, \rho, 2}(t) - M_{\omega_1, \rho, 1}(t)) \\
& + \frac{1}{\alpha} \left(\frac{\omega_1^2 \omega_2^2}{\omega_1^2 - \omega_2^2} \right) (\varepsilon_{\omega_2, \rho}(t) - \varepsilon_{\omega_1, \rho}(t)) \\
= & TEC(L) + DCB_{\rho, rec, \omega_1, \omega_2} + DCB_{\rho, sat, \omega_1, \omega_2} \\
& + M_{\rho, \omega_1, \omega_2}(t) + \varepsilon_{\rho, \omega_1, \omega_2}(t),
\end{aligned} \tag{6.25}$$

where the scaling of the differential error and bias terms converts them into TEC units. As it is convenient to use TEC units from here on, the terms are renamed on the last line. Variable *DCB* stands for *differential code bias* (DCB) and it is unknown. For the more precise P-code differential group measurement a 2 ns precision results in a 5.7 TECU precision. The possible multipath errors in the observations are in the scale of 10 m (Wells et al., 1986) resulting in almost 100 in TECU. Hence, in practice, the differential group delay TEC can be considered as an absolute measurement up to DCB, but contaminated with relatively large measurement noise.

Carrier phase observable

The *Carrier phase*, also known as *carrier beat phase*, *phase difference* and *phase advancement* measurement or observable, is used similarly for LEO beacon and GNSS differential carrier phase measurements. In the carrier phase measurement the difference is taken between the incoming signal phase and a constant reference frequency generated in the receiver. The measurement is the phase difference. When measuring the phase, the initial number of full phase difference cycles cannot be detected. Hence, a new bias term, *phase ambiguity* γ , needs to be included in the model. If the signal is lost during the measurement a new phase ambiguity bias term needs to be added into the model (Wells et al., 1986; Vierinen et al., 2014).

By including the phase ambiguity, with other additional bias and error parameters, in Equation (6.20), the *carrier phase observable* can be written as

$$\begin{aligned}
\phi(t) = & \omega (\tau_{rec} + \tau_{sat} + b_{\phi, rec, \omega} + b_{\phi, sat, \omega}) + \frac{\omega}{c} (L + T(L)) \\
& - \frac{\alpha}{c\omega} TEC(L) + M_{\phi, sat, rec, \omega}(t) + \gamma_{sat, rec, \omega} + \varepsilon_{\phi, sat, rec, \omega}(t),
\end{aligned} \tag{6.26}$$

where similarly to pseudorange measurements the clock and hardware bias terms τ and b are converted from seconds and range and tropospheric bias from metres to radians.

For a differential carrier phase measurement in GPS L1 frequency the precision rule of 1% from the wavelength results in a range precision of 2 mm (Wells et al., 1986). The corresponding TEC measurement precision is approximately 0.02 TECU. The TEC measurement precision improves with lower frequencies.

Differential carrier phase measurement

As the carrier phase observable is measured in radians, measurements in two frequencies need to be scaled to same frequency before the subtraction.

$$\begin{aligned}
\Delta\phi &= \phi_{\omega_2} - \frac{\omega_2}{\omega_1} \phi_{\omega_1} \\
&= \omega_2 (\tau_{rec} + \tau_{sat} + b_{\phi,rec,\omega_2} + b_{\phi,sat,\omega_2}) + \frac{\omega_2}{c} (L + T(L)) \\
&\quad - \frac{\alpha}{c\omega_2} TEC(L) + M_{\phi,sat,rec,\omega_2}(t) + \gamma_{sat,rec,\omega_2} + \varepsilon_{\phi,sat,rec,\omega_2}(t) \\
&\quad - \frac{\omega_2}{\omega_1} \left[\omega_1 (\tau_{rec} + \tau_{sat} + b_{\phi,rec,\omega_1} + b_{\phi,sat,\omega_1}) + \frac{\omega_1}{c} (L + T(L)) \right. \\
&\quad \left. + \frac{\alpha}{c\omega_1} TEC(L) + M_{\phi,sat,rec,\omega_1}(t) + \gamma_{sat,rec,\omega_1} + \varepsilon_{\phi,sat,rec,\omega_1}(t) \right] \\
&= \frac{\alpha}{c} \left(\frac{\omega_2^2 - \omega_1^2}{\omega_1^2 \omega_2} \right) TEC(L) \\
&\quad + \omega_2 (b_{\phi,rec,\omega_2} - b_{\phi,rec,\omega_1}) + \omega_2 (b_{\phi,sat,\omega_2} - b_{\phi,sat,\omega_1}) + M_{\phi,sat,rec,\omega_2}(t) - \frac{\omega_2}{\omega_1} M_{\phi,sat,rec,\omega_1}(t) \\
&\quad + \gamma_{sat,rec,\omega_2} - \frac{\omega_2}{\omega_1} \gamma_{sat,rec,\omega_1} + \varepsilon_{\phi,sat,rec,\omega_2}(t) - \frac{\omega_2}{\omega_1} \varepsilon_{\phi,sat,rec,\omega_1}(t)
\end{aligned} \tag{6.27}$$

Converting the measured difference to TEC units results in

$$\begin{aligned}
\frac{c}{\alpha} \left(\frac{\omega_1^2 \omega_2}{\omega_2^2 - \omega_1^2} \right) \Delta\phi &= TEC(L) + \frac{c}{\alpha} \left(\frac{\omega_1^2 \omega_2}{\omega_2^2 - \omega_1^2} \right) (\omega_2 (b_{\phi,rec,\omega_2} - b_{\phi,rec,\omega_1}) - \omega_2 (b_{\phi,sat,\omega_2} - b_{\phi,sat,\omega_1})) \\
&\quad + \frac{c}{\alpha} \left(\frac{\omega_1^2 \omega_2}{\omega_2^2 - \omega_1^2} \right) \left(M_{\phi,sat,rec,\omega_2}(t) - \frac{\omega_2}{\omega_1} M_{\phi,sat,rec,\omega_1}(t) \right) \\
&\quad + \frac{c}{\alpha} \left(\frac{\omega_1^2 \omega_2}{\omega_2^2 - \omega_1^2} \right) \left(\gamma_{sat,rec,\omega_2} - \frac{\omega_2}{\omega_1} \gamma_{sat,rec,\omega_1} \right) \\
&\quad + \frac{c}{\alpha} \left(\frac{\omega_1^2 \omega_2}{\omega_2^2 - \omega_1^2} \right) \left(\varepsilon_{\phi,sat,rec,\omega_2}(t) - \frac{\omega_2}{\omega_1} \varepsilon_{\phi,sat,rec,\omega_1}(t) \right) \\
&= TEC(L) + IFB_{\phi,rec,\omega_1,\omega_2} + IFB_{\phi,sat,\omega_1,\omega_2} + M_{\phi,sat,rec,\omega_1,\omega_2}(t) \\
&\quad + \gamma_{\phi,sat,rec,\omega_1,\omega_2}^* + \varepsilon_{\phi,sat,rec,\omega_1,\omega_2}^*(t),
\end{aligned} \tag{6.28}$$

where the scaled differential bias terms are in TEC units and renamed at the last line as *interfrequency bias* (IFB) denoted here with variable IFB , and the naming for the rest of the parameters is self-explanatory.

The phase ambiguity γ^* is an unknown constant for a continuous measurement between a satellite-receiver pair. The IFBs can be assumed constant for each individual receiver and satellite when using the same frequency pairs. Hence, the IFBs can be included in the phase ambiguity parameter as

$$\gamma_{\phi,sat,rec,\omega_1,\omega_2} := IFB_{\phi,rec,\omega_1,\omega_2} + IFB_{\phi,sat,\omega_1,\omega_2} + \gamma_{\phi,sat,rec,\omega_1,\omega_2}^*.$$

The multipath error is on a centimeter scale (Wells et al., 1986) and can be included in the measurement error term

$$\varepsilon_{\phi,sat,rec,\omega_1,\omega_2}(t) := M_{\phi,sat,rec,\omega_1,\omega_2}(t) + \varepsilon_{\phi,sat,rec,\omega_1,\omega_2}^*(t).$$

The differential carrier phase TEC measurement can then be written as

$$\frac{c}{\alpha} \left(\frac{\omega_1^2 \omega_2}{\omega_2^2 - \omega_1^2} \right) \Delta\phi = TEC(L) + \gamma_{\phi,sat,rec,\omega_1,\omega_2} + \varepsilon_{\phi,sat,rec,\omega_1,\omega_2}(t). \quad (6.29)$$

The precision of differential phase measurement at GPS frequencies is approximately 0.03 TECU and differential phase measurements with much lower LEO beacon frequencies are even more precise. However, as $\gamma_{\phi,sat,rec,\omega_1,\omega_2}$ is not known, the measurement remains relative.

6.2.4 Carrier phase leveling

The differential code measurement is absolute up to DCB, but has a low precision. The phase differential measurement is relative due to the unknown phase ambiguity; however the measurement precision is high.

To achieve the absolute scale for the more accurate carrier phase measurement, the differential carrier phase is fitted to the differential group delay measurement. The fitting is done by calculating the offset σ_{offset} between the two measurements using the high-elevation parts as the low-elevation measurements are more prone to multipath errors (Klobuchar, 1996; Horvath and Crozier, 2007). The curve fitting results in real TEC measurement, which is absolute with respect to phase ambiguity and has the accuracy of the phase measurements, but still contains hardware biases. The final measurement can be written as

$$\frac{c}{\alpha} \left(\frac{\omega_1^2 \omega_2}{\omega_2^2 - \omega_1^2} \right) \Delta\phi + \sigma_{\text{offset}} = TEC(L) + DCB_{\rho,rec,\omega_1,\omega_2} + DCB_{\rho,sat,\omega_1,\omega_2} + \varepsilon_{\phi,sat,rec,\omega_1,\omega_2}(t). \quad (6.30)$$

Differential code bias (DCB)

The reason why differential phase measurements with DCB, instead of phase ambiguity, are preferred is that the phase ambiguity remains the same only for each individual satellite-receiver lock. The DCBs are also unknown, however, it can often be assumed that each individual signal transmitted from a GNSS satellite has a DCB that is independent of the receiver. Correspondingly, it can be assumed that each receiver has a DCB for each different signal, independent of the transmitting satellite. In comparison to TEC, it can be assumed that the changes in DCBs are slow. However, the receiver DCB depend on e.g. temperature (Coster et al., 2013) and daily variations of over 8 TECU are reported (Dyrud et al., 2008). In the GLONASS system, each satellite transmits slightly different frequencies and thus each satellite-receiver combination has its unique DCBs for different signals between the pair. Detailed reviews about GNSS biases are provided by Dyrud et al. (2008); Håkansson et al. (2017). For bias calibration see e.g. Dyrud et al. (2008); Vierinen et al. (2016).

6.2.5 LEO beacon satellite measurement model

The ground-based measurement of LEO dual-frequency beacon signals is a differential carrier phase measurement as given in Section 6.2.3. The TEC measurements can then be written as a linear model given in Equation (3.3)

$$\mathbf{m}_{\text{LEO}} \approx \mathbf{A}_{\text{LEO}} \mathbf{f} + \mathbf{B}_{\gamma} \boldsymbol{\gamma} + \boldsymbol{\varepsilon}_{\text{LEO}}, \quad (6.31)$$

where the vector \mathbf{m}_{LEO} consists of individual relative slant TEC measurements given in (6.29) and correspondingly $\boldsymbol{\varepsilon}_{\text{LEO}}$ the measurement errors $\varepsilon_{\phi, \text{sat}, \text{rec}, \omega_1, \omega_2}$. Similarly to Equation (2.3), the rows of matrix \mathbf{A}_{LEO} are discrete approximations for the integral operators of slant TEC (6.22) for all the measured signal paths operating on the unknown electron density values $\mathbf{f} \in \mathbb{R}^N$ in the discretised three-dimensional domain. The vector $\boldsymbol{\gamma}$ consists of unknown phase ambiguity constants $\gamma_{\phi, \text{sat}, \text{rec}, \omega_1, \omega_2}$ and it needs to be taken into account as an additional unknown. As the phase ambiguity remains the same during each continuous observation, several individual measurements in the above model share a common

$\gamma_{\phi,sat,rec,\omega_1,\omega_2}$ parameter.

$$\mathbf{B}_\gamma = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 1 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix} \quad (6.32)$$

is a design matrix of zeros and ones, which picks the correct ambiguity parameter for each measurement.

6.2.6 GNSS satellite measurement model

TEC measurements of GNSS satellite signals are based on the levelled carrier phase measurement given in Equation (6.30). In the form of Equation (3.3), a vector of GNSS measurements can then be modelled as

$$\mathbf{m}_{\text{GNSS}} \approx \mathbf{A}_{\text{GNSS}} \mathbf{f} + \mathbf{B}_{\text{rec}} \mathbf{b}_{\text{rec}} + \mathbf{B}_{\text{GNSS}} \mathbf{b}_{\text{GNSS}} + \boldsymbol{\varepsilon}_{\text{GNSS}}, \quad (6.33)$$

where again the vector \mathbf{m}_{GNSS} consists of individual GNSS TEC measurements (6.30). The measurement error vector $\boldsymbol{\varepsilon}_{\text{GNSS}}$ consists of error terms $\varepsilon_{\phi,sat,rec,\omega_1,\omega_2}$. Similarly to LEO measurements, the matrix \mathbf{A}_{GNSS} is the discretisation of the integrals in slant TEC (6.22) acting on the discretised electron density values \mathbf{f} , as shown in Equation (2.3) for the two-dimensional measurements. The vectors \mathbf{b}_{rec} and \mathbf{b}_{sat} consists of DCBs of the measurements. \mathbf{B}_{rec} and \mathbf{B}_{sat} are design matrices, similar to \mathbf{B}_γ in Equation (6.32), picking the correct DCB for each measurement.

The altitudes of GNSS satellites are around 20,000 km, thus besides the ionosphere, most of the signal propagation takes place in the plasmasphere above. The electron density in plasmasphere is generally much lower than in the ionosphere; however, due to the long ray paths, the resulting contribution to electron content can be significant. In Lunt et al. (1999) it is reported that during solar minimum the plasmaspheric contribution over Europe is typically a few TEC units. At night, especially in winter it can constitute 50% or even more in GNSS measurements. The contribution decreases at higher latitudes and the proportional contribution decreases towards solar maximum.

If the whole domain spanned by the receivers and satellites should be modelled similarly, the resulting grid size can become unnecessarily high-dimensional. One technique to reduce

dimensions is by using an irregular grid, where voxel sizes increase towards the boundaries, particularly at high altitudes. Another scheme is to extend the grid only over the ionosphere and use plasmaspheric models (see e.g. Jakowski and Hoque (2018) and references therein) for the exceeding parts of the measurements. The plasmaspheric model can be introduced into the measurement model in (6.22), as

$$TEC(L) = \int_{L_0}^{L_{iono}} N_e(z)dz + \int_{L_{iono}}^L N_e(z)dz \approx \int_{L_0}^{L_{iono}} N_e(z)dz + \int_{L_{iono}}^L N_{e,pmodel}(z), \quad (6.34)$$

where L_{iono} is the upper boundary altitude of the reconstruction grid and $N_{e,pmodel}$ a plasmaspheric model. A straightforward approach is to assume a uniform but unknown plasmaspheric electron density, resulting in

$$TEC(L) \approx \int_{L_0}^{L_{iono}} N_e(z)dz + (L - L_{iono})N_{e,punif}, \quad (6.35)$$

where $N_{e,punif}$ the uniform plasmaspheric electron density constant. The TEC in (6.34) or (6.35) is then plugged into equation (6.30). The selection of (6.35) introduces an additional unknown in the final measurement model (6.33).

In ionospheric studies, the majority of GNSS TEC measurements are typically made with ground-based receivers, with fixed and known locations. However, in satellite *radio occultation* (RO) the GNSS measurement is carried out with a LEO satellite onboard receiver (Hajj et al., 1994). The main difference between the two is thus that in RO also the receiver is in motion. The equations above can be used for modelling both ground-based and RO measurements.

6.3 Ionosonde measurements

In 1924 Breit and Tuve (1926) proved the existence of an ionised layer in Earth's atmosphere by receiving ionospheric echoes from a transmitted *high-frequency* (HF, 3–30 MHz) signal. The seminal study laid the foundation for ionospheric soundings and ionosondes.

An ionosonde is practically a radar transmitting HF pulses and measuring the time it takes for a pulse to travel back and forth to the reflection altitude in the ionosphere. The reflection occurs when the refractive index reaches zero. For the ordinary mode that is the altitude where the plasma frequency matches the frequency of the propagating wave, while a signal with a higher frequency than the current maximum plasma frequency will penetrate through the ionosphere. Hence, the range of the transmitted frequencies should cover the current plasma frequency. Usual ionospheric peak electron densities range from 10^{10} to $10^{12} \frac{1}{m^3}$ (Klobuchar, 1985), with conversions given in Equation (6.5) the corresponding plasma frequency range from 0.9 to 9 MHz. A typical ionosonde covers the frequencies from 0.5 to 20 MHz.

When the signal frequency in use is close to the plasma frequency the earlier assumptions regarding collision and gyro frequencies are not valid. After the lowest frequencies, when the signal frequency is higher than 2 MHz (Klobuchar, 1985), the collisions can again be neglected, however, even at the highest frequencies used in ionosondes, the presence of the magnetic field and hence gyrofrequency needs to be taken into account in the refractive index given with Appleton-Lassen formula (6.3).

The pulses transmitted from an ionosonde travel at group velocity (6.9). Before the reflection occurs, the propagating wave is slowed down by the ionisation below the reflection altitude. Hence, deriving the altitude from the signal travel time assuming that the pulses propagate with the speed of light will result in so-called *virtual height*

$$h' = \frac{c}{2}\Delta t. \quad (6.36)$$

The modelling for the roundtrip time can be carried out more accurately by taking the group index into account as

$$\Delta t = \frac{2}{c} \int_0^h n_g(z) dz, \quad (6.37)$$

where the integral is defined along a line from ionosonde location at $z = 0$ to $z = h$, where h is the *real height* i.e. the actual reflection height. When the pulse frequency and plasma frequency up to current altitude are included, the virtual height can be written as

$$h'(\omega) = \int_0^{h(\omega)} n_g(\omega_p(z), \omega) dz. \quad (6.38)$$

The solution for $h(\omega)$ in Equation (6.38) is a nonlinear problem and cannot be solved within the linear framework provided in Chapter 4. In ionospheric imaging, usually real height profiles pre-analysed with some specific scaling algorithm, such as NhPC (Huang and Reinisch, 1996) within Automatic Real Time Ionogram Scaler with True Height (ARTIST) (Reinisch and Huang, 1983), POLynomial ANALysis (POLAN) (Titheridge, 1985), Autoscala (Pezzopane and Scotto, 2004) and NeXtYZ (Zabotin et al., 2006) are used.

An analysed real height profile is a vector of reflection altitudes for corresponding pulse frequencies. Typically it is assumed that the reflections take place directly above the instrument location. This results in a measurement model

$$h_{\text{ionos}}(\omega) = h(\omega) + \varepsilon_{\text{ionos}}(\omega), \quad (6.39)$$

where $\varepsilon_{\text{ionos}}$ are errors in altitude that are dependent on each other within each analysed profile. When converting the plasma frequencies and measurement errors to electron density and approximating them in model grid points, ionosonde measurements can be modelled in the form of Equation (3.3) as

$$\mathbf{m}_{\text{ionos}} \approx \mathbf{A}_{\text{ionos}} \mathbf{f} + \boldsymbol{\varepsilon}_{\text{ionos}}, \quad (6.40)$$

where $\mathbf{A}_{\text{ionos}}$ is a simple design matrix picking the column of \mathbf{f} closest to the instrument location up to the highest reflection altitude.

6.4 Incoherent scatter radar measurements

The principle of a basic radar is to transmit pulsed or continuous electromagnetic waves and to receive the signal reflected from a hard target, such as an aeroplane, a ship or a speeding car. Based on the travel time and the doppler shift of the signal the measurement typically gives the distance to the target as well as the speed in the radial direction from the radar.

The incoherent scatter radar (ISR) theory was first proposed by Gordon (1958) to investigate Earth's ionosphere. Intuitively the incoherent scatter can be understood as a number of small scatterers distributed randomly in a volume. However, instead of an actual reflection, the free electrons in the ionosphere will accelerate when illuminated with the incident electric field of the radar signal. As a result, the electrons start to re-radiate as Hertzian dipoles in the corresponding frequency. The physical phenomenon is called Thomson scattering. The movement of the electrons in ionospheric plasma is not completely free, but it is dominated by the significantly more massive positive ions. Thus, even though the incoherent backscatter is from the electrons, the measurement will include a Doppler effect originating from the ion velocities.

In contrast to a single hard target, the ionosphere is a continuous medium and the scattering will take place at several altitudes. If a plain continuous sine wave signal is transmitted, it is impossible to say from which distance the received signal is scattering. To overcome this so-called range ambiguity, some transmission modulation is required in the transmitted signal.

Due to spatial and temporal fluctuations in plasma, the measured backscattered field can be considered as a Gaussian random variable with zero mean. Hence, it is more informative to estimate the covariance of the measurements. The estimated autocovariance function can be represented also with its Fourier transform pair i.e. power spectral density. Based on the power spectral density, plasma parameters such as electron density, ion and electron temperature, ion mass ratio and ion velocity can be obtained.

For ionospheric imaging, the most important plasma parameter is the electron density. An individual electron density measurement can be modelled as

$$m_{\text{IS}} = N_e(z) + \varepsilon_{\text{IS}}, \quad (6.41)$$

where $z \in \mathbb{R}^3$ is the measurement location along the radar beam and ε_{IS} the corresponding measurement error. Similarly to ionosonde measurements, the incoherent scatter radar measurements can be interpolated to model grid points and modelled as direct measurements of unknown electron densities in the form of Equation (3.3) as

$$\mathbf{m}_{\text{IS}} \approx \mathbf{A}_{\text{IS}} \mathbf{f} + \varepsilon_{\text{IS}}, \quad (6.42)$$

where the design matrix \mathbf{A}_{IS} selects values of \mathbf{f} related to corresponding measurements.

6.5 Langmuir probe in situ measurements

Langmuir probe is named after Irving Langmuir, who pioneered the method at General Electric in the 1920s. It is one of the most straightforward ways of measuring plasma (Klobuchar, 1985). However, it is an in situ measurement performed inside the medium, hence, ionospheric plasma measurements require a vehicle with an access to the ionosphere, such as a satellite or a rocket.

Langmuir probe measurements are based on detection of electric current between two conducting surfaces interacting with the medium. Typically the current is measured between a plane, a cylindrical, or a spherical shaped electrode and the satellite surface. The measurements are carried out by changing the probe potential in small steps. The sweep over a range of realistic potentials produces a voltage-current curve. Plasma parameters, such as electron density, electric potential of plasma and electron temperature, can then be determined from the curve (Klobuchar, 1985; Hargreaves, 1992; Chen, 2003).

Similarly to ionosonde and incoherent scatter radar measurements, an individual electron density measurement provided by Langmuir probe is modelled here as

$$m_{\text{LP}} = N_e(z) + \varepsilon_{\text{LP}}, \quad (6.43)$$

where $z \in \mathbb{R}^3$ is the location of the probe. For a discretised system (3.3), a vector of measurements is modelled again as

$$\mathbf{m}_{\text{LP}} \approx \mathbf{A}_{\text{LP}} \mathbf{f} + \boldsymbol{\varepsilon}_{\text{LP}}, \quad (6.44)$$

where the design matrix \mathbf{A}_{LP} selects values of \mathbf{f} related to the probe location at the time of the measurement.

Chapter 7

Development of methodology in ionospheric imaging

The use of tomographic methods for ionospheric imaging was first suggested by Austen et al. (1986) and later published in Austen et al. (1988). The article presented a two-dimensional simulation study assuming LEO satellite measurements from a chain of receiver stations. Iterative ART and SIRT algorithms were used with a Chapman profile (Chapman, 1931) as an initial guess.

The first electron density reconstructions with real observations from LEO satellite transmissions were presented by Andreeva (1990). The relative nature of phase measurements was taken into account by solving the phase ambiguity within the inversion that was carried out with ART. Pryse and Kersley (1992) used independent EISCAT incoherent scatter radar data to validate reconstruction results obtained with a setup of two receivers and the SIRT algorithm.

Another early simulation study was carried out with MART by Raymund et al. (1990). In the aforementioned study the limitations of ionospheric measurements and necessity of prior information was acknowledged. The limitations of the satellite measurement geometry and the resulting ill-posedness were studied more explicitly later by Yeh and Raymund (1991) and Raymund et al. (1994b). Studies on resolution limits due to geometric limitations and effects on station spacing were later carried out by Na et al. (1995) and Sutton and Na (1995). Saksman et al. (1997) showed that due to restricted measurement geometry an infinite amount of ionospheric electron density functions can be defined that are invisible to such a set-up.

Work presented in Fremouw et al. (1992) had several major contributions for the field. It discussed the use of stochastic inversion presented in Tarantola and Valette (1982), used earlier mostly in geophysics. However, the inversion was carried out with weighted damped least squares (Menke, 1989), which is analogous to generalised Tikhonov regularisation. The approach also utilised basis functions, namely Fourier basis functions in the horizontal

direction and empirical orthogonal functions (EOF) in the vertical. The EOFs were based on model ionospheres.

The use of ionosonde measurements in ionospheric imaging was speculated on in Kersley et al. (1993). Raymund et al. (1994a) used scaled ionograms in a simulation study. Later Heaton et al. (1995) used ionosonde as prior information.

The inclusion of GNSS measurements to ionospheric imaging was already suggested by Yunck et al. (1988). The far-sighted speculations also acknowledged the lack of vertical information provided by ground-based satellite measurements and considered possibilities of satellite radio occultation (RO) measurements. A two-dimensional simulation study with GPS-to-LEO RO measurements was carried out by Hajj et al. (1994), where the TSVD approach was used to regularise the inversion. Singular values were also used to study the effect of improved measurement geometry. The first experimental results with GPS measurements were carried out by Rius et al. (1997). The approach used a three-dimensional spatial domain with four vertical layers and Kalman filter for the temporal dimension.

The use of basis functions in three dimensions was presented by Howe et al. (1998). The functions were constructed by combining spherical harmonics in the horizontal and EOF in the vertical direction. The simulation study considered GPS and LEO measurements and solved the GPS DCB within the procedure. The use of EOFs for three-dimensional tomography was continued later with Multi-Instrument Data Analysis System (MIDAS) by Mitchell and Spencer (2003); Bust et al. (2007); Chartier et al. (2012); Bruno et al. (2019).

In global-scale four-dimensional ionospheric imaging the amount and quality of measurements vary spatially, and in voxel-based approaches, the role of realistic physical prior information becomes even more pivotal. Due to the significant role of the prior distribution, these approaches are often referred to as data assimilation methods, and the prior distribution is then more commonly known as *background model*. The data assimilation methods and nomenclature originate from meteorology, oceanography and geophysics (Tarantola, 1987; Menke, 1989; Daley, 1991; Daley and Barker, 2000). In ionospheric imaging, most of the applied methods are variations of Kalman filter and can use any measurements that can be modelled as linearised functions of electron density. However, the large number of unknown parameters give rise to computational issues with the Kalman filter approach. The size of the covariance matrices in model space is $N \times N$ and computational complexity for school-book matrix multiplication and inversion grows as $\mathcal{O}(N^3)$. Hence, in the following methods there are practically two main differences: First, the selected background ionospheric model that can be anything from a simple climatology to complicated parametrised physical models. Second, how the algorithm handles the covariance matrices that are too large to fit in the computer memory.

Bust et al. (2004, 2007) derived the Ionospheric Data Assimilation Three-Dimensional (IDA3D) algorithm from Three-Dimensional Variational Data Assimilation (3DVAR) (Daley, 1991). The 3DVAR is a general approach allowing a nonlinear forward model, however,

when assuming a linear forward model and Gaussian measurement error and background distributions, the approach reverts to the Kalman filter. IDA3D uses ground-based GPS and LEO satellite, satellite RO, satellite in situ and ionosonde measurements. Several ionospheric models have been used as a background model, including International Reference Ionosphere (IRI) (Bilitza et al., 1993; Bilitza, 2001) and Parameterized Ionosphere Model (PIM) (Daniell et al., 1995). The background information is fed into the Kalman filter in the prediction step that is a mixture of earlier time step and the background model (Bust and Mitchell, 2008). According to Bust et al. (2004), IDA3D does not solve the posterior covariance, but only its diagonal i.e. the posterior variance.

Angling and Cannon (2004) presented an approach very similar to IDA3D, later entitled Electron Density Assimilative Model (EDAM) (Angling and Khattatov, 2006; Angling, 2008). EDAM uses PIM or IRI model as the prior mean and updates it with satellite RO and ground-based satellite measurements. EDAM is also a version of a Kalman filter, where a persistence model with exponential delay is used as a dynamic model. Only the diagonal of the posterior covariance matrix is solved and parametric correlations are given for non-diagonal entries for the following time step. Prior covariance matrix elements with the distance exceeding a predefined value are discarded resulting in a sparser covariance matrix. The method is also capable of solving DCBs within the tomographic analysis (Angling, 2008). Angling and Jackson-Booth (2011) added virtual height ionosonde measurements to EDAM in a nonlinear setting.

A method called GPS Ionospheric Inversion (GPSII) was presented by Fridman et al. (2006, 2009). It uses GPS and LEO satellite, GPS-LEO satellite RO, radio altimeter VTEC, LEO in situ, and ionosonde measurements. It uses a nonlinear model to obtain nonnegative solutions and the actual solution is obtained with the Newton-Kontorovich method. Within iterations, the GPSII uses a combination of generalized Tikhonov regularisation and Bayesian approach with a discrete multivariate Gaussian prior: The method uses a prior covariance matrix but it is also weighted with a regularisation parameter. The Kalman filter-type error covariance matrix propagation is not used. Both IRI 2000 and PIM have been used as a background model. It is stated that a factorised (separable) representation of the covariance matrix has a substantial positive effect on the memory requirements and computation speed. GPSII solves DCB within the inversion.

A method utilising the Gaussian random field prior/Kriging/Gaussian process was introduced by Minkwitz et al. (2015). In this approach the covariance is given as a three-dimensional function that can be integrated according to measurement geometry to obtain the covariance for the TEC measurements. The covariance function parameters are then estimated from the measurement data. For the actual inversion, the covariance needs to be evaluated then only between the reconstructed locations and the TEC measurements. Hence, the reconstruction could be carried out e.g. for an individual two-dimensional plane inside the actual three-dimensional domain. In Minkwitz et al. (2016) the approach was extended to a four-dimensional case by adding the temporal dimension to the covariance function.

An approach called the Global Assimilation of Ionospheric Measurements (GAIM) model (Schunk et al., 2004; Scherliess et al., 2004; Gardner et al., 2014; Scherliess et al., 2017) has been developed by Utah State University. There exist different versions of the approach that use a reduced-state Kalman filter and ensemble Kalman filter. Different background models have been used from more simple ionospheric models (Schunk et al., 2004) to a physical Ionosphere-Plasmasphere Model that utilises ionospheric drivers such as neutral densities and winds, magnetospheric and equatorial electric fields, and auroral precipitation (Scherliess et al., 2004).

The similarly named Global Assimilative Ionospheric Model (GAIM) has been developed by the University of Southern California and the Jet Propulsion Laboratory (USC/JPL) (Rosen et al., 2001; Hajj et al., 2004; Wang et al., 2004). GAIM USC/JPL, again, has a simpler version utilising a Kalman filter, where covariance matrix elements corresponding to distances over a preset value are discarded. A more complicated version uses the 4DVAR approach. 4DVAR (Courtier et al., 1994) is a general variational approach for data assimilation, which, in a case of linear measurement and dynamical models, reverts to Kalman smoother (Carrassi et al., 2018). The 4DVAR version of GAIM USC/JPL incorporates several ionospheric drivers that are estimated along the electron density within a range of time.

More recently, Elvidge and Angling (2019) presented a method called advanced ensemble electron density assimilation system (AENeAS). Similarly to different GAIM models, AENeAS is a physics-based data assimilation model and it seeks to predict the ionospheric state. AENeAS uses Thermosphere Ionosphere Electrodynamics General Circulation Model (TIE-GCM) (Qian et al., 2014) and NeQuick (Nava et al., 2008) as background models and local ensemble transform Kalman filter (LETKF) (Hunt et al., 2007). LETKF is a version of EnKF where the assimilation is performed only for local regions, which further reduces the state space of the model.

An early review on ionospheric imaging methods is provided by Raymund (1994) and a comparison of methods by Raymund (1995). Another introduction to ionospheric imaging and its early methods is given by Fehmers (1996). A book on ionospheric imaging with focus on iterative methods is provided by Kunitsyn et al. (2003). Bust and Mitchell (2008) provide a comprehensive review article where most of the present methods are already discussed.

7.1 TomoScand

TomoScand is a system for ionospheric imaging generated during the 2010s in the Finnish meteorological institute and Sodankylä geophysical observatory, University of Oulu. At this point it has been used mostly regionally over Northern Europe. It utilises measurements of GPS, GLONASS, GALILEO and LEO satellite signals, incoherent scatter radars, ionosondes, satellite in situ probings and GNSS-to-LEO RO. The GNSS DCBs can be

estimated within the system.

Similarly to methods such as IDA3D, EDAM etc. presented above, the TomoScand algorithm uses a simplified Kalman filter without solving the posterior covariance, or solving only its diagonal. Currently the dynamical model (5.8) in use is a persistence model with transition $\mathbf{H}^{(l-1)} = \lambda \mathbf{I}$ and attenuation $0 \leq \lambda \leq 1$. Alternatively, some ionospheric model can be used in the prediction step or directly as a prior mean.

The essential difference to aforementioned similar techniques is in the construction of the prior covariance. In particular, TomoScand uses GMRF correlation priors, presented in Section 4.3, for representing the prior distribution. TomoScand then relies heavily on sparse matrix implementations. Currently Multifrontal Massively Parallel sparse direct Solver (MUMPS) (Amestoy et al., 2001, 2019) with an R interface RMUMPS (<https://github.com/morispa/rmumps>) is used for solving the high-dimensional linear problem in parallel. The main steps of TomoScand are given in Algorithm 1. An example visualisation of TomoScand reconstruction from GNSS measurements is given in Figure 7.1.

Algorithm 1 TomoScand analysis

1. Set the spatial and temporal domain.
 - Grid (lat, long, alt) (geographic coordinates, possibly irregular).
 - Start and end time (UTC).
 2. Read data
 - Read measurements (GNSS, LEO, ionosonde, in situ, occultation,...).
 - Read GNSS satellite DCBs (Section 6.2.4).
 - Data quality control and filtering.
 3. Form measurement models
 - Formation of matrices $\mathbf{A}_{\text{GNSS}}, \mathbf{A}_{\text{LEO}}, \dots$ corresponding with the measurements in use (Chapter 6).
 - Measurement error estimation, if not provided with data.
 4. Form prior distribution (i.e. background model) (Chapter 4).
 - Set prior mean for the unknown electron density (i.e. background mean).
 - Prediction step (5.8)
 - Set prior covariance (i.e. background error covariance).
 - Set standard deviation/variance mask for the unknown electron density. Previous posterior variance can be used (see Step 6 below).
 - Set correlation lengths (in all 3 coordinate directions).
 - Form matrix \mathbf{L}_C (Section 4.3.1).
 - Set prior distributions for:
 - GNSS DCBs
 - LEO phase ambiguity
 - Uniform plasmaspheric contribution
 5. Solve MAP estimate (4.24) for the sparse linear system (RMUMPS).
 6. Optionally solve the posterior covariance, its diagonal, or parts of it (RMUMPS).
 7. Save and plot results.
 8. Set start and end time for the next step and start again from item 2 above.
-

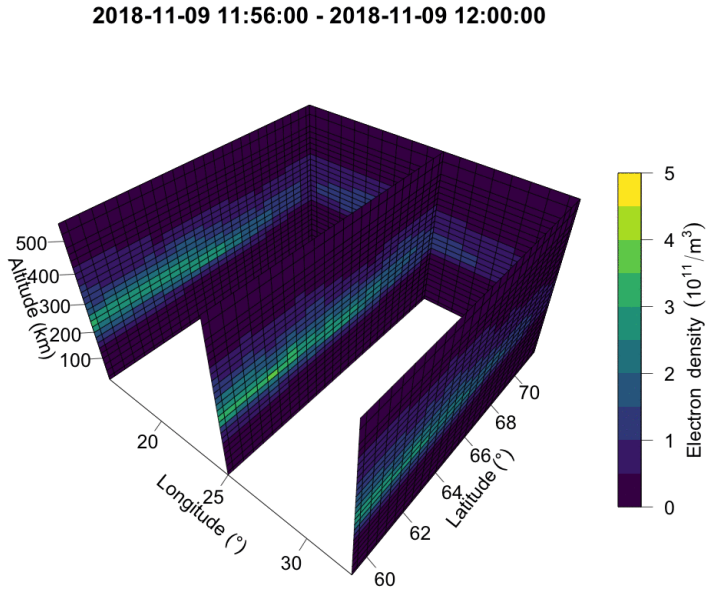
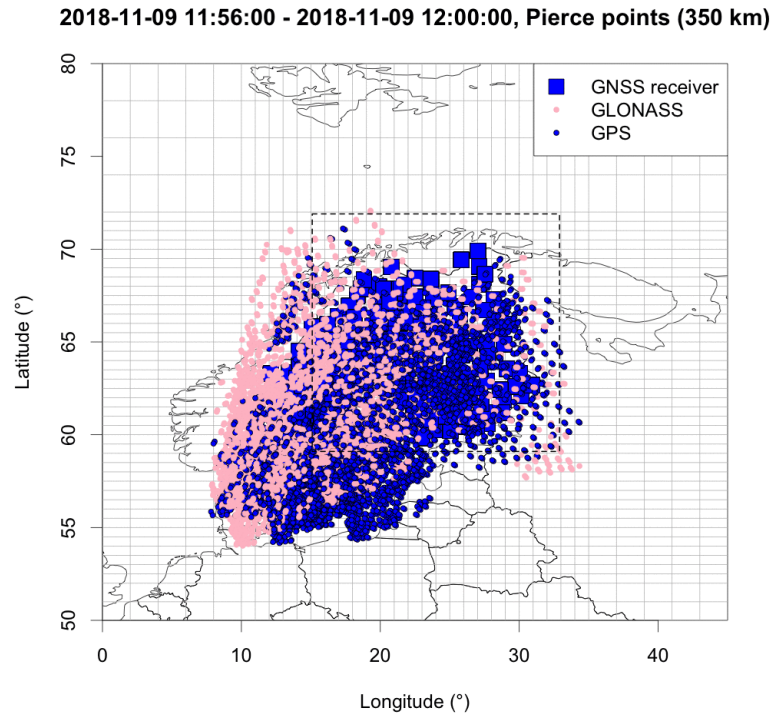


Figure 7.1: Example output from TomoScand analysis. Top: Tomographic domain, with irregular reconstruction grid and locations of the GPS and GLONASS satellite pierce points at an altitude of 350 km within a 2-min interval. Bottom: Three-dimensional reconstruction of ionospheric electron density.

Chapter 8

Discussion and conclusions

In this thesis an algorithm for four-dimensional multi-instrument ionospheric electron density imaging is developed. The algorithm uses a Bayesian approach for obtaining the most probable state of the ionospheric electron density, by updating the prior distribution i.e. the existing information of the ionospheric state with a set of new measurements. When used sequentially, the method is generally known as Kalman filter. In contrast to other Bayesian approaches used for ionospheric imaging, the prior distribution is essentially given as a Gaussian Markov Random Field correlation prior. The approach allows determining the prior covariance in an intuitive manner with a parametric function. However, for the numerical computations the covariance information is represented with a sparse precision matrix. Thus, instead of forming the $N \times N$ covariance matrix, approximately the same information is given with a precision matrix where the number of non-zero elements grows only as $\mathcal{O}(N)$. The precision matrix is also quick to construct and can easily be modified for different covariance structures. Effectively the same information can also be given for different discretisations of the domain as well as for irregular grids, as long as the discretisation lengths remain substantially shorter than the corresponding correlation lengths.

Unfortunately the sparsity cannot contribute to further steps when using full Kalman filter. Despite the initial sparsity, the resulting posterior covariance would again be a full and dense matrix that cannot be solved for high-dimensional problems. A solution for the diagonal of the posterior covariance i.e. posterior variance is possible. In this respect the approach is on par with the earlier methods.

To take into account the covariance from one time step to another, methods such as ensemble Kalman filter should be considered. However, on a regional scale, the dynamical transitions in ionospheric electron density can be substantial even in short timescales. Hence, even if the previous posterior covariance was solved, the role of the poorly known process noise covariance at the prediction step can be significant. Especially, when considering ionospheric imaging in an operational manner, how much can be achieved by putting much effort into advancing the covariance temporally is a relevant question. On the other

hand, for understanding the uncertainty related to any ionospheric electron density reconstructions, the examination of posterior covariances is essential.

For numerical computations, an R (R Core Team, 2017) implementation of the algorithm, called TomoScand, was written. Similarly to other imaging methods it can use any ionospheric electron density model as its background. However, on a regional scale the ionospheric electron density models can sometimes be severely flawed. Even in a case with dense receiver networks, the data assimilation can struggle if the background information is critically misleading. Hence, recently only a persistence model with attenuation coefficient has been used. Instead, emphasis is placed on the modelling of the prior/background error covariance that controls the uncertainty associated with the unknown electron density distribution. The approach improves the performance in regional imaging of the high-latitude ionospheric dynamics, providing an extension for the local measurements such as ISR and ionosonde measurements.

References

- P. R. Amestoy, I. S. Duff, J. Koster, and J.-Y. J.-Y. L'Excellent. A Fully Asynchronous Multifrontal Solver Using Distributed Dynamic Scheduling. *SIAM Journal on Matrix Analysis and Applications*, 23(1):15–41, 2001. ISSN 0895-4798. doi: 10.1137/S0895479899358194.
- P. R. Amestoy, A. Buttari, J. Y. L'Excellent, and T. Mary. Performance and scalability of the block low-rank multifrontal factorization on multicore architectures. *ACM Transactions on Mathematical Software*, 45(1):1–26, 2019. ISSN 15577295. doi: 10.1145/3242094.
- A. H. Andersen and A. C. Kak. Simultaneous Algebraic Reconstruction Technique (SART): A Superior Implementation of the ART Algorithm. *Ultrasonic imaging*, 6:81–94, 1984. ISSN 15596915. doi: 10.1145/2387358.2387363.
- E. S. Andreeva. Radio tomographic reconstruction of ionization dip in the plasma near the Earth. *J. Exp. Theor. Phys. Lett.*, 52:142–148, 1990.
- M. J. Angling. Annales Geophysicae First assimilations of COSMIC radio occultation data into the Electron Density Assimilative Model (EDAM). *Annales Geophysicae*, pages 353–359, 2008.
- M. J. Angling and P. S. Cannon. Assimilation of radio occultation measurements into background ionospheric models. *Radio Science*, 39:1–11, 2004. doi: 10.1029/2002RS002819.
- M. J. Angling and N. K. Jackson-Booth. A short note on the assimilation of collocated and concurrent GPS and ionosonde data into the Electron Density Assimilative Model. *Radio Science*, 46(4):1–7, 2011. ISSN 00486604. doi: 10.1029/2010RS004566.
- M. J. Angling and B. Khattatov. Comparative study of two assimilative models of the ionosphere. *Radio Science*, 41(April):1–11, 2006. doi: 10.1029/2005RS003372.
- J. Austen, S. Franke, C. Liu, and K. Yeh. Application of computerized tomography techniques to ionospheric research. In A. Tauriainen, editor, *International Beacon Satellite Symposium on Radio Beacon Contribution to the Study of Ionization and Dynamics of*

- the Ionosphere and to Corrections to Geodesy and Technical Workshop, Proceedings. Part 1 (A87-50101 22-46).*, pages 25–35., Oulu, Finland, 1986. University of Oulu.
- J. R. Austen, S. J. Franke, and C. H. Liu. Ionospheric imaging using computerized tomography Jeffrey. *Radio Science*, 23(3):299–307, 1988.
- P. A. Bernhardt, C. A. Selcher, S. Basu, G. S. Bust, and S. C. Reising. Atmospheric studies with the tri-band beacon instrument on the COSMIC constellation. *Terrestrial, Atmospheric and Oceanic Sciences*, 11(1):291–312, 2000. ISSN 10170839.
- D. Bilitza. International Reference Ionosphere 2000. *Radio Science*, 36(2):261–275, 2001. doi: 10.1029/2000RS002432.
- D. Bilitza, K. Rawer, L. Bossy, and T. Gulyaeva. International reference ionosphere - past, present, and future: II. Plasma temperatures, ion composition and ion drift. *Advances in Space Research*, 13(3):15–23, 1993. ISSN 02731177. doi: 10.1016/0273-1177(93)90241-3.
- G. Breit and M. A. Tuve. A test of the existence of the conducting layer. *Physical Review*, 28(3):554–575, 1926. ISSN 0031899X. doi: 10.1103/PhysRev.28.554.
- A. Brekke. *Physics of the Upper Polar Atmosphere*. Wiley-Praxis Series in Atmospheric Physics. Wiley, 1997. ISBN 9780471960188.
- J. Bruno, C. N. Mitchell, K. H. Bolmgren, and B. A. Witvliet. A realistic simulation framework to evaluate ionospheric tomography. *Advances in Space Research*, 2019. ISSN 02731177. doi: 10.1016/j.asr.2019.11.015. URL <https://linkinghub.elsevier.com/retrieve/pii/S0273117719308233>.
- K. G. Budden. *Radio Waves in the Ionosphere: The Mathematical Theory of the Reflection of Radio Waves from Stratified Ionised Layers*. Cambridge University Press, Cambridge, 1st edition, 1961. ISBN 9780521043632.
- G. S. Bust and C. N. Mitchell. History, current state, and future directions of ionospheric imaging. *Rev. Geophys.*, 46, 2008. doi: 10.1029/2006RG000212.
- G. S. Bust, T. W. Garner, T. L. Gaussiran II, and T. L. Gaussiran. Ionospheric Data Assimilation Three-Dimensional (IDA3D): A global, multisensor, electron density specification algorithm. *J. Geophys. Res.*, 109(A11):1–14, 2004. ISSN 21699402. doi: 10.1029/2003JA010234.
- G. S. Bust, G. Crowley, T. W. Garner, T. L. Gaussiran II, R. W. Meggs, C. N. Mitchell, P. S. J. Spencer, P. Yin, and B. Zapfe. Four-dimensional GPS imaging of space weather storms. *Space Weather*, 5(2), 2007. ISSN 15427390. doi: 10.1029/2006SW000237.

- D. Calvetti and E. Somersalo. *Introduction to Bayesian Scientific Computing: Ten Lectures on Subjective Computing (Surveys and Tutorials in the Applied Mathematical Sciences)*. Springer-Verlag, Berlin, Heidelberg, 2007. ISBN 0387733930.
- A. Carrassi, M. Bocquet, L. Bertino, and G. Evensen. Data assimilation in the geosciences: An overview of methods, issues, and perspectives. *Wiley Interdisciplinary Reviews: Climate Change*, 9(5):1–79, 2018. ISSN 17577799. doi: 10.1002/wcc.535.
- Y. Censor. Finite Series-Expansion Reconstruction Methods. 71(3), 1983.
- S. Chapman. The absorption and dissociative or ionizing effect of monochromatic radiation in an atmosphere on a rotating earth part {I} and part {II}. Grazing incidence. *Proceedings of the Physical Society*, 43(5):483–501, sep 1931. doi: 10.1088/0959-5309/43/5/302.
- A. T. Chartier, C. N. Mitchell, and D. R. Jackson. A 12 year comparison of MIDAS and IRI 2007 ionospheric Total Electron Content. *Advances in Space Research*, 49(9):1348–1355, may 2012. ISSN 02731177. doi: 10.1016/j.asr.2012.02.014.
- F. F. Chen. Lecture Notes on Langmuir Probe Diagnostics. *IEEE-ICOPS meeting*, pages 1–40, 2003.
- G. Christakos. *Random Field Models in Earth Sciences*. Dover Publications, Mineola N.Y., 2005. ISBN 0-486-43872-4.
- A. J. Coster, J. Williams, A. Weatherwax, W. C. Rideout, and D. Herne. Accuracy of GPS total electron content: GPS receiver bias temperature dependence. *Radio Science*, 48(2):190–196, 2013. ISSN 00486604. doi: 10.1002/rds.20011.
- P. Courtier, J. Thépaut, and A. Hollingsworth. A strategy for operational implementation of 4D-Var, using an incremental approach. *Quarterly Journal of the Royal Meteorological Society*, 120(519):1367–1387, 1994. ISSN 00359009. doi: 10.1256/smsqj.51911.
- R. Daley. *Atmospheric Data Analysis*. Cambridge Atmospheric and Space Science Series. Cambridge University Press, 1991. ISBN 9780521382151.
- R. Daley and E. Barker. *NAVDAS Source Book. NRL Atmospheric Variational Data Assimilation System*. 2000.
- R. E. Daniell, L. D. Brown, D. N. Anderson, M. W. Fox, P. H. Doherty, D. T. Decker, J. J. Sojka, and R. W. Schunk. Parameterized ionospheric model: A global ionospheric parameterization based on first principles models. *Radio Science*, 30(5):1499–1510, 1995. ISSN 1944799X. doi: 10.1029/95RS01826.
- K. Davies. *Ionospheric Radio Propagation*. National Bureau of Standards Monograph, 1965.

- K. Davies. *Ionospheric Radio*. IEE electromagnetic waves series. Peregrinus on behalf of the Institution of Electrical Engineers, London, 1990. ISBN 9780863411861.
- L. Dyrud, A. Jovancevic, A. Brown, D. Wilson, and S. Ganguly. Ionospheric measurement with GPS: Receiver techniques and methods. *Radio Science*, 43(6):1–11, 2008. ISSN 00486604. doi: 10.1029/2007RS003770.
- S. Elvidge and M. J. Angling. Using the local ensemble Transform Kalman Filter for upper atmospheric modelling. *Journal of Space Weather and Space Climate*, 9:A30, 2019. ISSN 2115-7251. doi: 10.1051/swsc/2019018.
- G. Evensen. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research*, 99: 143–162, 1994.
- G. Evensen. The Ensemble Kalman Filter: Theoretical formulation and practical implementation. *Ocean Dynamics*, 53(4):343–367, 2003. ISSN 16167341. doi: 10.1007/s10236-003-0036-9.
- G. Evensen. *Data Assimilation, the Ensemble Kalman Filter*. Springer-Verlag, Dordrecht Heidelberg London New York, 2nd edition, 2009. ISBN 9783642037108. doi: 10.1007/978-3-642-03711-5.
- G. C. Fehmers. *Tomography of the Ionosphere*. 1996. ISBN 9789038604381.
- E. J. Fremouw, J. A. Secan, and B. M. Howe. Application of stochastic inverse theory to ionospheric tomography. *Radio Science*, 27(5):721–732, 1992. ISSN 1944799X. doi: 10.1029/92RS00515.
- S. V. Fridman, L. J. Nickisch, M. Aiello, and M. Hausman. Real-time reconstruction of the three-dimensional ionosphere using data from a network of GPS receivers. *Radio Science*, 41(5):1–7, 2006. ISSN 00486604. doi: 10.1029/2005RS003341.
- S. V. Fridman, L. J. Nickisch, and M. Hausman. Personal-computer-based system for real-time reconstruction of the three-dimensional ionosphere using data from diverse sources. *Radio Science*, 44(3):1–12, 2009. ISSN 00486604. doi: 10.1029/2008RS004040.
- L. C. Gardner, R. W. Schunk, L. Scherliess, J. J. Sojka, and L. Zhu. Global assimilation of ionospheric measurements-gauss markov model: Improved specifications with multiple data types. *Space Weather*, 12(12):675–688, 2014. ISSN 15427390. doi: 10.1002/2014SW001104.
- G. H. Golub and C. F. Van Loan. *Matrix Computations*. Johns Hopkins Studies in the Mathematical Sciences. Johns Hopkins University Press, fourth edi edition, 2013. ISBN 9781421407944.

- R. Gordon, R. Bender, and G. T. Herman. Algebraic Reconstruction Techniques (ART) for three-dimensional electron microscopy and X-ray photography. *Journal of Theoretical Biology*, 29(3):471–481, 1970. ISSN 10958541. doi: 10.1016/0022-5193(70)90109-8.
- W. E. Gordon. Incoherent Scattering of Radio Waves by Free Electrons with Applications to Space Exploration by Radar. *Proceedings of the IRE*, 46(11):1824–1829, 1958. ISSN 00968390. doi: 10.1109/JRPROC.1958.286852.
- B. N. Hahn. Dynamic linear inverse problems with moderate movements of the object: Ill-posedness and regularization. *Inverse Problems and Imaging*, 9(2):395–413, 2015. ISSN 19308345. doi: 10.3934/ipi.2015.9.395.
- G. A. Hajj, R. Ibañez-Meier, E. R. Kursinski, and L. J. Romans. Studying the ionosphere with the global positioning system. *International Journal of Imaging Systems & Technology*, 5(2):174–187, 1994. ISSN 00486604. doi: 10.1002/ima.1850050214.
- G. A. Hajj, B. D. Wilson, C. Wang, X. Pi, and I. G. Rosen. Data assimilation of ground GPS total electron content into a physics-based ionospheric model by use of the Kalman filter. *Radio Science*, 39(1):1–17, 2004. ISSN 0048-6604. doi: 10.1029/2002rs002859.
- M. Håkansson, A. B. Jensen, M. Horemuz, and G. Hedling. Review of code and phase biases in multi-GNSS positioning. *GPS Solutions*, 21(3):849–860, 2017. ISSN 15211886. doi: 10.1007/s10291-016-0572-7.
- J. K. Hargreaves. *The solar-terrestrial environment: An introduction to geospace - the science of the terrestrial upper atmosphere, ionosphere and magnetosphere*. Cambridge University Press, Cambridge, UK; New York, USA; Melbourne, Australia, 1992. ISBN 0521327482.
- J. A. T. Heaton, S. E. Pryse, and L. Kersley. Improved background representation, ionosonde input and independent verification in experimental ionospheric tomography. *Ann. Geophys.*, 13(1297-1302):1297–1302, 1995.
- A. E. Hoerl and R. W. Kennard. Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics*, 12(1):55–67, 1970. ISSN 15372723. doi: 10.1080/00401706.1970.10488634.
- I. Horvath and S. Crozier. Software developed for obtaining GPS-derived total electron content values. *Radio Science*, 42(2), 2007. ISSN 00486604. doi: 10.1029/2006RS003452.
- B. M. Howe, K. Runciman, and J. A. Secan. Tomography of the ionosphere: Four-dimensional simulations. *Radio Science*, 33(1):109–128, 1998. ISSN 00486604. doi: 10.1029/97RS02615.

- J. Hsieh. *Computed Tomography: Principles, Design, Artifacts, and Recent Advances*. SPIE, John Wiley & Sons, Bellingham, 2 edition, 2009. ISBN 0819444251. doi: 10.1117/3.817303.
- X. Huang and B. W. Reinisch. Vertical electron density profiles from the digisonde network. *Advances in Space Research*, 18(6):121–129, 1996. ISSN 02731177. doi: 10.1016/0273-1177(95)00912-4.
- B. R. Hunt, E. J. Kostelich, and I. Szunyogh. Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter. *Physica D: Nonlinear Phenomena*, 230(1-2):112–126, 2007. ISSN 01672789. doi: 10.1016/j.physd.2006.11.008.
- N. Hyvönen, M. Kalke, M. Lassas, H. Setälä, and S. Siltanen. Three-dimensional dental X-ray imaging by combination of panoramic and projection data. *Inverse Problems and Imaging*, 4(2):257–271, 2010. ISSN 19308337. doi: 10.3934/ipi.2010.4.257.
- N. Jakowski and M. M. Hoque. A new electron density model of the plasmasphere for operational applications and services. 2018.
- J. H. Justice, A. Vassiliou, S. Singh, J. D. Logel, P. A. Hansen, B. R. Hall, P. R. Hutt, and J. J. Solanki. Acoustic tomography for monitoring enhanced oil recovery. 1989.
- S. Kaczmarz. Angenäherte Auflösung von Systemen linearer Gleichungen (english translation by Stockman, J.). *Bulletin International de l'Académie Polonaise des Sciences et des Lettres. Classe des Sciences Mathématiques et Naturelles. Série A, Sciences Mathématiques.*, 35:355 – 7, 1937.
- J. Kaipio and E. Somersalo. *Statistical and Computational Inverse Problems*. Applied Mathematical Sciences. Springer, New York, 2005. ISBN 9780387271323.
- J. Kaipio and E. Somersalo. Statistical inverse problems: Discretization, model reduction and inverse crimes. *Journal of Computational and Applied Mathematics*, 198(2):493–504, 2007. ISSN 03770427. doi: 10.1016/j.cam.2005.09.027.
- A. C. Kak and M. Slaney. *Principles of Computerized Tomographic Imaging*. IEEE Press, New York, electronic edition, 1988. ISBN 978-0-89871-494-4. doi: 10.1118/1.1455742.
- R. E. Kalman. A new approach to linear filtering and prediction problems. *Journal of Fluids Engineering, Transactions of the ASME*, 82(1):35–45, 1960. ISSN 1528901X. doi: 10.1115/1.3662552.
- E. D. Kaplan and C. J. Hegarty. *Understanding GPS - Principles and Applications*. Artech House, Inc., Norwood, 2nd edition, 2006. ISBN 1-58053-894-0. doi: 10.1016/S1364-6826(97)83337-8.

- L. Kersley, J. A. T. Heaton, S. E. Pryse, and T. D. Raymund. Experimental ionospheric tomography with ionosonde input and EISCAT verification. *Annales Geophysicae*, 11: 1064–1074, 1993.
- J. A. Klobuchar. Ionospheric radio wave propagation. In A. S. Jursa, editor, *HANDBOOK OF GEOPHYSICS AND THE SPACE ENVIRONMENT*, chapter 10, pages 1–111. AIR FORCE GEOPHYSICS LABORATORY, AIR FORCE SYSTEMS COMMAND, UNITED STATES AIR FORCE, 4 edition, 1985.
- J. A. Klobuchar. Ionospheric effects on GPS positioning. In B. W. Parkinson and J. J. Spilker, editors, *Global Positioning System : Theory and Applications*, vol. 1, chapter 12, pages 485–515. American Institute of Aeronautics and Astronautics, Inc., 1996. ISBN 978-1-60086-638-8. doi: <https://doi.org/10.2514/4.866388>.
- V. E. Kunitsyn, E. S. Andreeva, S. J. Franke, and K. C. Yeh. Tomographic investigations of temporal variations of the ionospheric electron density and the implied fluxes. *Geophysical Research Letters*, 30(16):1851, 2003. ISSN 0094-8276. doi: 10.1029/2003gl016908.
- N. Lunt, L. Kersley, and G. J. Bailey. The influence of the protonosphere on GPS observations: model simulations. *Radio Science*, 34(3):725–732, 1999. ISSN 00486604. doi: 10.1029/1999RS900002.
- W. Menke. *Geophysical Data Analysis: Discrete Inverse Theory*. Academic Press, San Diego, revised ed edition, 1989. ISBN 0080507328.
- D. Minkwitz, K. G. Van Den Boogaart, T. Gerzen, and M. Hoque. Tomography of the ionospheric electron density with geostatistical inversion. *Annales Geophysicae*, 33:1071–1079, 2015. doi: 10.5194/angeo-33-1071-2015.
- D. Minkwitz, K. G. Van Den Boogaart, T. Gerzen, M. Hoque, and M. Hernández-pajares. Ionospheric tomography by gradient-enhanced kriging with STEC measurements and ionosonde characteristics. *Annales Geophysicae*, 34:999–1010, 2016. doi: 10.5194/angeo-34-999-2016.
- C. N. Mitchell and P. S. J. Spencer. A three-dimensional time-dependent algorithm for ionospheric imaging using GPS. *Annals of Geophysics*, 46(4):687–696, 2003. ISSN 15935213. doi: 10.4401/ag-4373.
- J. L. Mueller and S. Siltanen. *Linear and Nonlinear Inverse Problems with Practical Applications*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 2012. doi: 10.1137/1.9781611972344.
- H. Na, B. Hall, and E. Sutton. Ground station spacing effects in ionospheric tomography. *Annales Geophysicae*, 13:1288–1296, 1995.

- F. Natterer and F. Wübbeling. *Mathematical Methods in Image Reconstruction*. SIAM, Philadelphia, 2001. ISBN 0898716225. doi: 10.1118/1.1455744.
- B. Nava, P. Coisson, and S. M. Radicella. Atmospheric and Solar-Terrestrial Physics A new version of the NeQuick ionosphere electron density model. 70:1856–1862, 2008. doi: 10.1016/j.jastp.2008.01.015.
- B. W. Parkinson, J. J. Spilker Jr., M. Aparicio, P. Brodie, L. Doyle, J. Rajan, P. Torrione, A. J. Van Dierendonck, P. Axelrad, S. G. Francisco, J. A. Klobuchar, M. S. Braasch, J. F. Zumberge, W. I. Bertiger, F. Van Graas, N. Ashby, L. Kruczynski, and F. D. Natali. *Global Positioning System: Theory and Applications*, volume 1. American Institute of Aeronautics and Astronautics, Inc., Washington, 1996. ISBN 156347106X.
- M. Pezzopane and C. Scotto. Software for the automatic scaling of critical frequency f_0F_2 and MUF(3000)F2 from ionograms applied at the Ionospheric Observatory of Gibilmanna. *Annals of Geophysics*, 47(6):1783–1790, 2004. ISSN 15935213. doi: 10.4401/ag-3375.
- D. L. Phillips. A Technique for the Numerical Solution of Certain Integral Equations of the First Kind. *J. ACM*, 9(1):84–97, jan 1962. ISSN 0004-5411. doi: 10.1145/321105.321114.
- S. E. Pryse and L. Kersley. A preliminary experimental test of ionospheric tomography. *Journal of Atmospheric and Terrestrial Physics*, 54(7-8):1007–1012, 1992. ISSN 00219169. doi: 10.1016/0021-9169(92)90067-U.
- L. Qian, A. G. Burns, B. A. Emery, B. Foster, G. Lu, A. Maute, A. D. Richmond, R. G. Roble, S. C. Solomon, and W. Wang. TIE-GCM: A community model of the coupled thermosphere/ ionosphere system. In J. Huba, R. Schunk, and G. Khazanov, editors, *Modeling the ionosphere-thermosphere system*, pages 73–84. 2014. doi: 10.1029/2012GM001297.
- R Core Team. R: A Language and Environment for Statistical Computing, 2017. URL <https://www.r-project.org/>.
- C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning*. MIT Press, Cambridge, 2006. ISBN 026218253X. doi: 10.1142/S0129065704001899.
- T. D. Raymund. Ionospheric tomography algorithms. *International Journal of Imaging Systems and Technology*, 5(2):75–85, 1994. ISSN 10981098. doi: 10.1002/ima.1850050204.
- T. D. Raymund. Comparison of several ionospheric tomography algorithms. *Annales Geophysicae*, 13:1254–1262, 1995.
- T. D. Raymund, J. R. Austen, S. J. Franke, C. H. Liu, J. A. Klobuchar, and J. Stalker. Application of computerized tomography to the investigation of ionospheric structures. *Radio Science*, 25(5):771–789, 1990. ISSN 1944799X. doi: 10.1029/RS025i005p00771.

- T. D. Raymund, Y. Bresler, D. N. Anderson, and R. E. Daniell. Model-assisted ionospheric tomography: A new algorithm. *Radio Science*, 29(6):1493–1512, 1994a.
- T. D. Raymund, S. J. Franke, and K. C. Yeh. Ionospheric tomography: its limitations and reconstruction methods. *Journal of Atmospheric and Terrestrial Physics*, 56(5), 1994b. ISSN 00219169. doi: 10.1016/0021-9169(94)90104-X.
- B. W. Reinisch and X. Huang. Automatic calculation of electron density profiles from digital ionograms: 3. Processing of bottomside ionograms. *Radio Science*, 18(3):477–492, 1983. ISSN 1944799X. doi: 10.1029/RS018i003p00477.
- A. Rius, G. Ruffini, and L. Cucurull. Improving the vertical resolution of ionospheric tomography with GPS occultations. *Geophysical Research Letters*, 24(15):2291–2294, 1997. ISSN 0094-8276. doi: 10.1029/97gl52283.
- L. Roininen, M. S. Lehtinen, S. Lasanen, M. Orispää, and M. Markkanen. Correlation priors. *Inverse Probl. and Imag.*, 5(1):167–184, 2011. doi: 10.3934/ipi.2011.5.167.
- L. Roininen, P. Piironen, and M. S. Lehtinen. Constructing continuous stationary covariances as limits of the second-order stochastic difference equations. *Inverse Problems and Imaging*, 7(2):611–647, may 2013. ISSN 19308337. doi: 10.3934/ipi.2013.7.611.
- I. G. Rosen, C. Wang, G. A. Hajj, X. Pi, and B. Wilson. An adjoint method based approach to data assimilation for a distributed parameter model for the ionosphere. *Proceedings of the IEEE Conference on Decision and Control*, 5:4406–4408, 2001. ISSN 01912216. doi: 10.1109/CDC.2001.980895.
- H. Rue and L. Held. *Gaussian Markov Random Fields: Theory And Applications (Monographs on Statistics and Applied Probability)*. Chapman & Hall/CRC, 2005. ISBN 1584884320.
- J. M. Rüeger. Refractive Index Formulae for Radio Waves. *Integration of Techniques and Corrections to Achieve Accurate Engineering*, pages 1–13, 2002.
- E. Saksman, T. Nygrén, and M. Markkanen. Ionospheric structures invisible in satellite radiotomography. *Radio Science*, 32(2):605–616, 1997. ISSN 00486604.
- S. Särkkä. Linear operators and stochastic partial differential equations in Gaussian process regression. In *Artificial Neural Networks and Machine Learning - ICANN 2011 - 21st International Conference on Artificial Neural Networks, Espoo, Finland, June 14-17, 2011, Proceedings, Part II*, pages 1–8, 2011. doi: 10.1007/978-3-642-21738-8-20.
- S. Särkkä. *Bayesian filtering and smoothing*. Cambridge University Press, Cambridge, UK, 2013. ISBN 978-1-107-61928-9. doi: 10.1017/CBO9781139344203.

- L. Scherliess, R. W. Schunk, J. J. Sojka, and D. C. Thompson. Development of a physics-based reduced state Kalman filter for the ionosphere. *Radio Science*, 39(1):1–12, 2004. ISSN 0048-6604. doi: 10.1029/2002rs002797.
- L. Scherliess, R. W. Schunk, L. C. Gardner, J. V. Eccles, L. Zhu, and J. J. Sojka. The USU-GAIM-FP data assimilation model for ionospheric specifications and forecasts. In *2017 32nd General Assembly and Scientific Symposium of the International Union of Radio Science, URSI GASS 2017*, volume 2017-Janua, pages 1–4, 2017. ISBN 9789082598704. doi: 10.23919/URSIGASS.2017.8104978.
- R. W. Schunk, L. Scherliess, J. J. Sojka, D. C. Thompson, D. N. Anderson, M. Codrescu, C. Minter, T. J. Fuller-Rowell, R. A. Heelis, M. Hairston, and B. M. Howe. Global Assimilation of Ionospheric Measurements (GAIM). *Radio Science*, 39(1):1–11, 2004. ISSN 0048-6604. doi: 10.1029/2002rs002794.
- E. Sutton and H. Na. Comparison of geometries for ionospheric tomography. 30(1):115–125, 1995.
- K. Tanabe. Projection method for solving a singular system of linear equations and its applications. *Numerische Mathematik*, 17(3):203–214, 1971. ISSN 0029599X. doi: 10.1007/BF01436376.
- A. Tarantola. *Methods for Data Fitting and Parameter Estimation*. Inverse Problem Theory. Elsevier Science Publishers B. V., Netherlands, 1987.
- A. Tarantola and B. Valette. Generalized nonlinear inverse problems solved using the least squares criterion. *Reviews of Geophysics*, 20(2):219, 1982.
- A. N. Tikhonov and V. Y. Arsenin. *Solutions of ill-posed problems*. V. H. Winston & Sons, Washington, D.C.: John Wiley & Sons, New York, 1977.
- J. E. Titheridge. Ionogram analysis with the generalised program POLAN. Technical report, WORLD DATA CENTER A for Solar-Terrestrial Physics, Auckland, New Zealand, 1985.
- J. Vierinen, J. Norberg, M. S. Lehtinen, O. Amm, L. Roininen, A. Väänänen, P. J. Erickson, and D. McKay-Bukowski. Beacon satellite receiver for ionospheric tomography. *Radio Science*, 49(12):1141–1152, 2014. ISSN 1944799X. doi: 10.1002/2014RS005434.
- J. Vierinen, A. J. Coster, W. C. Rideout, P. J. Erickson, and J. Norberg. Statistical framework for estimating GNSS bias. *Atmospheric Measurement Techniques*, 9(3):1303–1312, 2016. ISSN 18678548. doi: 10.5194/amt-9-1303-2016.
- C. Wang, G. A. Hajj, X. Pi, I. G. Rosen, and S. Wing. Development of the Global Assimilative Ionospheric Model. *Radio Science*, 39(1):1–11, 2004. ISSN 00486604. doi: 10.1029/2002RS002854.

- D. Wells, N. Beck, D. Delikaraoglou, A. Kleusberg, E. J. Krakiwsky, G. Lachapelle, R. B. Langley, M. Nakiboglu, K.-P. Schwarz, J. M. Tranquilla, and P. Vaníček. *Guide to GPS Positioning*. CANADIAN GPS ASSOCIATES, New Brunswick, Canada, 1986. ISBN 0920114733.
- M. Yamamoto. Digital beacon receiver for ionospheric TEC measurement. *Earth, Planets and Space*, 60(3):21–24, 2008. ISSN 1343-8832. doi: 10.1186/BF03353137.
- K. C. Yeh and T. D. Raymund. Limitations of ionospheric imaging by tomography. *Radio Science*, 26(6):1361–1380, 1991. ISSN 1944799X. doi: 10.1029/91RS01873.
- T. P. Yunck, G. F. Lindal, and C.-H. Liu. The Role of GPS in Precise Earth Observation. In *IEEE PLANS '88., Position Location and Navigation Symposium, Record. 'Navigation into the 21st Century'*, pages 251–258, Orlando, FL, USA, USA, 1988. IEEE. doi: 10.1109/PLANS.1988.195491.
- N. A. Zabolotin, J. W. Wright, and G. A. Zhabankov. NeXtYZ: Three-dimensional electron density inversion for dynasonde ionograms. *Radio Science*, 41(6):1–12, dec 2006. ISSN 00486604. doi: 10.1029/2005RS003352.

Publication I

J. Norberg, L. Roininen, J. Vierinen, O. Amm, D. McKay-Bukowski, and M. S. Lehtinen, Ionospheric tomography in Bayesian framework with Gaussian Markov random field priors, *Radio Sci.*, 50(2): 138–152, 2015. ISSN 1944799X. doi: 10.1002/2014RS005431.

RESEARCH ARTICLE

10.1002/2014RS005431

Special Section:

Beacon Satellite Symposium
2013

This article is a companion
to Vierinen et al. [2014]
doi:10.1002/2014RS005434.

Key Points:

- We present a novel ionospheric tomography reconstruction method
- The method is based on Bayesian inference with the use of GMRF priors
- The prior distribution is built based on physical and empirical information

Correspondence to:

J. Norberg,
johannes.norberg@fmi.fi

Citation:

Norberg, J., L. Roininen, J. Vierinen, O. Amm, D. McKay-Bukowski, and M. Lehtinen (2015), Ionospheric tomography in Bayesian framework with Gaussian Markov random field priors, *Radio Sci.*, 50, 138–152, doi:10.1002/2014RS005431.

Received 14 MAR 2014

Accepted 21 JAN 2015

Accepted article online 28 JAN 2015

Published online 18 FEB 2015

Ionospheric tomography in Bayesian framework with Gaussian Markov random field priors

J. Norberg¹, L. Roininen², J. Vierinen³, O. Amm¹, D. McKay-Bukowski², and M. Lehtinen²
¹Finnish Meteorological Institute, Helsinki, Finland, ²Sodankylä Geophysical Observatory, University of Oulu, Sodankylä, Finland, ³Haystack Observatory, Massachusetts Institute of Technology, Westford, Massachusetts, USA

Abstract We present a novel ionospheric tomography reconstruction method. The method is based on Bayesian inference with the use of Gaussian Markov random field priors. We construct the priors as a system of stochastic partial differential equations. Numerical approximations of these equations can be represented with linear systems with sparse matrices, therefore providing computational efficiency. The method enables an interpretable scheme to build the prior distribution based on physical and empirical information on the structure of the ionosphere. We show through synthetic test cases in a two-dimensional setup of latitude-altitude slices how this method can be applied to satellite-based ionospheric tomography and how information about the structure of the ionosphere can be implemented in the prior. The technique is capable of being easily extended to multifrequency tomographic analysis or used for the inclusion of other data sets of ionospheric electron density, such as ground-based observations by radars or ionosondes.

1. Introduction

The object of ionospheric tomography is to study the electron density of Earth's upper atmosphere between altitudes of 60 and 1000 km. The idea was originally proposed by Austen et al. [1988], and the statistical approach for the problem was first presented by Fremouw et al. [1992]. The first to utilize Bayesian statistical theory and Gaussian prior information for ionospheric tomography was Markkanen et al. [1995]. Later studies in the field are concentrated mostly in the three-dimensional cases and real-time estimations [Hansen, 2002; Mitchell and Spencer, 2003] where the problem is essentially the same as that in two dimensions, but the sparseness of the data and increasing computational demand become even more crucial factors. More recently, Seemala et al. [2014] have used algorithm similar to Markkanen et al. [1995] for three-dimensional ionospheric tomography. The method development of ionospheric tomography is well documented by Bust and Mitchell [2008]. Standard ionospheric tomography techniques are covered in text books, such as Kunitsyn and Tereshchenko [2003].

The main data component in ionospheric tomography is usually provided with ground-based measurements of signals transmitted by Global Positioning System (GPS) satellites or Low Earth Orbit (LEO) beacon satellites. Here we concentrate on a simple case that models the overflight of one LEO satellite and is illustrated in Figure 1. However, the method presented here can utilize various types of ionospheric measurements.

The unknown object of interest is the electron density in the ionosphere. We assume that it is constant in time during the flyby, which for an overflight of a LEO satellite over a regional tomographic receiver network on the ground, like the TomoScand receivers in Fenno-Scandinavia, is approximately 15 min Vierinen et al. [2014].

Mathematically, ionospheric tomography is an ill-posed problem which can be considered a sparse limited-angle tomography and an inverse problem [Kaipio and Somersalo, 2005]. The low information provided by the measurement setup in ionospheric tomography has been discussed in several articles. The restrictions caused by the limited elevation angle and limited number of receivers have been studied by Yeh and Raymund [1991], and the connection between measurement geometry and possible lattice resolutions are analyzed by Na and Sutton [1994]. Saksman et al. [1997] show that the ionosphere can contain structures that remain completely invisible for a certain class of receiver geometries.

As the information provided by the receiver-satellite geometry is very limited, the unknown electron density cannot be solved without stabilizing/regularizing the problem with some additional information.

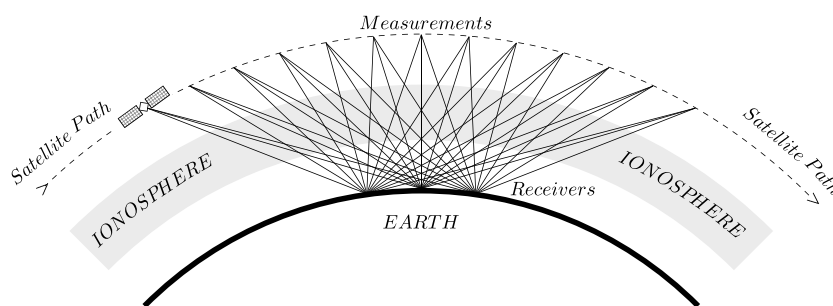


Figure 1. Schematic plot of ionospheric tomography.

The most commonly used methods in ionospheric tomography include iterative methods such as the Algebraic Reconstruction Technique [Andreeva, 1990] and similar iterative techniques [Raymund *et al.*, 1990; Pryse *et al.*, 1993], use of base functions in the vertical direction [Na and Lee, 1990; Fremouw *et al.*, 1992], and smoothness constraints [Markkanen *et al.*, 1995].

Raymund [1995] compared different algorithms and their variations for ionospheric tomography. While there were significant differences between the methods, it was also reported that the well performing algorithms did not have consistency over different cases. As one calibration of the model can produce good solution with some data, it can be very unsuitable under different conditions.

As the ionospheric tomography reconstructions are dominated by the additional information of the algorithm, it is important to understand what this information is and how strongly it affects the solution. In the Bayesian statistical approach, the stabilization for the inverse problem is given as an *a priori probability distribution*. The *a priori probability distribution* contains the information on the unknown parameters one has before actual measurements. The solution, i.e., the *a posteriori probability distribution* is constructed from the *a priori distribution* and the likelihood density functions. The latter, given the measurements, states the likelihood of unknown parameters. For the remainder of the paper we refer to these probability distributions simply as *prior* and *posterior* and to parameters of these distributions with the premodifiers *prior* and *posterior*.

Markkanen *et al.* [1995] used Bayesian approach for ionospheric tomography where the stabilization was provided by constraining the differences between neighboring pixels with an altitude-dependent regularization parameter profile. This approach gives a computationally efficient way to stabilize the problem by requiring smoothness from the reconstruction. However, this kind of smoothness constraint does not result in a proper prior probability distribution; i.e., the covariance matrix does not exist, and the regularizing effect is controlled with a regularization parameter that has no clear physical interpretation.

Natterer [1986] considered general computerized tomography where an isotropic exponential model can be understood as a prior covariance function for a Radon transform problem. For ionospheric tomography, Arikani *et al.* [2007] used similar approach with Gaussian random field priors with proper covariance matrices. This method is a truly Bayesian approach for ionospheric tomography, where the prior information can be given as a proper distribution. Instead of hiding the stabilization inside the algorithm, it is given with a clear physical interpretation. In practice, the prior distribution can be formed by choosing the prior mean and covariance structure based on, e.g., International Reference Ionosphere model, Chapman profile, other ionospheric models, or additional measurements such as profiles from ionosonde measurements. The variance of the system should then reflect the credibility of the given parameters.

Despite the evident advances of the Bayesian approach, the shortcoming, that was remarked also by Arikani *et al.* [2007], is that due to the resulting full covariance matrix, the problem becomes easily computationally unfeasible. This is the case especially in the three-dimensional ionospheric tomography where the number of unknown parameters rises rapidly.

Here we follow Roininen *et al.* [2011, 2013, 2014] to utilize similar smoothness constraints as in Markkanen *et al.* [1995] but parametrized in a more advanced manner so that the given constraints form a Gaussian Markov random field (GMRF). The GMRFs are implemented with sparse matrices providing computational efficiency, yet the approach provides a known and controllable covariance structure. Hence, the prior

probability distribution can be defined by prior mean and covariance, which are given as physical quantities in units of electrons per volume and length.

The GMRF prior provides also discretization invariance, i.e., the information given with the prior, and hence, the resulting reconstructions are essentially not affected by the change of the lattice used for the modeling. However, most importantly, the GMRFs provide a computationally efficient way for truly Bayesian statistical inversion with proper prior distributions. To our knowledge, this is the first time that GMRF priors are used to address the problem of ionospheric tomography.

The paper is organized as follows: in section 2, we review the basic concepts of the Bayesian statistical inversion. The construction of prior distribution with GMRF is explained in section 3, and the altitude information and the corresponding parametric fields for GMRFs are covered in section 4. In section 5 we show how the Bayesian statistical inversion with the GMRF priors is applied to ionospheric tomography in practice. Following section 6, which has our conclusions and discussion, we briefly consider the future plans on the topic in section 7. The first results with real measurement data are presented in the companion paper *Vierinen et al. [2014]*.

2. Bayesian Statistical Inversion

Figure 1 shows a schematic plot of ionospheric tomography. The dual-frequency signal is transmitted from the beacon satellites and received on the ground. The relative difference of ionospheric propagation delays of the two different frequencies can be measured tracking the phase of the two carriers. When subtracting one phase curve from a scaled version of the other, the contribution of tropospheric refraction, clock drifts, and uncertainties in the satellite position to measurements are canceled out and the relative electron content along the signal path remains, up to an unknown constant. The actual measurement can then be modeled as a line integral of electron density

$$m_t = \gamma_i + \int_{L_t} N_e(z) dz_t, \quad (1)$$

where m_t is the measured relative total electron content, t is the time, $N_e(z)$ the electron density of the domain, and $z := (z_1, z_2) \in \mathbb{R}^2$ the coordinate. The integral is defined over the measurement signal path L_t with coordinates z_t . The unknown constant γ_i is due to the 2π ambiguity of the initial phase difference and the index i refers to a certain receiver-satellite combination. For details see *Davies [1965]*. We assume a ray-casting approximation for the paths, i.e., the path from satellite to receiver is always a straight line. We call equation (1) the continuous observation model.

When the electron density and phase constants are known, we call the solution for m_t of equation (1) the forward problem. The inverse problem is then to solve the electron density $N_e(z)$ and phase constants γ_i , when other parameters of the forward model are known.

For computations, we need a discrete approximation of the continuous observation model (1). We set $z_{j,k} := (z_{1,j}, z_{2,k}) := (jh_1, kh_2)$, where h_1 and h_2 are the discretization lengths, i.e., the horizontal and vertical dimensions of the pixel in the discretized two-dimensional domain, row index $j = 1, 2, \dots, n_{\text{row}}$ and column index $k = 1, 2, \dots, n_{\text{col}}$. It is well known that a discretization of a linear integral equation can be obtained by a discrete linear approximation. Let us write the approximation as a linear matrix equation

$$\mathbf{m} = \mathbf{A}\mathbf{x} \quad (2)$$

where $\mathbf{m} := (m_1, \dots, m_{n_m})^T$ is the measurement vector containing the measured electron contents, \mathbf{A} the theory matrix, and the vector $\mathbf{x} := (N_e(z_{1,1}), \dots, N_e(z_{n_{\text{row}}, n_{\text{col}}}), \gamma_1, \dots, \gamma_n)^T$ contains the unknown electron densities and phase constants. The index n_m is the total number of measurements and n the total number of phase constants. The theory matrix \mathbf{A} is formed according to the measurement geometry, and it approximates the line integrals given in equation (1). Here the matrix $\mathbf{A} \in \mathbb{R}^{(n_{\text{row}} n_{\text{col}}) \times (n_{\text{row}} n_{\text{col}} + n_\gamma)}$ also includes the design matrix part which adds a correct γ_i parameter for the corresponding measurement.

In reality, the measurements are noisy. The noise can come from various sources, for example, radio frequency interference, thermal noise of the measurement device itself, and from small scale-changes in plasma that cannot be captured within the model resolution. For more detailed discussion on the error sources we refer to our companion paper *Vierinen et al. [2014]*.

In statistical inverse problems also all the other variables are viewed as random variables. The randomness describes our degree of information concerning their realizations [Kaipio and Somersalo, 2005]. Hence, we change the paradigm from a deterministic matrix model of equation (2) to a stochastic matrix model

$$\mathbf{M} = \mathbf{A}\mathbf{X} + \mathbf{e}, \quad (3)$$

where $\mathbf{e} = (e_1, \dots, e_{n_m})^T$ is the measurement noise vector and all vectors \mathbf{M} , \mathbf{X} , and \mathbf{e} are considered random variables. For the observed measurements we use \mathbf{m} as a realization of \mathbf{M} . We assume that the \mathbf{e} is normally distributed with zero mean and Σ covariance. We denote this as $\mathbf{e} \sim \mathcal{N}(\mathbf{0}, \Sigma)$. We also assume that we know the covariance matrix Σ , but we do not know realizations of the noise \mathbf{e} pathwise. In practice the mean and covariance of the measurement error distribution are estimated from the data.

The solution to the Bayesian statistical inverse problem is the posterior probability distribution given with the Bayes' formula [Kaipio and Somersalo, 2005].

$$D_{\text{post}}(\mathbf{X}) = \frac{D(\mathbf{M}|\mathbf{X})D_{\text{pr}}(\mathbf{X})}{D(\mathbf{M})} \propto D(\mathbf{M}|\mathbf{X})D_{\text{pr}}(\mathbf{X}) \propto \exp\left(-\frac{1}{2}(\mathbf{M} - \mathbf{A}\mathbf{X})^T \Sigma^{-1}(\mathbf{M} - \mathbf{A}\mathbf{X})\right) D_{\text{pr}}(\mathbf{X}),$$

where we use \propto for proportionality, D is a Gaussian normal density function with subscript "post" for posterior and "pr" for prior. $D(\mathbf{M}|\mathbf{X})$ is a normal density function for a conditional distribution of \mathbf{M} when \mathbf{X} is given, and here it is obtained from the discrete approximation of equation (3). $D(\mathbf{M}|\mathbf{X})$ can also be understood as a likelihood function for realizations of parameter \mathbf{X} given observed measurements $\mathbf{M} = \mathbf{m}$.

From now on, we assume that the continuous unknown electron density lies in \mathbb{R}^2 . We further assume that the prior distribution of the discrete unknown electron density \mathbf{X} is a Gaussian field, i.e., a multivariate normal distribution with a mean value μ and a covariance Σ_{pr} . Then the unnormalized posterior density is

$$D_{\text{post}}(\mathbf{X}) \propto \exp\left(-\frac{1}{2}\left((\mathbf{m} - \mathbf{A}\mathbf{X})^T \Sigma^{-1}(\mathbf{m} - \mathbf{A}\mathbf{X}) + (\mathbf{X} - \mu)^T \Sigma_{\text{pr}}^{-1}(\mathbf{X} - \mu)\right)\right),$$

from which we can calculate the maximum a posteriori (MAP) estimator, i.e., the most probable state of the ionosphere, given the prior information and measurements. We give the maximum a posteriori estimator by the matrix equation

$$\mathbf{x}_{\text{MAP}} = \Sigma_{\text{post}} \left(\mathbf{A}^T \Sigma^{-1} \mathbf{m} + \Sigma_{\text{pr}}^{-1} \mu \right), \quad (4)$$

where

$$\Sigma_{\text{post}} = \left(\mathbf{A}^T \Sigma^{-1} \mathbf{A} + \Sigma_{\text{pr}}^{-1} \right)^{-1} \quad (5)$$

is the posterior covariance estimator. We note that the maximum a posteriori estimator is analytically the same as the generalized Tikhonov regularized solution but, in this case, carries a statistical interpretation [Kaipio and Somersalo, 2005].

The prior covariance matrix Σ_{pr} can be formed with a desired structure as in Natterer [1986] and Arikan et al. [2007]; however, the resulting covariance matrix Σ_{pr} is a full matrix, and therefore, the computation of the posterior covariance becomes computationally heavy.

Let us assume that we know a sparse matrix \mathbf{L} which satisfies $\mathbf{L}^T \mathbf{L} = \Sigma_{\text{pr}}^{-1}$. Then we do not need to invert the full covariance matrix. On the other hand, we can then represent the same information as a virtual measurement equation

$$\mathbf{L}(\mathbf{X} - \mu) = \mathbf{W} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}) \quad (6)$$

where the vector \mathbf{W} is normally distributed, the mean $\mathbf{0}$ a vector of zeros, and the covariance \mathbf{I} an identity matrix. Then the model can be written as a system of matrix equations

$$\begin{bmatrix} \mathbf{M} \\ \mathbf{L}\mu \end{bmatrix} = \begin{bmatrix} \mathbf{A} \\ \mathbf{L} \end{bmatrix} \mathbf{X} + \begin{bmatrix} \mathbf{e} \\ \mathbf{W} \end{bmatrix}. \quad (7)$$

It is then trivial to note that the weighted least squares solution of the system of equations (7) gives the maximum a posteriori estimate of equation (4). This is a critical point in the development of the method expounded in this paper, as we can present the full covariance matrix by a sparse matrix \mathbf{L} . In the following two sections, we consider how to construct this matrix \mathbf{L} as a sparse matrix.

3. Gaussian Markov Random Field Prior

The main goal of this article is to build the matrix Σ_{pr}^{-1} of equation (4) as a sparse matrix, but in such a way that it together with the prior mean μ defines a proper normal distribution with known covariance structure. Here we do this with the formalism of the GMRFs.

The GMRF can be understood as a multivariate normal distribution, but instead of mean and covariance, the distribution is defined by its mean and precision matrix, i.e., the inverse covariance matrix Σ_{pr}^{-1} . Here we concentrate on cases where the positive definite precision matrix is sparse. The zeros in the precision matrix indicate that the corresponding unknown is conditionally independent from these elements, given the nonzero entries. This is known as the Markov property. For a detailed definition of GMRFs we refer to *Rue and Held [2005]*.

Here we follow *Roininen et al. [2011, 2013]* to build the matrix \mathbf{L} of precision matrix $\mathbf{L}^T \mathbf{L} = \Sigma_{\text{pr}}^{-1}$ in (4) so that the resulting GMRF covariance matrix approximates a continuous Gaussian-shaped covariance function. We first define and parametrize the distribution we are aiming at and then show how the corresponding matrix \mathbf{L} is formed.

Let $X(z)$, $z \in \mathbb{R}^2$ be a real-valued Gaussian random function with finite moments of second order. We set

$$\begin{aligned} \mathbf{E}(X(z)) &= \mu(z) = \mu, \\ C(s) &:= \mathbf{E}(X(z)X(y)) = \mathbf{E}(X(z-y)X(0)), \end{aligned}$$

where \mathbf{E} stands for the expectation operator and $s := (s_1, s_2) := z - y$, and $s, z, y \in \mathbb{R}^2$. Let us assume that $C(s)$ is an absolutely integrable function, then $X(z)$ is a weakly stationary random field. We then choose the prior distribution to be a weakly stationary Gaussian random field with mean μ and a squared exponential autocovariance function

$$C(s) = \alpha \exp\left(-\left(\frac{s_1^2}{l_1^2} + \frac{s_2^2}{l_2^2}\right)\right), \quad (8)$$

We call parameter α the scaling factor of the variance, and l_1, l_2 correlation lengths to respective coordinate directions.

We then present how the corresponding GMRF is constructed. We first define the discretized white noise as a discrete approximation of the continuous white noise W at a lattice point (j, k)

$$W_{j,k} \sim \mathcal{N}\left(0, \frac{1}{h_1 h_2}\right), \quad (9)$$

where h_1 and h_2 are the discretization lengths, i.e., the horizontal and vertical pixel dimensions of an uniform lattice. We denote the finite approximation of the continuous mean μ by vector $\bar{\mu} = (\bar{\mu}_1, \dots, \bar{\mu}_N)$. Then, according to *Roininen et al. [2011, 2013]*, we can approximate the continuous spatial differential operator equations as a system of stochastic partial difference equations

$$\begin{cases} X_{j,k} - \bar{\mu} = \sqrt{\frac{al_1 l_2}{c_0 h_1 h_2}} W_{j,k}^{(0,0)} \\ X_{j,k} - X_{j-1,k} = \sqrt{\frac{ah_1 l_2}{c_1 l_1 h_2}} W_{j,k}^{(1,1)} \\ X_{j,k} - X_{j,k-1} = \sqrt{\frac{ah_1 l_2}{c_1 l_1 h_2}} W_{j,k}^{(1,2)} \\ \frac{l_1^2}{h_2^2} (X_{j+1,k} - 2X_{j,k} + X_{j-1,k}) + \\ \frac{l_2^2}{h_1^2} (X_{j,k+1} - 2X_{j,k} + X_{j,k-1}) = \sqrt{\frac{al_1 l_2}{c_2 h_1 h_2}} W_{j,k}^{(2,0)}, \end{cases} \quad (10)$$

where $W^{(p,p')}$ are independent discrete white noise processes as in (9), where the $p \in \{0, 1, 2, \dots\}$ gives the order of differences and $p' \in \{0, 1, 2\}$ indicates the direction of differences with correspondence 1 = horizontal, 2 = vertical, and 0 = both directions; e.g., $W^{(1,2)}$ is a white noise process corresponding to first-order differences in vertical direction. Parameters c_p are known scaling constants chosen as explained in *Roininen et al. [2013]*.

In *Roininen et al.* [2011, 2013] the equations in (10) were defined on whole space \mathbb{R} or \mathbb{R}^2 . For computations we change the paradigm from the whole space to a bounded domain. This is done via the introduction of periodic boundary conditions

$$X_{j,k} = X_{j+n_{\text{row}},k}, \text{ and } X_{j,k} = X_{j,k+n_{\text{col}}}. \quad (11)$$

Finally, we can write the discrete approximation in equation (10) as a stacked matrix equation corresponding to equation (6) as

$$\begin{bmatrix} \mathbf{L}^{(0,0)} \\ \mathbf{L}^{(1,1)} \\ \mathbf{L}^{(1,2)} \\ \mathbf{L}^{(2,0)} \end{bmatrix} (\mathbf{X} - \boldsymbol{\mu}) = \mathbf{W} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}), \quad (12)$$

where the matrices $\mathbf{L}^{(p,p')} \in \mathbb{R}^{n_{\text{row}} n_{\text{col}} \times n_{\text{row}} n_{\text{col}}}$ are defined as follows: The white noise matrix

$$\mathbf{L}^{(0,0)} = \sqrt{\frac{c_0 h_1 h_2}{\alpha l_1 l_2}} \mathbf{I}.$$

Let us denote by \otimes the Kronecker product and an identity matrix by $\mathbf{I}_{n_{\text{row}} \times n_{\text{row}}} \in \mathbb{R}^{n_{\text{row}} \times n_{\text{row}}}$. Then the first-order difference matrices are

$$\mathbf{L}^{(1,1)} = \sqrt{\frac{c_1 h_1 h_2}{\alpha h_1 l_2}} \mathbf{I}_{n_{\text{row}} \times n_{\text{row}}} \otimes \begin{bmatrix} -1 & 1 & & \\ & \ddots & \ddots & \\ & & -1 & 1 \\ 1 & & & -1 \end{bmatrix}_{n_{\text{col}} \times n_{\text{col}}}$$

and

$$\mathbf{L}^{(1,2)} = \sqrt{\frac{c_1 h_1 l_2}{\alpha l_1 h_2}} \begin{bmatrix} -1 & 1 & & \\ & \ddots & \ddots & \\ & & -1 & 1 \\ 1 & & & -1 \end{bmatrix}_{n_{\text{row}} \times n_{\text{row}}} \otimes \mathbf{I}_{n_{\text{col}} \times n_{\text{col}}}.$$

The second-order differences can be presented as

$$\mathbf{L}^{(2,0)} = \sqrt{\frac{c_2 h_1 h_2}{\alpha l_1 l_2}} \left(\mathbf{I}_{n_{\text{row}} \times n_{\text{row}}} \otimes \begin{bmatrix} -2 & 1 & & 1 \\ 1 & -2 & 1 & \\ & \ddots & \ddots & \\ & & 1 & -2 & 1 \\ 1 & & & 1 & -2 \end{bmatrix}_{n_{\text{col}} \times n_{\text{col}}} + \begin{bmatrix} -2 & 1 & & 1 \\ 1 & -2 & 1 & \\ & \ddots & \ddots & \\ & & 1 & -2 & 1 \\ 1 & & & 1 & -2 \end{bmatrix}_{n_{\text{row}} \times n_{\text{row}}} \otimes \mathbf{I}_{n_{\text{col}} \times n_{\text{col}}} \right).$$

In *Roininen et al.* [2013] it is shown that by taking the order of difference matrices and the discretization interval into the limit, i.e., $p \rightarrow \infty$ and $h_1, h_2 \rightarrow 0$, the resulting covariance converges toward the continuous Gaussian-shaped covariance function given in equation (8). However, it is numerically verified that the matrices $\mathbf{L}^{(p,p')}$ can be used to form an adequate discrete approximation for this covariance already with $\max(p) = 2$, when h_1 and h_2 are shorter enough than the desired correlation lengths of the field.

As the discretization length is taken into account, the prior is discretization invariant and essentially the same information can be used for different lattice sizes. Also, the construction of priors on irregular lattices can be done similarly to the method used in this section. Any regular finite-difference approximation can be used in these cases. For the sake of notational clarity, we will not go through the construction of these priors on irregular lattices.

In *Roininen et al.* [2014] a similar idea was applied to a nonlinear Bayesian inverse problem, the electrical impedance tomography, but with finite element methods.

4. Altitude Information for the GMRF Prior

The electron density of the ionosphere varies spatially. For that, we change the constant parameters μ, α, l_1, l_2 introduced in the previous section to continuous, differentiable, and positive functions $\mu(z), \alpha(z), l_1(z), l_2(z)$ for locations z , where $z \in \mathbb{R}^2$.

Then the discrete GMRF approximation is

$$\begin{cases} X_{j,k} - \mu_{j,k} = \sqrt{\frac{\alpha_{j,k} l_{1,j,k} l_{2,j,k}}{c_0 h_1 h_2}} W_{j,k}^{(0)} \\ (X_{j,k} - \mu_{j,k}) - (X_{j-1,k} - \mu_{j-1,k}) = \sqrt{\frac{\alpha_{j,k} h_1 l_{2,j,k}}{c_1 l_{1,j,k} h_2}} W_{j,k}^{(1,1)} \\ (X_{j,k} - \mu_{j,k}) - (X_{j,k-1} - \mu_{j,k-1}) = \sqrt{\frac{\alpha_{j,k} l_{1,j,k} h_2}{c_1 h_1 l_{2,j,k}}} W_{j,k}^{(1,2)} \\ \frac{l_{1,j,k}^2}{h_1^2} ((X_{j+1,k} - \mu_{j+1,k}) - 2(X_{j,k} - \mu_{j,k}) + (X_{j-1,k} - \mu_{j-1,k})) + \frac{l_{2,j,k}^2}{h_2^2} ((X_{j,k+1} - \mu_{j,k+1}) - 2(X_{j,k} - \mu_{j,k}) \\ + (X_{j,k-1} - \mu_{j,k-1})) = \sqrt{\frac{\alpha_{j,k} l_{1,j,k} l_{2,j,k}}{c_1 h_1 h_2}} W_{j,k}^{(2)}, \end{cases} \quad (13)$$

where the parameters $\mu_{j,k}, \alpha_{j,k}, l_{1,j,k}$, and $l_{2,j,k}$ are the values of the corresponding continuous functions at the location $z_{j,k}$. The approximation can then be presented with a similar stacked matrix representation as in equation (12).

In section 5, we show through numerical simulations that in practice the prior distribution can be determined with the parameters $\mu_{j,k}, \alpha_{j,k}, l_{1,j,k}$, and $l_{2,j,k}$ and that the effect of the parameters is interpretable. With the resulting GMRF prior, we can model structural properties of the ionosphere; i.e., we have an inhomogeneous prior. This approach is somewhat similar to the work of Bardsley [2013] and to Kaipio *et al.* [1999], in that we model the structural properties of the priors.

We note that from the purely mathematical point of view, we have not verified that there is a continuous GMRF prior in the discretization limit $h_1, h_2 \rightarrow 0$. However, numerical simulations support this idea, and we conjecture that the resulting GMRF prior is discretization invariant. Hence, the reconstructions computed are essentially independent of the computational lattice used.

5. Numerical Examples

In this section we show how the prior distributions are built with GMRFs and demonstrate with synthetical examples how the Bayesian approach with GMRF priors is implemented for ionospheric tomography. The first results with real data are presented in our companion paper Vierinen *et al.* [2014].

In the examples, we first simulate samples from the constructed prior distributions to show that the distributions actually incorporate the information given with the parameters of previous sections. We then show how well the unknown electron density, with a certain altitude characteristics and scale size of structures, can be reconstructed in an optimal case where we know the exact prior distribution. However, as with real measurements, the exact prior distribution is never actually known; we then show how well the unknown electron density can be reconstructed with a loose prior that does not restrict the solution with strict information that we actually do not have.

We take the theory presented in previous sections and build the inverted prior covariance matrix with matrix

$$\mathbf{L} = \begin{bmatrix} \mathbf{L}^{(0,0)} \\ \mathbf{L}^{(1,1)} \\ \mathbf{L}^{(1,2)} \\ \mathbf{L}^{(2,0)} \end{bmatrix},$$

where matrices $\mathbf{L}^{(0,0)}, \mathbf{L}^{(1,1)}, \mathbf{L}^{(1,2)}$, and $\mathbf{L}^{(2,0)}$ are constructed by following equation (13). In the vertical direction, we give the mean value $\boldsymbol{\mu} = (\mu_{1,1}, \mu_{1,2}, \dots, \mu_{n_{\text{row}}, n_{\text{col}}})^T$ and standard deviation $\sqrt{\boldsymbol{\alpha}} = (\sqrt{\alpha_{1,1}}, \sqrt{\alpha_{1,2}}, \dots, \sqrt{\alpha_{n_{\text{row}}, n_{\text{col}}}})^T$ of the field with a Chapman profile [Brekke, 1997].

Chapman profiles are altitude profiles of the ionospheric electron density, which result from the balance between electron production and loss processes in the ionosphere. Specifically, as a production process the

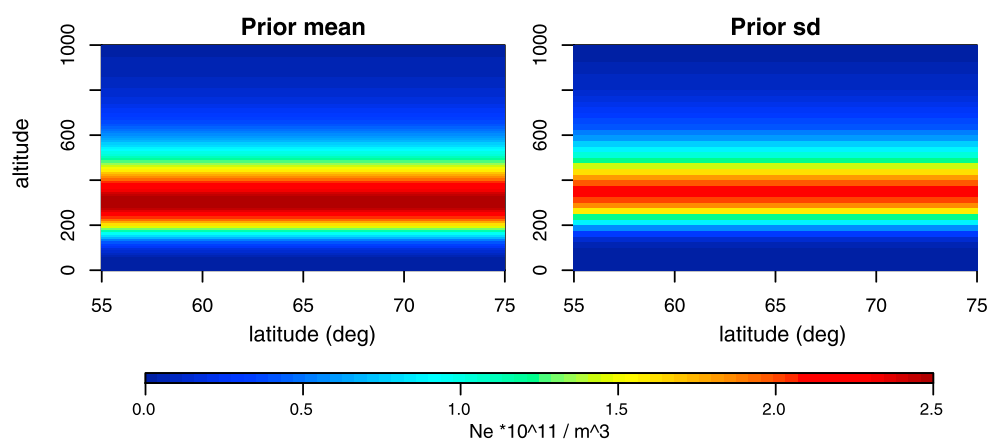


Figure 2. Parameter fields for prior distribution used in Prior 1.

Chapman profiles include electrochemical ionization due to energy input into the neutral atmosphere, and as a loss process recombination due to interaction of electrons with ions and neutral particles. The shape of the Chapman profile is defined with scale height, peak value, and peak altitude. The latter two of these parameters define what the expected maximum of the electron density is and its altitude. Scale height gives

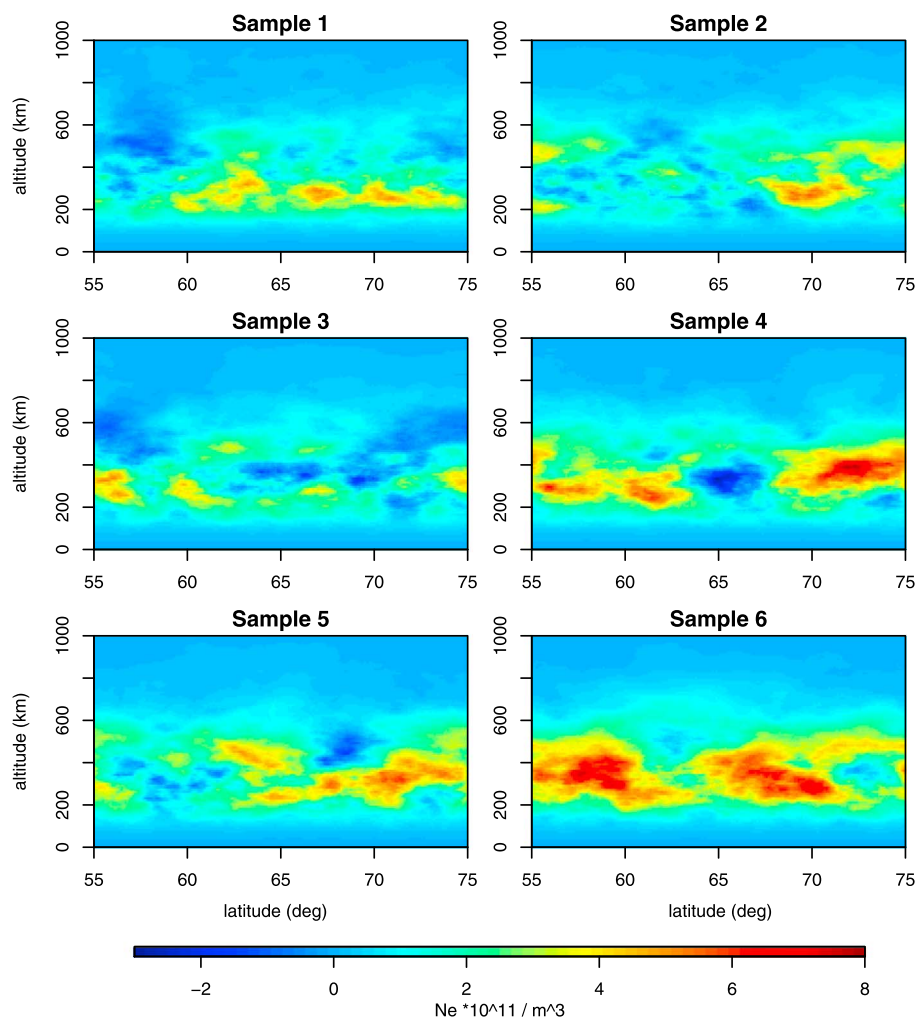


Figure 3. Six samples from the Prior 1 distribution.

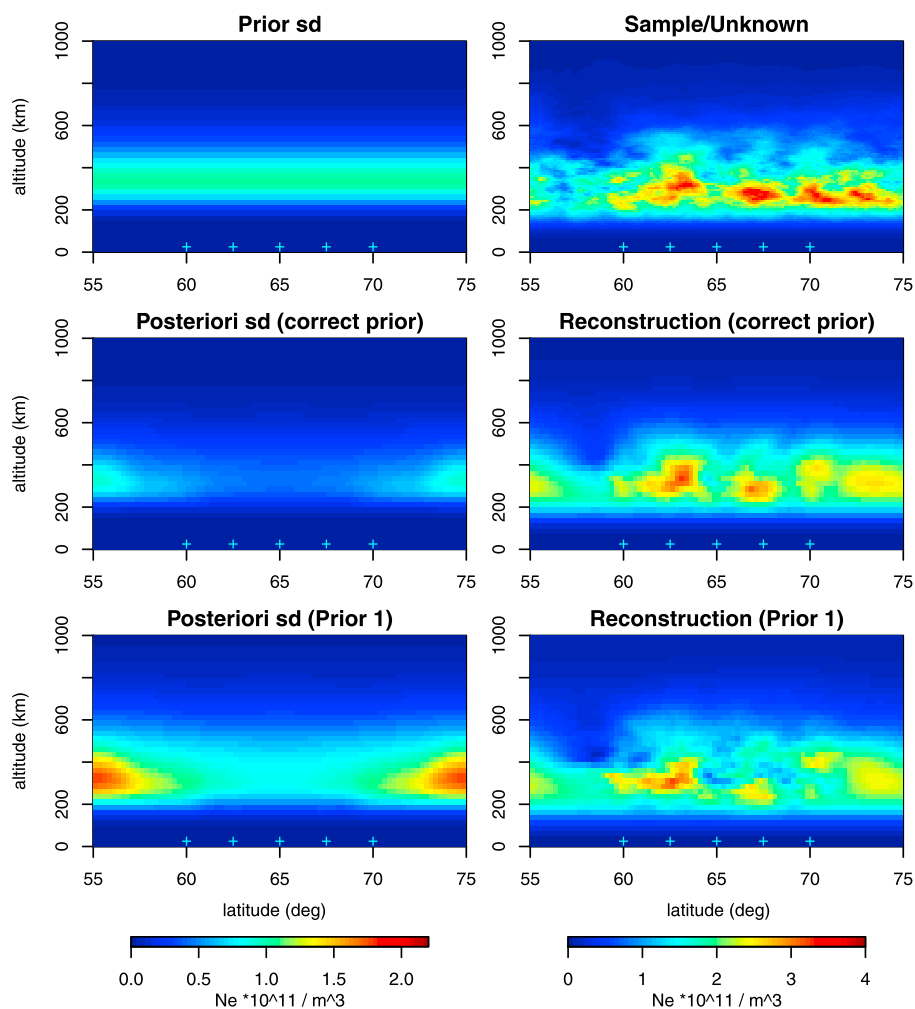


Figure 4. Case 1. (top row) The new prior standard deviation and the corresponding sample. (middle row) The posterior standard deviation and maximum a posteriori estimate with the correct prior. (bottom row) The posterior standard deviation and maximum a posteriori estimate with Prior 1. Receivers marked with cyan plus signs.

the vertical width of the profile. Shorter scale heights, close to 100 km or under, gives a profile where the electron density is expected to be concentrated strictly around the peak altitude. Larger scale heights such as 145 km used in the simulations here spreads the profile of expected electron density wider to low and high altitudes.

The mean value is our initial guess on the unknown electron density before the measurements, and the standard deviation states how much we trust this initial guess. In addition to that, we set a correlation length for both horizontal l_1 and vertical directions l_2 . We define here the correlation length as a distance where the correlation between two points drops to 10% of variance. These parameters can be interpreted as how large details we expect to detect in the reconstruction. Very long correlation lengths correspond to a very smooth ionosphere.

The periodicity conditions of equation (11) are given for computational reasons. As these conditions cannot be justified physically, in the vertical direction we have cut the correlation length close to the vertical dimension of the pixel size at the boundaries. This could also be done for the horizontal direction, but in practice the reconstructions have to be done in domains larger than it is reasonable to use for analysis. Hence, the possible boundary effects are cut out from the resulting domain for ionospheric analysis.

In Figure 2 we present the mean value and standard deviations for an example case, from now on referred to as Prior 1. The peak for mean value is $2.5 \cdot 10^{11} \text{ Ne/m}^3$ at 300 km altitude. For the standard deviation the peak is $2 \cdot 10^{11} \text{ Ne/m}^3$, also at 300 km altitude. The scale height for the standard deviation profile is 120 km

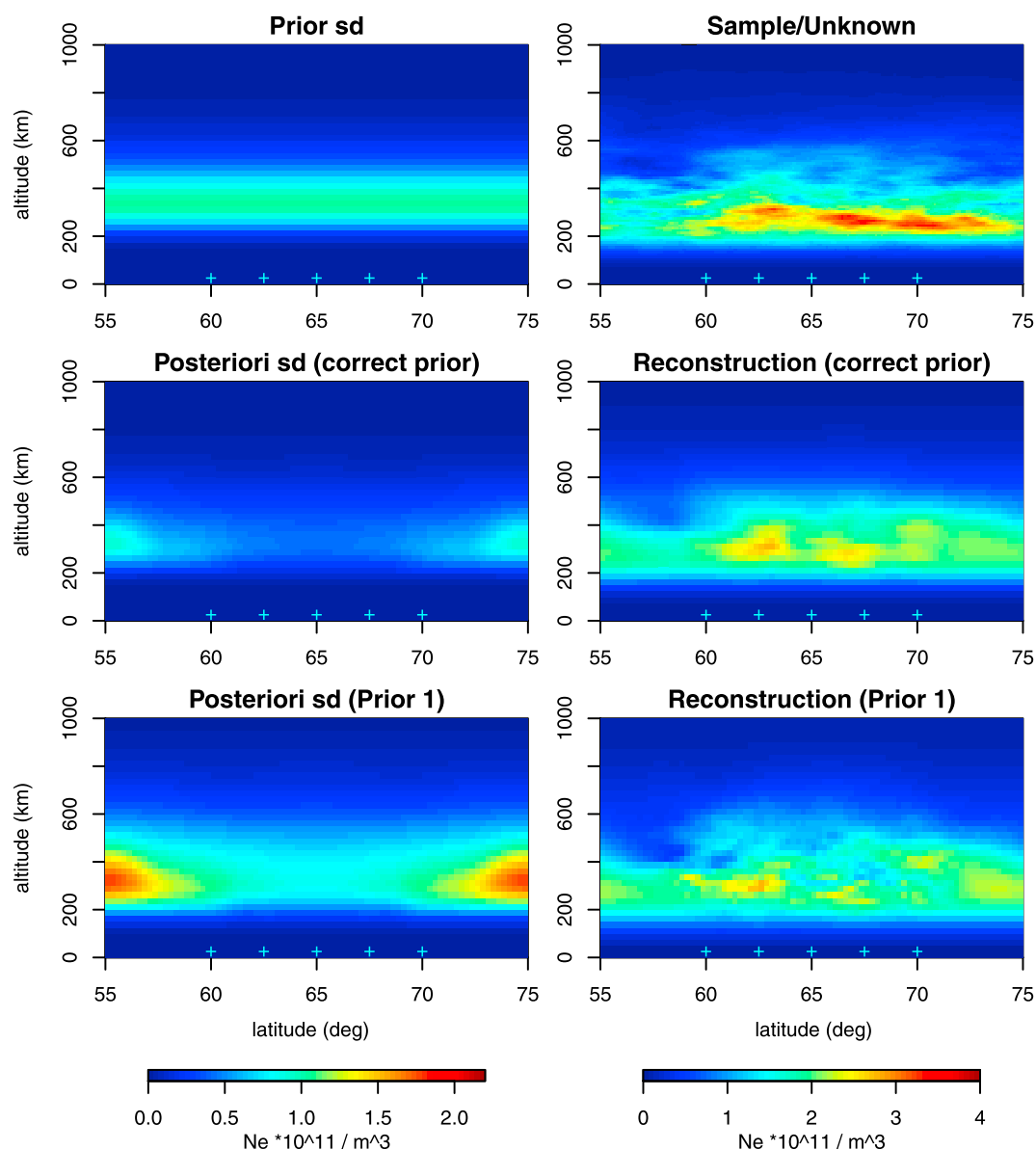


Figure 5. Case 2. (top row) The new prior standard deviation and the corresponding sample. (middle row) The posterior standard deviation and maximum a posteriori estimate with the correct prior. (bottom row) The posterior standard deviation and maximum a posteriori estimate with Prior 1. Receivers marked with cyan plus signs.

and 145 km for the mean. The horizontal correlation length is set to 10° and vertical to 400 km. In Figure 3 we have drawn six sample realizations from Prior 1 distribution. The samples are simulated by multiplying a standard normal random vector with Cholesky decomposition of the prior covariance matrix, then adding the prior mean value [Gentle, 2005].

The samples can be considered as likely realizations of the ionospheric electron density based only on our existing knowledge before the measurements. The samples provide us with a way to understand the prior distribution and to observe that the information stated in previous paragraphs is really incorporated in the samples. At this point we are not using any positivity constraints and, as the standard deviation of the prior is so large, there are some areas with negative values. We do not claim that these samples would be physically realistic realizations of the ionosphere but that we can clearly characterize structural information with the distributions as the samples clearly correspond to the chosen parameter values. The idea here is to

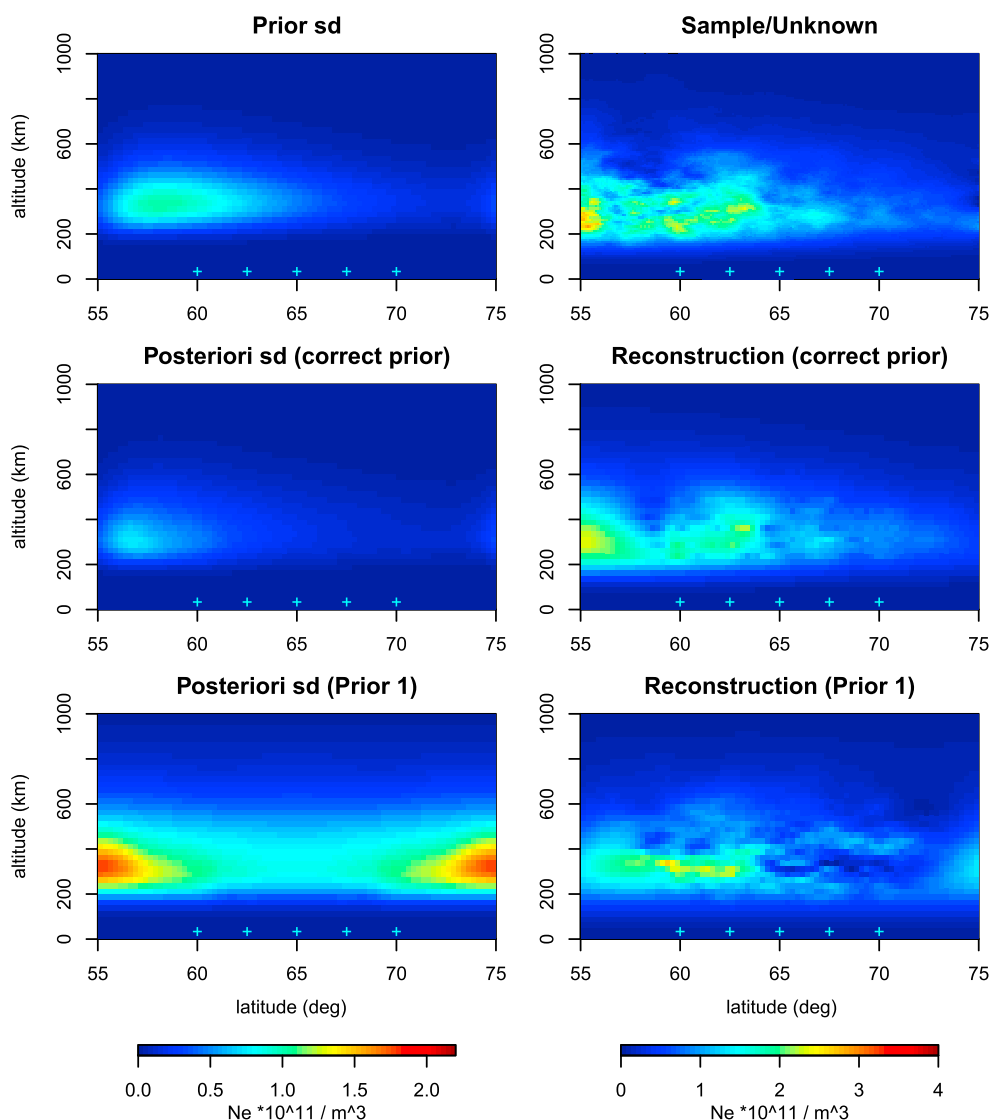


Figure 6. Case 3. (top row) The new prior standard deviation and the corresponding sample. (middle row) The posterior standard deviation and maximum a posteriori estimate with the correct prior. (bottom row) The posterior standard deviation and maximum a posteriori estimate with Prior 1, where the peak value of the mean is decreased to $1.3 \cdot 10^{11} \text{ Ne/m}^3$. Receivers marked with cyan plus signs.

build the Prior 1 with a large variance so that it would contain only the very basic information we have on the ionosphere.

We then study four different cases illustrated in Figures 4–7. In the first three we form a stricter prior distribution than Prior 1, by dropping the peak standard deviation to half, and generate a random sample from that distribution. The sample is then taken as our unknown electron density. The samples in the example cases are all simulated with the same random seed as sample 1 of Figure 3; hence, the effect of different parameter values can be observed by comparing the samples. In the fourth case in Figure 7, we build a simple electron density with vertical Chapman profile and arbitrary horizontal profile without known prior distribution.

The simulation of measurements is done by solving equation (3) as a forward problem. The domain for the electron density is given in latitudes from 55° to 75° . Vertically, it is limited from ground level to 1000 km altitude. We assume one satellite overflight and five receivers spread uniformly between the latitudes of 60° and 70° . The elevation angle limit for the observations that are used in the tomographic inversion is set to 10° . Given this measurement geometry, we first form the matrix \mathbf{A} , take the simulated sample as \mathbf{x} ,

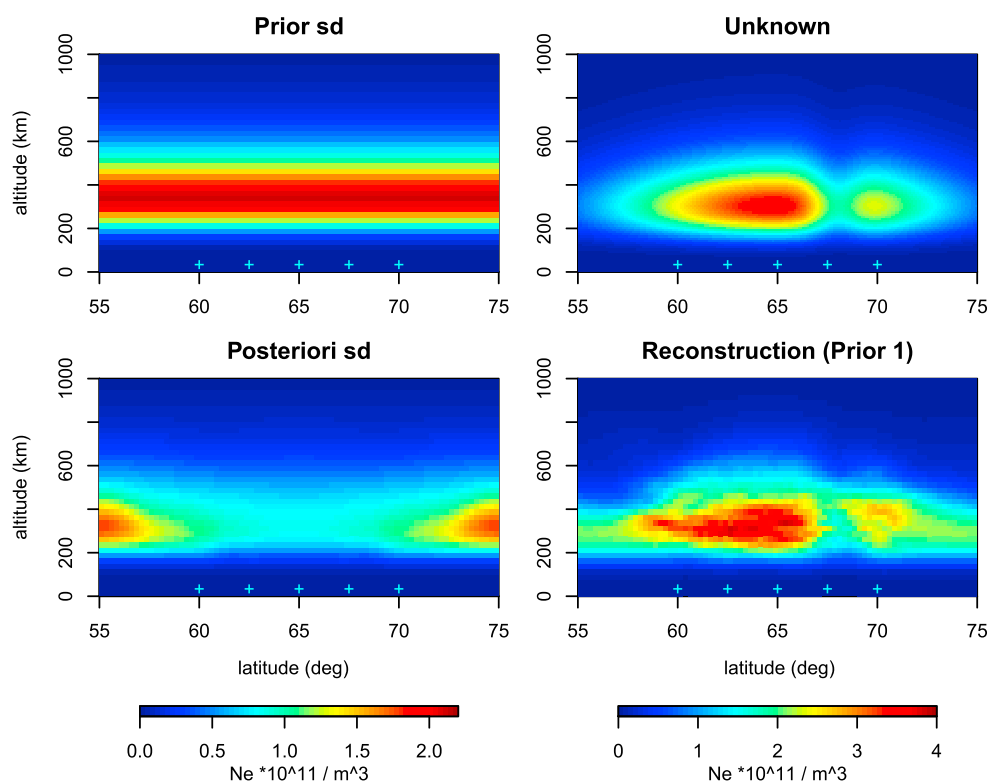


Figure 7. Case 4. (top row) The standard deviation of Prior 1 and a simple synthetical sample. (bottom row) The posterior standard deviation with Prior 1 and the maximum a posteriori estimate. Receivers marked with cyan plus signs.

and compute \mathbf{m} of equation (2). We then simulate the measurement noise \mathbf{e} as white noise, with standard deviation $\sigma = 0.01 \max(\mathbf{m})$ and add this to the measurements to get the relative total electron content measurements \mathbf{M} of equation (3). As explained after equation (1), the real-life ground measurements are relative measurements. Hence, for each satellite-receiver combination we also add an additional random constant γ_i to model the absolute level of total electron content. Here we generate these constants from a Gaussian distribution with a standard deviation of $0.1 \max(\mathbf{m})$. As we are assuming one satellite overflight, this gives us five additional unknown parameters, one for each receiver.

From the simulated measurements, we then solve the “unknown” electron density and phase ambiguity parameters with the estimators given in equations (4) and (5). For the three first cases, the reconstructions are done with both, the correct prior used for sampling and the less restrictive Prior 1 distribution. For the fourth case, the reconstruction is done only with the Prior 1 distribution. In Figures 4–7 the posterior standard deviations are also presented. The posterior standard deviations show the uncertainties of the solution, given the prior distribution and the measurements.

The samples for electron densities are done for 100×200 lattices that correspond to $10 \text{ km} \times 0.1^\circ$ pixel size in this example. For the reconstructions we have considered 40×80 lattices with $25 \text{ km} \times 0.25^\circ$ pixels.

Case 1. In Figure 4 the scale height of the prior standard deviation is set to 100 km and the peak value to $1 \cdot 10^{11} \text{ Ne/m}^3$. The scale height of mean value is also decreased to 125 km. Other parameters are left as in Prior 1. When comparing the sample produced with these parameters to “sample 1” from Figure 3, it can be seen that the main structures remain intact, but the high density is limited to a thinner layer and the variation is lower.

In reconstruction with the correct prior the main structures are recovered, peak densities being slightly lower than in the original electron density. The posterior standard deviation decreases approximately to half from the prior. In reconstruction with Prior 1 the main structures are still somewhat recovered but the overall quality is weaker. As the Prior 1 is less informative, the posterior standard deviation is larger than with the correct prior.

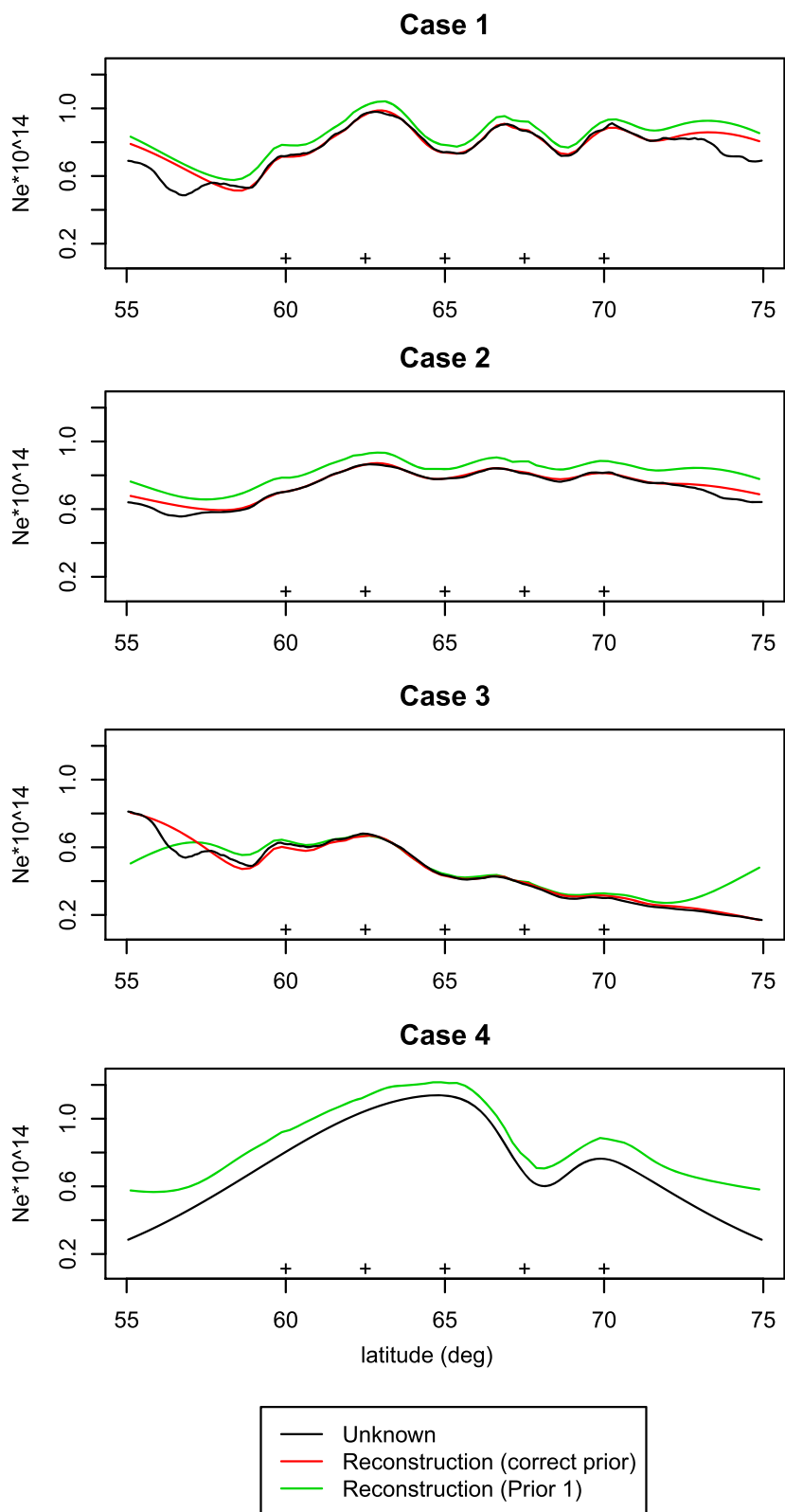


Figure 8. Comparisons of vertically integrated electron content of the original electron density (black) and reconstructions with correct prior (red) and Prior 1 (green). Receivers marked with plus signs.

Case 2. In Figure 5 the parameters are unchanged from Case 1 except for the horizontal correlation length which is set to 20° . When comparing the sample to the sample of the previous case, the effect is clearly visible as the horizontal correlation has increased. With the correct prior, the main features of the reconstructed electron density are reasonably visible, but smoothing is slightly larger than in the previous case. As the horizontal correlation length of Prior 1 is 10° , the Prior 1 can be considered even less informative for Case 2 than it is for Case 1. However, part of the main features is still visible in this reconstruction. As the Prior 1 was used and the measurement geometry was unchanged, the information obtained from the system remains the same as in the previous case, and therefore, the posterior standard deviation is also unchanged.

Case 3. In Figure 6 the parameters are again the same as in Case 1, but the profiles for the prior standard deviation and prior mean have maximum latitude at 57° from where they are horizontally decreasing. The results behave accordingly. As the standard deviation is decreasing northward we are giving stricter information to that direction and the reconstruction with correct prior is therefore also more accurate. This can be seen from the posterior standard deviation which gets very low northward when the correct prior is used. In the first reconstruction trials with Prior 1, in large parts of the area the mean value is significantly larger than in the original electron density; also, some negative densities were produced in reconstruction. Hence, for the reconstruction with Prior 1 in Figure 6 the mean of Prior 1 was scaled down to $1.3 \cdot 10^{11} \text{ Ne/m}^3$ until the reconstruction was strictly positive. The posterior standard deviation remains the same compared to previous cases.

Case 4. In Figure 7, a simple electron density is constructed by multiplying a vertical Chapman profile with an arbitrary two-peak horizontal profile. Hence, in this case there is no exact prior knowledge and we rely only on the Prior 1 distribution. In the reconstruction, the main features of the electron density are recovered rather well; however, some values are present that are higher than those in the original electron density. Similar to previous cases, the posterior standard deviation is unchanged as the information provided by the system is the same.

In Figure 8, we have compared the vertically integrated electron content between the original electron density and different reconstructions. In all of the cases the Prior 1 seems to overestimate the electron content slightly. Overall the plots show reasonable agreement between integrated electron contents. In *Vierinen et al.* [2014] the method is applied successfully also to real measurements.

6. Conclusion and Discussion

It is well known that the ionospheric tomography problem, as presented here, is ill posed and cannot be solved uniquely. In order to get a reasonable tomographic reconstruction, additional information is needed to stabilize/regularize the problem. To give this additional information in an easily interpretable manner, we have used Bayesian statistical inversion, where the stabilization is given as a prior distribution.

As the Bayesian approach results in computational difficulties with the full covariance matrices that need to be inverted, we have used GMRF priors to overcome these computational difficulties. We have demonstrated that with GMRFs we can incorporate physically interpretable information in the prior distribution. Most importantly, the GMRFs are implemented with extremely sparse matrices. The sparse matrices provide computational efficiency making the Bayesian statistical inversion approach feasible with simple matrix computations.

In Bayesian statistical inversion, when working with real data, the prior distribution can be chosen based on other measurements or models. However, our uncertainty on this guess should be then reflected in the variance of the prior distribution. The possibility to choose the prior somewhat subjectively might feel unsound, but we stress that all the different tomographic algorithms need the stabilizing/regularizing information and in this Bayesian approach this information is just made visible and interpretable.

In synthetic test cases, where the true electron density is known, the correspondence between the individual reconstruction and the underlying truth can be enhanced considerably by tuning the parameters through trial and error. Instead of using such an approach, the examples presented in section 5 give an overall understanding of how well the ionospheric tomography problem can be solved in extreme cases, with regard to the scale of details in the electron density and to the uncertainty of the state of the ionosphere before the measurements.

In the Bayesian framework, the posterior covariance also provides us with understanding of the uncertainties of the resulting reconstruction. In cases with the correct prior, credible intervals can be computed from the posterior standard deviations. As the solution is highly dependent on the prior in use, one has to be cautious before jumping to conclusions about credible intervals when using real data. However, when prior modeling is done carefully, even this kind of reasoning can be justified.

7. Future Work

We are currently building a new ionospheric tomography receiver network in Fenno-Scandinavia. The future work will include applications of the method presented in this paper to real dual-frequency beacon data. The framework presented in the paper is also well suited for multidata source inversion, allowing the inclusion of measurements from other ionospheric remote sensing instruments, such as ionosondes, GPS receivers, and incoherent scatter radars. We also plan to extend the method for three-dimensional ionospheric tomography.

Acknowledgments

This work has been funded by Academy of Finland (grants 132694 and 250215) and European Regional Development Fund (Regional Council of Lapland, grant A31221). The data used in this study are simulated in R environment and can be regenerated by following the examples.

References

- Andreeva, E. S. (1990), Radio tomographic reconstruction of ionisation dip in the plasma near the Earth, *J. Exp. Theor. Phys.*, *54*, 142–148.
- Arikan, O., F. Arikan, and C. B. Erol (2007), 3-D Computerized ionospheric tomography with random field priors, in *Mathematical Methods in Engineering*, pp. 325–334, Springer, Netherlands.
- Austen, J. R., S. J. Franke, and C. H. Liu (1988), Ionospheric imaging using computerized tomography, *Radio Sci.*, *23*(3), 299–307.
- Bardsley, J. M. (2013), Gaussian Markov random field Priors for inverse problems, *Inverse Problem Imaging*, *7*(2), 397–416.
- Brekke, A. (1997), *Physics of the Upper Polar Atmosphere*, Wiley, West Sussex.
- Bust, G. S., and C. N. Mitchell (2008), History, current state, and future directions of ionospheric imaging, *Rev. Geophys.*, *46*, RG1003, doi:10.1029/2006RG000212.
- Davies, K. (1965), *Ionospheric Radio Propagation*, U.S. Dep. of Commer., Washington, D. C.
- Fremouw, E. J., J. A. Secan, and B. M. Howe (1992), Application of stochastic inverse theory to ionospheric tomography, *Radio Sci.*, *27*(5), 721–732.
- Gentle, J. E. (2005), *Random Number Generation and Monte Carlo Methods—Statistics and Computing*, Springer, New York.
- Hansen, A. J. (2002), Tomographic estimation of the ionosphere using terrestrial GPS sensors, PhD dissertation, Stanford Univ., Stanford, Calif.
- Kaipio, J., and E. Somersalo (2005), *Statistical and Computational Inverse Problems*, Springer, New York.
- Kaipio, J., V. Kolehmainen, M. Vauhkonen, and E. Somersalo (1999), Inverse problems with structural prior information, *Inverse Problems*, *15*, 713–729.
- Kunitsyn, V. E., and E. D. Tereshchenko (2003), *Ionospheric Tomography*, Springer, Berlin.
- Markkanen, M., M. Lehtinen, T. Nygrén, J. Pirttilä, P. Helenius, E. Vilenius, E. D. Tereshchenko, and B. Z. Khudukon (1995), Bayesian approach to satellite radiotomography with applications in the Scandinavian sector, *Ann. Geophys.*, *13*, 1277–1287.
- Mitchell, C. N., and P. S. J. Spencer (2003), A three-dimensional time-dependent algorithm for ionospheric imaging using GPS, *Ann. Geophys.*, *46*, 687–696.
- Na, H., and H. Lee (1990), Resolution analysis of tomographic reconstruction of electron density profiles in the ionosphere, *Int. J. Imaging Syst. Technol.*, *2*, 209–218.
- Na, H., and E. Sutton (1994), Resolution Analysis of ionospheric tomography systems, *Int. J. Imaging Syst. Technol.*, *5*, 169–173.
- Natterer, F. (1986), *The Mathematics of Computerized Tomography*, Wiley, Cichester, U. K., and New York.
- Pryse, S. E., L. Kersley, D. L. Rice, C. D. Russell, and I. K. Walker (1993), Tomographic imaging of the ionospheric midlatitude trough, *Ann. Geophys.*, *11*, 144–149.
- Raymund, T. D. (1995), Comparisons of several ionospheric tomography algorithms, *Ann. Geophys.*, *13*, 1254–1262.
- Raymund, T. D., J. R. Austen, S. J. Franke, C. H. Liu, J. A. Klobuchar, and J. Stalker (1990), Application of computerised tomography to the investigation of the ionospheric structures, *Radio Sci.*, *25*, 771–789.
- Roininen, L., M. Lehtinen, S. Lasanen, M. Orispää, and M. Markkanen (2011), Correlation priors, *Inverse Problem Imaging*, *5*(1), 167–184.
- Roininen, L., P. Piiroinen, and M. Lehtinen (2013), Constructing continuous stationary covariances as limits of the second-order stochastic difference equations, *Inverse Problem Imaging*, *7*(2), 611–647.
- Roininen, L., J. M. J. Huttunen, and S. Lasanen (2014), Whittle-Matérn priors for Bayesian statistical inversion with applications in electrical impedance tomography, *Inverse Problems Imaging*, *8*(2), 561–586.
- Rue, H., and L. Held (2005), *Gaussian Markov Random Fields—Theory and Applications*, Chapman and Hall/CRC, Boca Raton, Fla.
- Saksman, E., T. Nygrén, and M. Markkanen (1997), Ionospheric structures invisible in satellite tomography, *Radio Sci.*, *32*(2), 606–616.
- Seemala, G. K., M. Yamamoto, A. Saito, and C.-H. Chen (2014), Three-dimensional GPS ionospheric tomography over Japan using constrained least squares, *J. Geophys. Res. Space Physics*, *119*, 3044–3052, doi:10.1002/2013JA019582.
- Vierinen, J., J. Norberg, M. S. Lehtinen, O. Amm, L. Roininen, A. Väänänen, P. J. Erickson, and D. McKay-Bukowski (2014), Beacon satellite receiver for ionospheric tomography, *Radio Sci.*, *49*, 1141–1152, doi:10.1002/2014RS005434.
- Yeh, K. C., and T. D. Raymund (1991), Limitations of ionospheric imaging by tomography, *Radio Sci.*, *26*(6), 1361–1380.

Publication II

J. Vierinen, J. Norberg, M. S. Lehtinen, O. Amm, L. Roininen, A. Väänänen, P. J. Erickson and D. McKay-Bukowski. Beacon satellite receiver for ionospheric tomography *Radio Sci.*, 49(12):1141-1152, 2014. ISSN 1944799X. doi: 10.1002/2014RS005434.

RESEARCH ARTICLE

10.1002/2014RS005434

Special Section:

Beacon Satellite
Symposium 2013

This article is a companion
to *Norberg et al.* [2014]
doi:10.1002/2014RS005431.

Key Points:

- New low-cost beacon satellite receiver
- Measurement equations and error terms for ionospheric tomography

Correspondence to:

J. Vierinen,
x@mit.edu

Citation:

Vierinen, J., J. Norberg, M. S. Lehtinen, O. Amm, L. Roininen, A. Väänänen, P. J. Erickson, and D. McKay-Bukowski (2014), Beacon satellite receiver for ionospheric tomography, *Radio Sci.*, 49, 1141–1152, doi:10.1002/2014RS005434.

Received 20 MAR 2014

Accepted 28 SEP 2014

Accepted article online 2 OCT 2014

Published online 1 DEC 2014

Beacon satellite receiver for ionospheric tomography

J. Vierinen¹, J. Norberg², M. S. Lehtinen³, O. Amm², L. Roininen³, A. Väänänen³,
P. J. Erickson¹, and D. McKay-Bukowski³
¹Haystack Observatory, Massachusetts Institute of Technology, Westford, Massachusetts, USA, ²Finnish Meteorological Institute, Helsinki, Finland, ³Sodankylä Geophysical Observatory, University of Oulu, Sodankylä, Finland

Abstract We introduce a new coherent dual-channel beacon satellite receiver intended for ionospheric tomography. The measurement equation includes neutral atmosphere and ionosphere propagation effects, relative errors in satellite and receiver clocks, and residual Doppler shifts caused by errors in the satellite ephemeris. We also investigate the distribution of errors for phase curve measurements and the use of phase curve measurements for limited angle tomography using the framework of statistical linear inverse problems. We describe the design of our beacon satellite receiver software and present one possible hardware configuration. Finally, we present results obtained using a network of four newly developed receivers and compare the results with those of an existing ionospheric tomography network at Sodankylä Geophysical Observatory.

1. Introduction

Ground-based ionospheric tomography is a method for estimating ionospheric electron density using measurements of ionospheric delay as electromagnetic waves transmitted from satellites are received at multiple ground stations [Austen *et al.*, 1988; Raymund *et al.*, 1990; Kunitsyn *et al.*, 1994; Markkanen *et al.*, 1995; Nygrén *et al.*, 1996; Mitchell *et al.*, 1997; Bernhardt *et al.*, 1998; Kunitsyn and Tereshchenko, 2003; Garner *et al.*, 2009]. There are two main families of satellites that can be used for this purpose: low Earth orbit beacons and global navigation satellites. We will focus in this work on low Earth orbit (LEO) satellites that transmit phase coherent 150/400 MHz signals. These are of interest especially for local high-resolution tomography, as lower frequencies are more affected by the ionosphere than global navigation satellite frequencies. LEO satellites also have fast ground station transit times, providing significantly greater independent measurements of line integrals within a given time interval. These measurement qualities provide enhanced time resolution of derived ionospheric density structures over that possible with, e.g., GPS satellite networks [Rideout and Coster, 2006]. Bernhardt and Siefring [2006] provide an overview of coherent beacon satellite theory and operation.

The Sodankylä Geophysical Observatory (SGO) in Finland operates an existing tomography receiver network with receivers that were custom built by Invers Ltd. These receivers rely on a Russian Tsykada navigation satellite system [Kunitsyn *et al.*, 1994] limited to three currently operational transmitting platforms. In order to upgrade these receivers for modern observations, we require a new receiver that can function autonomously, observe multiple satellites simultaneously, and support all currently operational 150/400 MHz beacon satellites.

One practical way of implementing an ionospheric tomography receiver is using a software-defined radio [Mitola III, 2000; Reed, 2002], in which most of the signal processing is performed with software on a general purpose computer. The main advantage of this approach is that the receiver software can be continuously improved, even after deployment of the instrument. Scientific numerical algorithms can also be used, instead of fixed purpose signal processing hardware. An implementation of an open-source software-defined beacon satellite receiver is described by Yamamoto [2008]. In this paper, we will introduce a new beacon receiver that builds on the work of Yamamoto and is specifically intended for autonomous operation of receivers in a large network configuration.

In the following sections, we derive equations of radio propagation through the atmosphere and describe the fundamentals of a beacon satellite measurement. We then outline the basic principle of tomographic reconstruction of ionospheric electron density using relative propagation time differences measured using a ground-based receiver network. After a description of the technical details of our hardware and software,

we conclude by presenting results from several newly deployed ionospheric tomography receivers. These results will be compared with data from the existing ionospheric tomography receiver chain operated by the SGO.

2. Atmospheric Propagation

The phase velocity of radio waves in a medium is defined as

$$v_p = cn^{-1}, \quad (1)$$

where n is the refractive index of the medium. In the ionosphere, the refractive index of the medium is given by the Appleton-Lassen magnetoionic propagation equation [Appleton and Chapman, 1932; Lassen, 1927; Hartree, 1929]. With the frequencies used in VHF and UHF beacon satellites, it is possible to ignore electron-neutral collisions and the magnetic field. In this case, the refractive index provided by the Appleton-Lassen equation can be approximated using

$$n = \left(1 - \omega_p^2/\omega^2\right)^{\frac{1}{2}}, \quad (2)$$

where $\omega_p = \sqrt{\frac{N_e e^2}{\epsilon_0 m_e}}$ is the electron plasma frequency (rad/s), and ω is the frequency of the radio wave (rad/s). The term N_e is electron density (m^{-3}), e is charge of an electron, ϵ_0 is the permittivity of free space, and m_e is the mass of an electron.

Below 50 km, the refractive index of plasma approaches 1 as the number density of electrons approaches 0. However, the refractive index of the neutral atmosphere starts becoming significant as the density of air grows. The lower atmospheric refractive index n_{tr} of radio waves below 20 GHz is assumed to be nondispersive and can be related to the pressure, temperature, and humidity of air using the following formula [Rüeger, 2002]

$$n_{\text{tr}} = 1 + 10^{-6} \left(77.68 \frac{p_d}{T} + 71.30 \frac{p_w}{T} + 3.75 \cdot 10^{-5} \frac{p_w}{T^2} \right), \quad (3)$$

where p_d is the partial pressure of dry air (hPa), T is the temperature (K), and p_w is the partial pressure of water vapor (hPa). We will not discuss inversion of tropospheric water vapor parameters [Bernhardt et al., 2000] in this work, but we will investigate the effect of the tropospheric propagation for the relative phase curve measurement. For this, it is sufficient that we represent the refractive index in the simple form

$$n_{\text{tr}} = 1 + \Delta n_{\text{tr}}, \quad (4)$$

where the small perturbation in the refractive index Δn_{tr} is equivalent to the second term on the right-hand side of equation (3).

We assume that an electromagnetic wave propagates from the satellite to a ground station along the axis z , with distance L between the satellite and ground station. In this case, with the satellite at $z = 0$ and phase velocity $v_p(z)$, the electric field amplitude can be expressed as

$$E(L, t) = E_0 \cos \left(\omega \left(t - \int_0^L dz/v_p(z) \right) \right). \quad (5)$$

From this we can determine the phase $E(L, t) = \cos(\phi'(t))$ of a signal at a ground station at a time-dependent distance $L(t)$ from the satellite as

$$\phi'(t) = \omega t - \frac{\omega}{c} \int_0^{L(t)} n(z) dz. \quad (6)$$

However, a radio receiver is typically tuned to the center frequency ω , which cancels out the term ωt and leaves

$$\phi(t) = \phi'(t) - \omega t = -\frac{\omega}{c} \int_0^{L(t)} n(z) dz. \quad (7)$$

In order to convert this into an integral of $N_e(z)$, we use the approximative form of the refractive index in equation (2) to determine the electron density as a function of the refractive index

$$N_e(z) = -\frac{\epsilon_0 m_e \omega^2}{e^2} (n(z)^2 - 1). \quad (8)$$

By approximating this as a first-order Taylor polynomial expanded around $n(z) = 1$ we obtain

$$N_e(z) \approx -\frac{2\epsilon_0 m_e \omega^2}{e^2} (n(z) - 1), \quad (9)$$

which is a good approximation when $\omega \gg \omega_p$ [Davies, 1965]. Using this, we get a formula for $n(z)$:

$$n(z) \approx 1 - \frac{e^2}{2\epsilon_0 m_e \omega^2} N_e(z). \quad (10)$$

If we now combine equations (10) and (4) and insert them into equation (7), we get

$$\phi(t) = -\frac{\omega}{c} L(t) + \frac{a}{\omega} \int_{L_0(t)}^{L(t)} N_e(z) dz - \frac{\omega}{c} \int_0^{L_0(t)} \Delta n_{tr}(z) dz, \quad (11)$$

where $a = \frac{e^2}{2\epsilon_0 m_e c}$. The above equation relates the measured phase to the line integral of electron density and the line integral of the perturbations in the refractive index caused by lower atmospheric propagation. Here the term $L(t)$ contributes to the Doppler shift of the signal in vacuum. This is typically known to some accuracy, and it can also be removed by adjusting the center frequency of the receiver correspondingly. However, we will leave this term in order to show how errors in $L(t)$ will affect the signal. The range $L_0(t)$ corresponds to the range where the refractive index transitions from tropospheric refractive index to ionospheric refractive index. We can make this assumption, as there is no appreciable overlap between those altitudes where ionospheric refractive index and the neutral atmosphere refractive index is significantly different from 1.

When receiving a signal for the first time after the satellite rises above the horizon, practical measurements of $\phi(t)$ have the difficulty of unknown initial phase γ due to the inherent 2π phase ambiguity. Additional errors are also introduced by unknown relative clock drifts of the satellite and the ground station, as well as uncertainties in the distance between the satellite and the ground station, which are all here denoted by a random variable $r(t)$. In practice, our measured phase curve will be of the form

$$\phi(t) = \gamma - \frac{\omega}{c} (L(t) + r(t)) + \frac{a}{\omega} \int_{L_0(t)}^{L(t)} N_e(z) dz - \frac{\omega}{c} \int_0^{L_0(t)} \Delta n_{tr}(z) dz. \quad (12)$$

Using $L(t)$ for the upper bound of the integral is a sufficiently good approximation as the errors in range are insignificant $r(t) \ll L(t)$. We can also write this in more concise form as follows:

$$\phi(t) = \gamma + \omega \zeta(t) + \frac{a}{\omega} \int_{L_0(t)}^{L(t)} N_e(z) dz, \quad (13)$$

where

$$\zeta(t) = -\frac{1}{c} \left(L(t) + r(t) + \int_0^{L_0(t)} \Delta n_{tr}(z) dz \right). \quad (14)$$

When estimating total electron content using dual-frequency beacon satellites, two simultaneous measurements of $\phi(t)$ are made at two different frequencies, i.e.,

$$\begin{aligned} \phi_1(t) &= \gamma_1 + a\omega_1^{-1} \int_{L_0(t)}^{L(t)} N_e(z) dz + \zeta(t)\omega_1 \\ \phi_2(t) &= \gamma_2 + a\omega_2^{-1} \int_{L_0(t)}^{L(t)} N_e(z) dz + \zeta(t)\omega_2. \end{aligned} \quad (15)$$

When we multiply $\phi_1(t)$ with $\frac{\omega_2}{\omega_1}$ and subtract $\phi_2(t)$ from it, the term $\zeta(t)$ is canceled out up to an unknown constant $\gamma' = \frac{\omega_2}{\omega_1} \gamma_1 - \gamma_2$ and we obtain

$$\frac{\omega_2}{\omega_1} \phi_1(t) - \phi_2(t) = \gamma' + a \left(\frac{\omega_2}{\omega_1^2} - \frac{1}{\omega_2} \right) \int_{L_0(t)}^{L(t)} N_e(z) dz,$$

where the contribution of $\zeta(t)$ is completely eliminated. This also has the benefit of eliminating lower atmospheric propagation effects, relative clock errors, and trajectory errors. We can express this in another way as

$$\Delta\phi_{12}(t) = \gamma'' + \int_{L_0(t)}^{L(t)} N_e(z) dz = \frac{1}{a} \left(\frac{\omega_2}{\omega_1^2} - \frac{1}{\omega_2} \right)^{-1} \left(\frac{\omega_2}{\omega_1} \phi_1(t) - \phi_2(t) \right), \quad (16)$$

where $\Delta\phi_{12}(t)$ is the so-called relative total electron content, comprising the line integral of electron density along the path between the satellite and the receiver, up to an unknown constant γ'' . This is the basic measurement that goes into the ionospheric tomography inversion procedure, which determines γ'' and $N_e(z)$ from simultaneous observations at multiple stations.

In practice, relative total electron content is typically measured at 1–50 Hz sampling rate during the pass of a satellite. We assume that $\Delta\phi_{12}(t)$ changes slowly as a function of time, which allows us to unwrap 2π phase ambiguities by selecting the smallest of the possible phase differences. However, if one of the satellite signals is lost for a sufficiently long time, removing this phase ambiguity becomes more difficult. In this case, one sensible option is to split the phase curves into two phase curves with independent unknown constants.

Using a similar well-known subtraction procedure [Danchik, 1998], we can also estimate $\zeta(t)$, which contains information about the troposphere, the clock errors, and the satellite trajectory errors. In this case, we multiply $\phi_1(t)$ with ω_1 and $\phi_2(t)$ with ω_2 and then we subtract the first from the second to obtain

$$\zeta(t) = \gamma + \frac{\omega_1 \phi_1(t) - \omega_2 \phi_2(t)}{\omega_1^2 - \omega_2^2}, \quad (17)$$

where γ is again an unknown constant. This measurement has the considerable benefit of eliminating ionospheric effects. If the trajectory is known to a good enough accuracy ($r(t)$ is small), then it is possible to make inferences of the tropospheric refractive index. The measurement $\zeta(t)$ can also be used to perform accurate orbital elements using multiple receiver stations [Izsak, 1960].

3. Measurement Errors

Correct modeling of errors in phase curve measurements is important for ionospheric tomography. To correctly express phase errors, we first inspect a discretized baseband measurement of a signal $s_t = ae^{i\phi_t} \in \mathbb{C}$, contaminated with noise that affects the amplitude and phase of the signal. This is in essence the measurement $m_t \in \mathbb{C}$ of a beacon satellite transmission

$$m_t = s_t + \xi_t = (a + \xi_t^a) e^{i(\theta_t + \xi_t^\theta)}. \quad (18)$$

Here ξ_t^a and ξ_t^θ are real-valued random processes that describe the amplitude and phase noise of the measurement. When assuming small enough additive Gaussian noise $\xi_t \sim N_{\mathbb{C}}(0, \sigma^2)$, ξ_t^a and ξ_t^θ can also be approximated as Gaussian noise.

When estimating the differential phase curve $\phi_t \in \mathbb{R}$, we examine the phase increment associated with each measurement m_t and cumulatively sum this together

$$\phi_t = \sum_{n=0}^t \text{Arg}(m_n/m_{n-1}), \quad (19)$$

or

$$\phi_t = \sum_{n=0}^t \theta_n - \theta_{n-1} + \xi_n^\theta - \xi_{n-1}^\theta. \quad (20)$$

Assuming that $\theta_n - \theta_{n-1} + \xi_n^\theta - \xi_{n-1}^\theta \ll 2\pi$, the cumulative summation does not lead to a cumulative error. To see this, we first investigate what happens at the initial step, $t = 1$

$$\phi_1 = \theta_1 - \theta_0 + \xi_1^\theta - \xi_0^\theta. \quad (21)$$

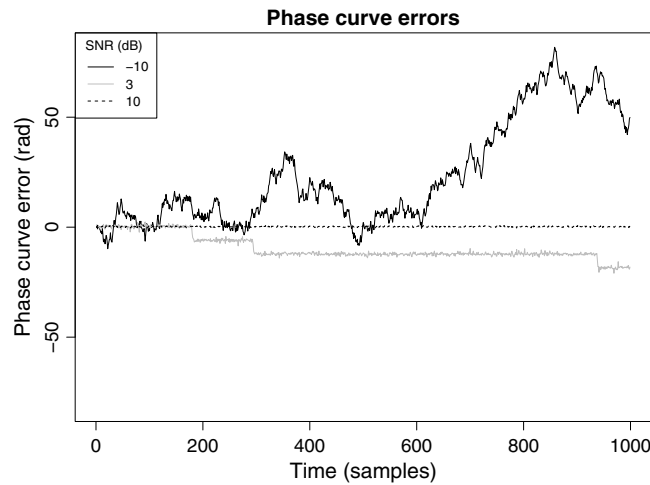


Figure 1. Phase errors with unwanted $\pm 2\pi$ jumps. In the case of low SNR (-10 dB), the jumps are nearly random jumps uniformly between $[-\pi, \pi]$. These cannot be easily detected. In the case of medium SNR (3 dB), there are a few discrete $\pm 2\pi$ jumps. In the case of high SNR (10 dB), random jumps of $\pm 2\pi$ are extremely rare.

At $t = 2$, we get

$$\phi_2 = \sum_{n=0}^2 (\theta_n - \theta_{n-1}) + \xi_2^\theta - \xi_0^\theta, \quad (22)$$

as ξ_1^θ cancels out. Through induction, we can then see that on step $t + 1$, we only have two noise terms, composed of initial error ξ_0^θ and a new term ξ_{t+1}^θ . All of the intermediate noise terms have canceled out in the successive cumulative sums.

$$\phi_{t+1} = \sum_{n=0}^{t+1} (\theta_n - \theta_{n-1}) + \xi_{t+1}^\theta - \xi_0^\theta. \quad (23)$$

Thus, for small enough errors, we can assume that the phase curve error is an identically distributed and independent Gaussian random variable. However, if the random errors are too large, there will be a 2π ambiguity that cannot be removed.

Simulated phase curve errors for several different signal to noise ratios are depicted in Figure 1. These were done by using equation (18) with constant signal $s_t = 1$ and added complex Gaussian random noise with variance $\sigma_t \sim N_{\mathbb{C}}(0, \text{SNR}^{-1})$. In this simulation, it is apparent that with a SNR of 10 dB, the errors are zero mean and they are not cumulative. With a SNR of 3 dB, the additive noise occasionally causes a 2π jump in either direction. In the case of a SNR of -10 dB, the signal resembles a random walk with uniformly distributed $[-\pi, \pi]$ steps in phase.

In order to properly handle cases where there is a significant drift of phase due to poor SNR, one possibility is to examine the SNR and to simply determine that the phase curve is not continuous anymore at the location where large erroneous drifts in phase are expected. In this case, when performing tomographic inversion, there will be multiple independent phase curves during each satellite overpass. This is discussed in section 4.

In practice, there are also other signal degrading factors that cause errors when estimating the phase curve. Signals from other beacon satellites, and other man-made radio noise can cause large erroneous drifts in the phase curve. In practice, these cannot be detected from SNR alone. In this case, one option is to search for large sudden jumps in the phase curve.

4. Tomographic Reconstruction

The purpose of a tomographic reconstruction of the ionosphere is to estimate the two- or three-dimensional electron density function $N_e(\mathbf{p})$ from phase curves measured during a satellite pass on multiple receiver stations. Here \mathbf{p} represents a spatial coordinate. A conceptual diagram of a beacon satellite measurement is shown in Figure 2.

This problem, which is sometimes called the limited angle tomography problem [Kaipio and Somersalo, 2005], is ill posed in the sense that the measurements do not contain enough information so that a unique solution for the unknown can be found. In order to get an estimate for the unknown, some additional information is therefore needed to stabilize the problem. In the case of ionospheric tomography, multiple possible approaches exist. One common method is to assume that the solution is spatially smooth and temporally unchanged during the time when the satellite passes over the receivers.

Numerically, the problem is solved by parametrizing $N_e(\mathbf{p})$ in some way, i.e., using the form $N_e(\mathbf{p}; \boldsymbol{\theta})$ where $\boldsymbol{\theta} \in \mathbb{R}^{N_p}$ denotes the finite set of model parameters that define the electron density within the volume, and N_p is the number of parameters.

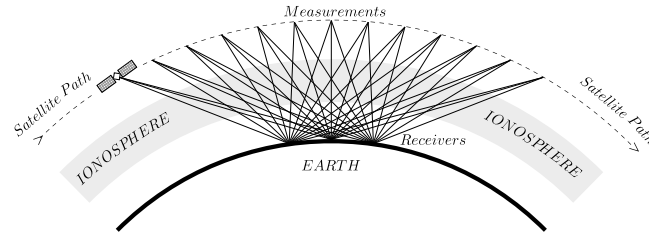


Figure 2. A conceptual diagram of an ionospheric tomography measurement. A satellite passes above ground-based receivers, which measure the propagation delay for the signal transmitted by the satellite.

station producing an independent phase curve is indexed here with n . Thus, our measurement equation for each observation becomes

$$m_{t,n} = \frac{1}{N} \sum_{i=1}^N N_e(\mathbf{l}_i^{t,n}; \boldsymbol{\theta}) + \gamma_n + \xi_{t,n}, \quad (24)$$

where $\mathbf{l}_i^{t,n}$ denotes the location of the i th point along the Riemann sum with N partitions and t denotes a time point on the measured phase curve.

If the relation between the model parameters and the unknown is linear, then the Riemann sum can be written in the following form, where $f(\mathbf{p}) \in \mathbb{R}^{N_p}$ is a vector-valued weighting function that evaluates $N_e(\mathbf{p}; \boldsymbol{\theta})$ at the given point \mathbf{p} according to the interpolation and integration rules. This allows the measurement equation to be written as

$$m_{t,n} = \frac{1}{N} \sum_{i=1}^N f(\mathbf{l}_i^{t,n}) \cdot \boldsymbol{\theta} + \gamma_n + \xi_{t,n}, \quad (25)$$

where \cdot denotes a dot product between $f(\mathbf{p})$ and $\boldsymbol{\theta}$.

An equation of linear form can be expressed as a linear matrix equation

$$\mathbf{m} = \mathbf{A}\mathbf{x} + \boldsymbol{\xi}, \quad (26)$$

where \mathbf{m} is a vector containing all the phase curve measurements $m_{n,t}$. The vector \mathbf{x} contains the unknown parameters $\boldsymbol{\theta}$ as well as the unknown constant offsets γ_n . The theory matrix \mathbf{A} contains elements that make equations (25) and (26) mathematically equivalent. The vector $\boldsymbol{\xi}$ contains the measurement errors, which are assumed to be normally distributed.

Such a linear problem can be solved within the framework of statistical inverse problems [Kaipio and Somersalo, 2005; Calvetti and Somersalo, 2007]. The maximum a posteriori estimator for the unknown parameters \mathbf{x} can be written as

$$\hat{\mathbf{x}}_{\text{MAP}} = \boldsymbol{\Sigma}_{\text{post}} \mathbf{A}^T \boldsymbol{\Sigma}^{-1} \mathbf{m}, \quad (27)$$

where

$$\boldsymbol{\Sigma}_{\text{post}} = (\boldsymbol{\Sigma}_{\text{pr}}^{-1} + \mathbf{A}^T \boldsymbol{\Sigma}^{-1} \mathbf{A})^{-1}. \quad (28)$$

In this case, the prior covariance assumption $\boldsymbol{\Sigma}_{\text{pr}}$ is used to stabilize the problem by enforcing smoothness to the solution. This approach is also used by the ionospheric tomography chain operated by the SGO [Markkanen et al., 1995] and has also been studied, e.g., by Nygrén et al. [1997].

5. Architecture

Our beacon satellite receiver design can be divided into the following parts: (1) Analog radio front end, (2) coherent dual-channel digital receiver, (3) general purpose personal computer, and software for extracting phase curves from dual-channel digital baseband signals.

Table 1. Locations of New Software-Defined Ionospheric Tomography Receivers

Name	Latitude (N)	Longitude (E)
LYR	78.15	16.04
KEV	69.76	27.01
TRO	69.58	19.22
SOD	67.37	26.63
MEK	62.77	30.97
TAR	58.26	26.46

The main task of the analog front end is to provide amplified signals from the antenna to the digital receiver. The digital receiver then takes analog signals on two channels (150 and 400 MHz) and produces coherently downconverted digitized baseband signals with sufficient bandwidth to include all beacon satellites of interest. This signal is then transferred to the computer.

This approach has several advantages compared to integrated hardware-based solutions. The software is easily customizable and allows more advanced signal processing methods to be used. The overall cost for developing a software-based system also turns out to be cheaper than developing an integrated hardware-based receiver, as pointed out by Yamamoto [2008].

We have implemented a new receiver based on these underlying principles. All of these programs associated with the receiver can be run in real time on a standard Linux PC with modest processing capacity. This collection of programs has been released under the General Public License as a project named Jitter (GNU Ionospheric Tomography Receiver). The software can be downloaded from <http://www.sgo.fi/~j/jitter/web>. This software package will autonomously produce phase curves from satellite passes. (Subsequent tomographic inversion analysis as described in this paper is scheduled for a future software release.)

It should be noted that Professor Mamoru Yamamoto has also described a relative total electron content estimation hardware and software system [Yamamoto, 2008] with many similarities to our system. The main difference is that Jitter is designed for autonomous operation in a large network of receivers and is also capable of simultaneously receiving multiple satellites.

6. Hardware

We have recently deployed six 150/400 MHz beacon satellite receivers as part of Finnish Meteorological Institute's TomoScand project to develop a new 3-D ionospheric tomography capability. Four of these receivers are used in the tomography example in this paper. Two more receivers have since been installed (KEV and LYR) but are not used in the results presented in this paper. The locations of all receivers are listed in Table 1. We plan to expand the receiver network in the near future.

The hardware for these receivers consists of a dual-band quadrifilar helix antenna by Nagara Denshi Ltd., a dual-band preamplifier for the antenna, and an RF front-end stage that is located near the computers. The 150 and 400 MHz channels are received with a single USRP N2x0 series digital receiver equipped with a dual-channel TVRX2 daughterboard (50–860 MHz). A block diagram of the receiver hardware is described in Figure 3.

As the data recording portion of the software is written using the Gnuradio open software framework, it is also possible to use other digital receivers with at least two coherent receiver channels, such as the USRP1, USRP E100, or USRP2. The software also supports alternate daughterboard configurations, e.g., WBX or BasicRX daughterboards for receiving the signals. While we discuss dual-band 150/400 MHz satellite observations in this paper, the software itself is general enough that it can be configured to record other frequency pairs as well such as 150/1066 MHz, provided a suitable antenna, RF front end, and digital receiver is used.

7. Software

The software is divided into different subsystems run as separate processes. One process records narrow-band signals to disk on a scheduled basis whenever a satellite is passing a station. Scheduling information including predicted pass times is provided by another process using standard satellite ephemeride routines. A third process is used to analyze the narrowband recordings, with the help of ephemeris-predicted Doppler shift information to further narrow the bandwidth of the individual channels and to calculate a relative phase difference curve. All of the software written for this package was designed with autonomous long-term operation in mind.

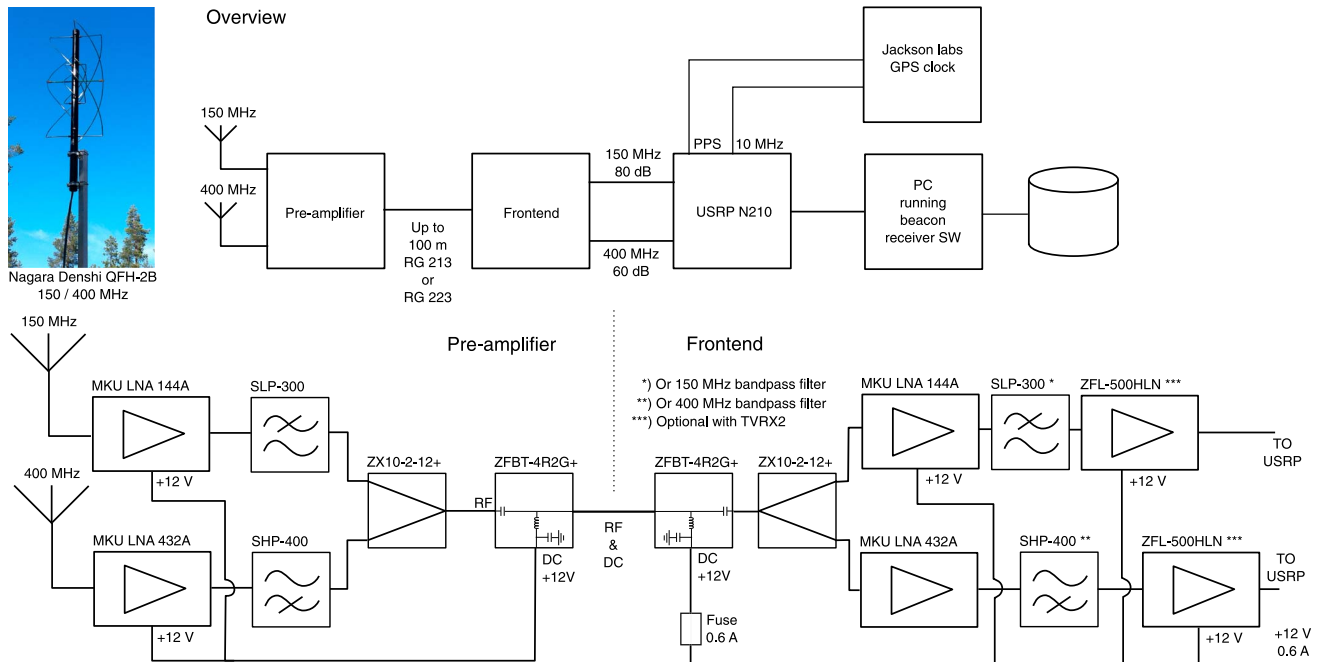


Figure 3. A block diagram of the RF front end used in the FMI beacon receiver chain.

Each day, Jitter automatically retrieves new two-line element orbital ephemeris files over the internet from publicly available sources (e.g., CelesTrak) and calculates satellite overflights for the next 24 h for the configured satellites. This information is written in a compact schedule file that is read by the data recording process. The ephemeris file retrieval and satellite overflight prediction programs are written in Python. The prediction program uses the PyEphem ephemeris package for calculating satellite positions.

The data recording process continuously receives wideband (typically 1 MHz) signals around the requested frequencies (typically 150 and 400 MHz). This allows the receiver to simultaneously record narrowband signals from multiple satellites transmitting different center frequencies, as long as their center frequency and Doppler shift fits within the wide receiver band. The data recorder process digitally tunes to the center frequency of each satellite when a predicted satellite pass occurs. The signal is digitally band-pass filtered and decimated, and the signal from an individual pass is stored on local media in complex single-precision floating point with a sample rate of 40 kHz, which is still large enough to accommodate the Doppler shift of an individual satellite pass. The data recording program is written in C++ using the Gnuradio framework.

The phase curve analysis process reads in newly recorded data files for a satellite pass. It then estimates the Doppler shift of the signal based on the ephemeris file (i.e., $L(t)$) for the satellite. This information is then used to remove the bulk of the Doppler shift of the satellite signal. After this, the remaining Doppler shift is estimated from the data and corrected. This is done by least squares fitting an outlier-removed seventh-order polynomial to the peak of the dynamic spectrum of the satellite signal. These operations are performed in an identical (albeit frequency scaled) way for both receiver channels to maintain their phase coherence. Finally, the data are band-pass filtered and decimated with a very narrowband filter (typically 50 Hz), and the differential phase curve $\Delta\phi_t$ is formed by subtracting the frequency scaled phase of one channel from the other $\Delta\phi_t = \frac{3}{8}\phi_t^{400} - \phi_t^{150}$ (last term of equation (16)). The resulting relative phase curve and satellite trajectory is stored in a data file with appropriate metadata (e.g., satellite ID, timestamp, and receiver location). The phase curve analysis program and plotting routines are written in Python, and the data are stored in HDF5 [Folk et al., 2011] format.

8. Satellites

A number of beacon satellites currently exist on orbit with transmissions that can be received with the designed system, but not all of these transmit above a given receiver location for various reasons (e.g., incompatible orbital inclination). In Scandinavia, we have successfully received signals from COSMOS 2407,

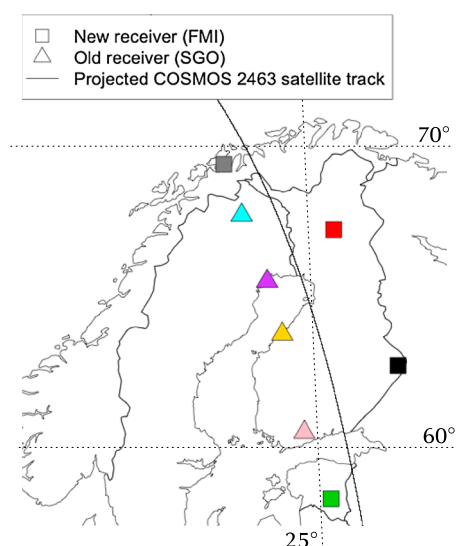


Figure 4. The new TomoScand receiver stations deployed by FMI, the old SGO receiver chain, and the projected satellite path.

COSMOS 2414, COSMOS 2429, COSMOS 2463, COSMOS 2454, RADCAL, and DMSP F15. With one receiver located in Ethiopia, we have also recorded the beacon on the Communications/Navigation Outage Forecasting System satellite. We have also recently received the Coherent Electromagnetic Radio Tomography beacon on the newly launched ePOP satellite [Bernhardt and Siefing, 2006].

9. Example Measurement

We perform an example study with COSMOS 2463 satellite overflight measurements collected from Scandinavia on 2013-10-23 at 09:50 UTC and produce a tomographic reconstruction of ionospheric electron density from the collected data. We chose this particular case due to the projection of the overflight passes between the receivers, as can be seen in Figure 4.

Data were collected with beacon satellite receivers designed as described in this paper and installed between 2011 and 2013 by the Finnish Meteorological Institute on the following four stations: TRO, SOD, MEK, and TAR (Table 1). The locations are also shown with colored squares on the map in Figure 4.

The satellite overflight takes approximately 15 min, and we assume a temporally stationary ionosphere for that duration. In two-dimensional tomography, we also assume a longitudinally uniform ionosphere in the observation domain.

The software employed for two-dimensional tomography in this example is described in the companion paper by J. Norberg et al., (Ionospheric tomography in Bayesian framework with Gaussian Markov random field priors, submitted to *Radio Science*, 2014). The software has been developed as part of Finnish Meteorological Institute (FMI)'s TomoScand project with the goal of three-dimensional ionospheric tomography using different data sources, e.g., beacon, GPS satellite, and ionosonde measurements.

In addition to showing results from new receivers, we compare the results obtained from the same satellite overflight with an existing receiver chain of four receivers operated by SGO, shown in Figure 4 with triangles.

9.1. Measurements

The phase difference curves measured by each receiver are shown in Figure 5. As explained in section 2, the presented curves are relative measurements. Hence, at this point we cannot make conclusions from the different absolute levels of the measurements, and instead focus on their relative shape. From the SGO curves (Figure 5, top) using existing receivers, we see that the measurement from Luleå (violet) and Kiruna (cyan) are noisier and contain probable measurement artifacts. For example, the Kiruna data close to 70° latitude shows a probable phase jump. The measurements with the new receivers shown in Figure 5 (bottom) are consistent when compared to each other. When comparing the two data sets, we find good overall correspondence between adjacent measurements of different receiver types.

To process with tomographic inversion, we first reduce the phase curve sampling rate by averaging. However, the high-resolution phase curves are simultaneously used to estimate the variances of the measurement error. Thus, we get an error variance estimate for each measurement point of the low-resolution phase curve used in the actual inversion computation. While estimating the measurement error variance, we also obtain information that can be used to handle the phase jump errors. Specifically, we set a threshold value for the variance and if it is exceeded, we assume that the signal has experienced disturbances that have caused unknown phase jumps. As mentioned at the end of section 3, we consider the measurements after the disturbance as a new independent observation with a new unknown phase offset.

9.2. Inversion

The computational domain for the tomographic reconstruction is defined by the measurement quality and geometry. For this work, we restrict the tomographic latitude range to 40°–78° latitude. The minimum

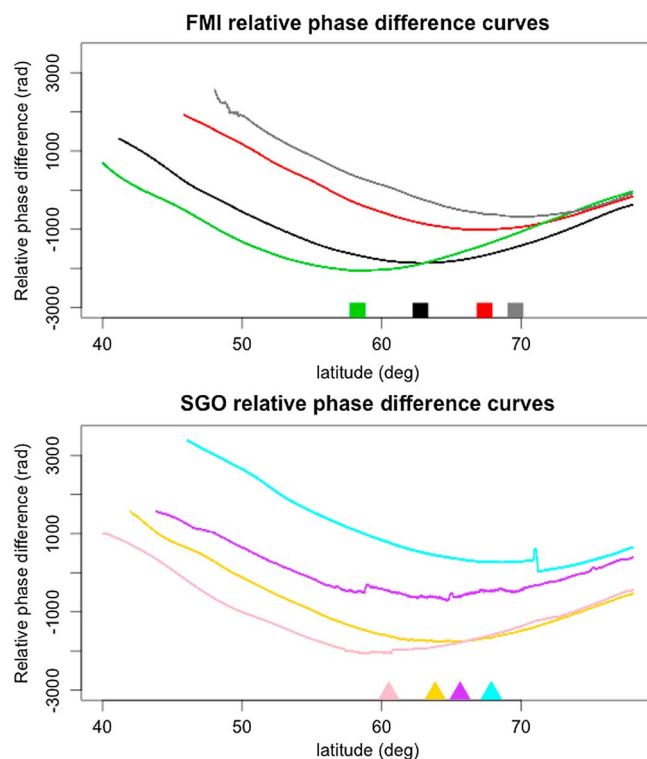


Figure 5. The relative phase difference curves from old receivers of SGO and new receivers of TomoScand/FMI receivers. Latitudes of each receiver station are marked by symbols.

values, and then fine tune the parameters so that the reconstruction starts to resemble the existing SGO receiver network reconstruction with the same data. This is mostly to show that regardless of the differences between the algorithms, it should be possible to produce reasonably similar reconstructions with them, when the same data are used.

We begin with a scale height of 140 km and a peak altitude of 280 km for both the mean and the standard deviation. The peak value for mean prior electron density is set to $4 \cdot 10^{11} \text{ m}^{-3}$ and the standard deviation to $2 \cdot 10^{11} \text{ m}^{-3}$. The higher the standard deviation is, the less we are forcing the reconstruction to follow the prior mean. The correlation length is defined here as the distance where the correlation between two points drops under one tenth of the variance. The horizontal correlation length is set to 4° and the vertical to 400 km. This gives us a rather loose prior where large variations over a given mean value are likely.

After initial results, we then change the scale height to 120 km, the peak electron density to $2 \cdot 10^{11} \text{ m}^{-3}$, the peak of the standard deviation to $1 \cdot 10^{10} \text{ m}^{-3}$, and the horizontal correlation length to 8° . This results in a much more restrictive prior distribution.

The SGO tomography algorithm is described by Markkanen *et al.* [1995]. It uses the statistical inversion method described in section 4 with slightly different prior assumptions than the ones in this study.

9.3. Results

As the information provided by the measurements becomes very small when the distance between satellite and receivers increases, reconstructions in Figure 6 are carried out in a smaller grid than is used for the actual computation.

When comparing the results between the two data sets, it is important to keep in mind that the receiver locations have significant differences in the longitudinal direction, as can be seen from Figure 4. However, we assume longitudinal uniformity of the ionosphere during the overflight based on separate measurements from the Magnetometers-Ionospheric Radars-All-sky Cameras Large Experiment magnetometer network (not shown).

elevation angle for measurements is set to 10° . The maximum altitude is set to 1000 km to accommodate the satellite orbit altitude. The pixel size is 0.25° in the horizontal direction and 25 km in the vertical direction. This results in a 160×40 grid. The theory matrix connecting the measurements to grid points is then formed by discrete linear approximation of the line integrals of satellite signal paths.

We form the inverse prior covariance matrix Σ_{pr}^{-1} of equation (28) following the approach introduced by Roininen *et al.* [2011, 2013]. The prior distribution is parametrized with mean value, standard deviation, and directional correlation lengths. In the vertical direction, the mean value and the standard deviation are given here using Chapman profiles [Brekke, 1997]. The Chapman ionization profile is determined by its peak value, peak altitude, and scale height.

As the prior parameters have an actual physical meaning, we first choose arbitrary but realistic prior parameter

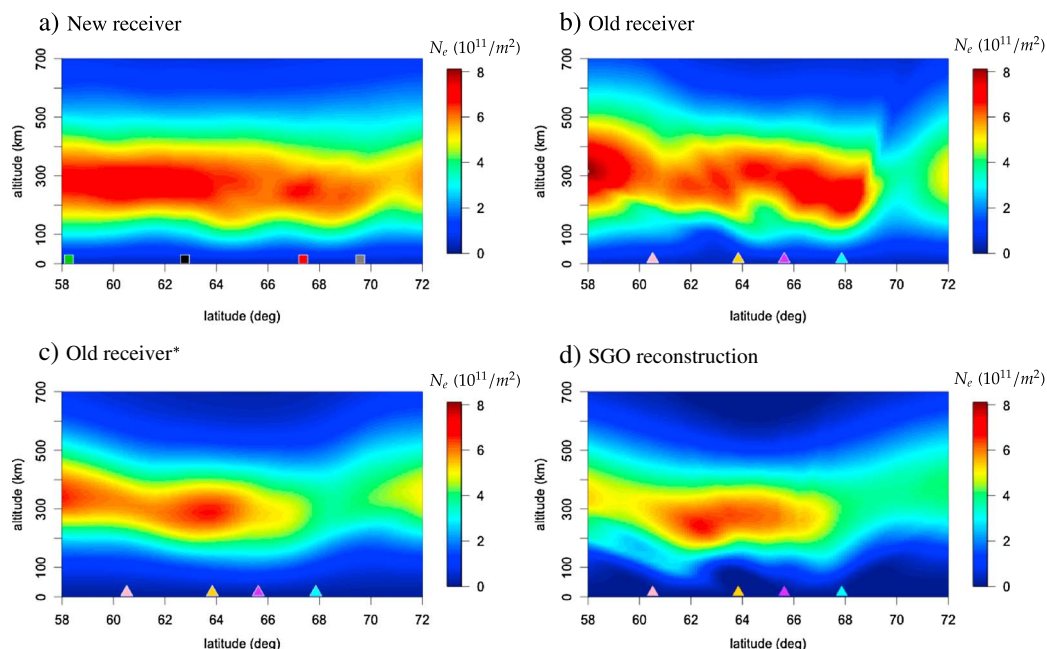


Figure 6. Tomographic reconstructions using new and old receiver data. (a) A reconstruction is made with the new algorithm with the initial prior and from the new receiver data. (b) The new algorithm with the initial prior and with the old receiver data. (c) The new algorithm with a stronger prior (smoother) and old receiver data. (d) A reconstruction with SGO's algorithm and old receiver data.

Figure 6a shows the tomographic reconstruction with the new algorithm using looser prior parameter values, based on data collected with the new receivers. Figure 6b plots the reconstruction with the same algorithm and parameters from SGO receiver data. The reconstructions have considerable similarities, but the reconstruction from the new receiver data is generally smoother throughout the reconstructed ionosphere. With the data from SGO receivers, in the reconstruction between the latitudes of 61° and 63° , there are also some faint structures in the ray direction that are quite clearly processing artifacts.

Figure 6c plots the results from the new algorithm with the SGO receiver data, but now with the stronger prior assumption. Figure 6d plots the reconstruction with SGO's production tomography algorithm, the SGO receiver data. When comparing these Figures 6c and 6d, we see that the algorithms are consistent when correspondingly strong prior distribution is used. In general, both reconstructions suffer from wave-like structure around the receivers. This would suggest that the prior is pulling the electron density too low on areas with less information. This could probably be corrected by adjusting the prior distribution.

10. Conclusions

We have implemented a new beacon satellite receiver system for 150/400 MHz dual-band transmissions. The current software is used for routine operational measurements. Once configured and operational, the receiver software automatically downloads the satellite ephemeris files and records all satellite passes. The software then performs a phase curve analysis on these measurements with another background process.

We have also produced realistic tomographic reconstruction of the ionospheric electron density with the data provided by our newly developed and newly deployed receivers. We have also shown that the tomographic algorithm used in this study gives results comparable to the ones produced by the existing SGO receiver system. The measured phase curves of the new receiver typically have fewer phase jumps than the SGO system, which indicates that the new receiver is more sensitive.

11. Future Work

After a more rigorous validation of the system using direct electron density profiles from incoherent scatter radar observations, future work will focus on deploying more 150/400 MHz beacon satellite receivers to obtain better geographical coverage. We will also focus on improving our tomographic algorithms in

such a way that they can incorporate auxiliary information about electron densities from ionosondes, incoherent scatter radars, and GPS receivers. Eventually, the developed algorithm will be extended to the three-dimensional case.

Acknowledgments

This work has been funded by Academy of Finland (applications 132694 and 250215) and European Regional Development Fund (Regional Council of Lapland, application A31221). We would also like to thank Paavo Törmänen and Tomi Teppo for help in building the beacon satellite receivers.

References

- Appleton, E. V., and F. W. Chapman (1932), The collisional friction experienced by vibrating electrons in ionized air, *Proc. Phys. Soc. London*, 44(243), 246–254.
- Hartree, D. R. (1929), Magneto-ionic propagation of electro-magnetic waves in the ionosphere, *Proc. Cambridge Philos. Soc.*, 25, 47.
- Austen, J. R., S. J. Franke, and C. H. Liu (1988), Ionospheric imaging using computerized tomography, *Radio Sci.*, 23(3), 299–307.
- Bernhardt, P. A., and C. L. Siefring (2006), New satellite-based systems for ionospheric tomography and scintillation region imaging, *Radio Sci.*, 41, RS5S23, doi:10.1029/2005RS003360.
- Bernhardt, P. A., et al. (1998), Two-dimensional mapping of the plasma density in the upper atmosphere with computerized ionospheric tomography (CIT), *Phys. Plasmas*, 5, 2010–2021.
- Bernhardt, P. A., C. A. Selcher, S. Basu, G. R. Bust, and S. C. Reising (2000), Atmospheric studies with the Tri-Band Beacon Instrument on the COSMIC constellation, *TAO*, 11(1), 291–312.
- Brekke, A. (1997), *Physics of the Upper Polar Atmosphere*, Wiley, West Sussex, England.
- Calvetti, D., and E. Somersalo (2007), *Introduction to Bayesian Scientific Computing—Ten Lectures on Subjective Computing*, Springer, New York.
- Danchik, R. J. (1998), An overview of transit development, *Johns Hopkins APL Technical Digest*, 19(1).
- Davies, K. (1965), *Ionospheric Radio Propagation*, Natl. Bur. of Stand., Washington, D. C.
- Folk, M., et al. (2011), An overview of the HDF5 technology suite and its applications, in *Proceedings of the EDBT/ICDT 2011 Workshop on Array Databases*, ACM, New York.
- Garner, T. W., T. L. Gaussiran II, J. A. York, D. M. Munton, C. M. Slack III, and A. M. Mahrous (2009), Ionospheric tomography network of Egypt: A new receiver network in support of the international heliophysical year, *Earth, Moon, Planets*, 104, 227–235.
- Izsak, I. G. (1960), Orbit determination from simultaneous dopplershift measurements, *SAO Spec. Rep.* 38.
- Kaipio, J., and E. Somersalo (2005), *Statistical and Computational Inverse Problems*, Springer, New York.
- Kunitsyn, V. E., and E. D. Tereshchenko (2003), *Ionospheric Tomography*, Springer, Berlin.
- Kunitsyn, V. E., E. S. Andreeva, E. D. Tereshchenko, B. Z. Khudukon, and T. Nygrn (1994), Investigations of the ionosphere by satellite radiotomography, *Int. J. Imaging Syst. Technol.*, 5(2), 112–127.
- Lassen, H. (1927), Über den Einfluss des Erdmagnetfeldes auf die Fortpflanzung der elektrischen Wellen der drahtlosen Telegraphie in der Atmosphäre, *Elektr. Nachr. Technik*, 4, 324–334.
- Markkanen, M., M. S. Lehtinen, T. Nygrén, J. Pirttilä, P. Helenius, E. Vilenius, E. D. Tereshchenko, and B. Z. Khudukon (1995), Bayesian approach to satellite radiotomography with applications in the Scandinavian sector, *Ann. Geophys.*, 13, 1277–1287.
- Mitchell, C. N., L. Kersley, J. A. T. Heaton, and S. E. Pryse (1997), Determination of the vertical electron-density profile in ionospheric tomography: Experimental results, *Ann. Geophys.*, 15, 747–752.
- Mitola, J., III (2000), *Software Radio Architecture: Object Oriented Approaches to Wireless Systems Engineering*, John Wiley, New York.
- Nygrén, T., M. Markkanen, M. S. Lehtinen, E. D. Tereshchenko, B. Z. Khudukon, O. V. Khudukon, O. V. Evstafiev, and P. Pollari (1996), Comparison of F-region electron density observations by satellite radio tomography and incoherent scatter methods, *Ann. Geophys.*, 14, 1422–1428.
- Nygrén, T., M. Markkanen, M. S. Lehtinen, E. D. Tereshchenko, and B. Z. Khudukon (1997), Stochastic inversion in ionospheric radiotomography, *Radio Sci.*, 32(6), 2359–2372.
- Rideout, W., and A. Coster (2006), Automated GPS processing for global total electron content data, *GPS Solutions*, 10, 219–228.
- Reed, J. H. (2002), *Software Radio*, Communications Engineering and Emerging Technologies Series, Prentice-Hall, Upper Saddle River, N. J.
- Roininen, L., M. S. Lehtinen, S. Lasanen, M. Orispää, and M. Markkanen (2011), Correlation priors, *Inverse Prob. Imag.*, 5(1), 167–184.
- Roininen, L., P. Piiroinen, and M. S. Lehtinen (2013), Constructing continuous stationary covariances as limits of the second-order stochastic difference equations, *Inverse Prob. Imag.*, 7(2), 611–647.
- Raymund, T. D., J. R. Austen, S. J. Franke, C. H. Liu, J. A. Klobuchar, and J. Stalker (1990), Application of computerized tomography to the investigation of ionospheric structures, *Radio Sci.*, 25(5), 771–789.
- Rüeger, J. M. (2002), Refractive index formulae for radio waves, *Proc. FIG XXII International Congress*, Washington, D. C.
- Yamamoto, M. (2008), Digital beacon receiver for ionospheric TEC measurement developed with GNU Radio, *Earth Planets Space*, 60, e21–e24.

Publication III

J. Norberg, I. I. Virtanen, L. Roininen, J. Vierinen, M. Orispää, K. Kauristie, and M. S. Lehtinen, Bayesian statistical ionospheric tomography improved by incorporating ionosonde measurements, *Atmos. Meas. Tech.*, 9(4): 1859–1869, 2016. ISSN 18678548. doi: 10.5194/amt-9-1859-2016.



Bayesian statistical ionospheric tomography improved by incorporating ionosonde measurements

Johannes Norberg^{1,2}, Ilkka I. Virtanen³, Lassi Roininen^{2,4}, Juha Vierinen⁵, Mikko Orispää^{2,4}, Kirsti Kauristie¹, and Markku S. Lehtinen^{2,4}

¹Finnish Meteorological Institute, Helsinki, Finland

²Sodankylä Geophysical Observatory, University of Oulu, Sodankylä, Finland

³Ionospheric Physics, University of Oulu, Oulu, Finland

⁴Department of Mathematics, Tallinn University of Technology, Tallinn, Estonia

⁵Haystack Observatory, Massachusetts Institute of Technology, Westford, MA, USA

Correspondence to: Johannes Norberg (johannes.norberg@fmi.fi)

Received: 26 August 2015 – Published in Atmos. Meas. Tech. Discuss.: 21 September 2015

Revised: 15 March 2016 – Accepted: 17 March 2016 – Published: 28 April 2016

Abstract. We validate two-dimensional ionospheric tomography reconstructions against EISCAT incoherent scatter radar measurements. Our tomography method is based on Bayesian statistical inversion with prior distribution given by its mean and covariance. We employ ionosonde measurements for the choice of the prior mean and covariance parameters and use the Gaussian Markov random fields as a sparse matrix approximation for the numerical computations. This results in a computationally efficient tomographic inversion algorithm with clear probabilistic interpretation.

We demonstrate how this method works with simultaneous beacon satellite and ionosonde measurements obtained in northern Scandinavia. The performance is compared with results obtained with a zero-mean prior and with the prior mean taken from the International Reference Ionosphere 2007 model. In validating the results, we use EISCAT ultra-high-frequency incoherent scatter radar measurements as the ground truth for the ionization profile shape.

We find that in comparison to the alternative prior information sources, ionosonde measurements improve the reconstruction by adding accurate information about the absolute value and the altitude distribution of electron density. With an ionosonde at continuous disposal, the presented method enhances stand-alone near-real-time ionospheric tomography for the given conditions significantly.

1 Introduction

In ionospheric satellite tomography the electron density distribution of the ionosphere is reconstructed from ground-based measurements of satellite-transmitted radio signals. The use of tomographic methods for ionospheric research was first suggested by (Austen et al., 1988). (Bust and Mitchell, 2008) provide a good overview on the development of the topic.

Mathematically ionospheric tomography is an ill-posed inverse problem and cannot be solved without some additional stabilization or regularization information. In ionospheric tomography the additional information is often incorporated with the use of iterative reconstruction algorithms such as algebraic reconstruction technique with a strong initial model for the ionosphere (Andreeva, 1990). Bayesian statistical inversion was applied to ionospheric tomography first by (Markkanen et al., 1995). The Bayesian approach provides an interpretable approach for the stabilization as the additional information is given as a prior probability distribution of unknown parameters. However, in the work of (Markkanen et al., 1995), the prior distribution is not defined by its covariance, but by an assumption of smoothness resulting from the limiting of the differences of neighboring pixels. This is an often valid assumption, but the relation between the prior parameters and the physical quantities is not clear. Recently, (Norberg et al., 2015) have described a method in which the prior can be built in a computationally efficient

way as a probability distribution with a known covariance structure. The prior is parameterized with physical units and can be understood as a probability distribution for realizations of the ionosphere.

Regardless of the tomographic algorithm in use, the information provided by satellite to ground measurements is poor in the vertical direction. This is due to the limited measurement geometry, namely the lack of horizontal signal paths. Consequently, the peak altitude and the vertical gradient of the reconstructed ionosphere will be determined mostly by the regularizing prior assumptions that are built in to the employed tomography algorithm. In this study we employ the ionosonde measurements to give these assumptions for the vertical profile.

An ionosonde is a radar used to investigate the ionosphere. An ionosonde transmits electromagnetic frequency pulses, sweeping through the high-frequency (HF) range, and receives the signals reflected from an altitude where the radar frequency matches a critical frequency (Breit and Tuve, 1926). For ordinary mode polarization the critical frequency is the plasma frequency of the local electron density. Because refractive index along the signal path differs significantly from that of a vacuum, conversion of signal travel time into reflection height is not trivial, but the electron density profile along the path needs to be taken into account. The reflections and the travel times at multiple frequencies can be used to estimate an electron density profile of the ionosphere. Because the ionosonde relies on reflection, it can directly measure only the bottom side of the ionospheric altitude profile up to the peak of the electron density profile. Also, it is not very effective for observing local minima, e.g., the valley region between the E and F regions of the ionosphere. Ionosonde measurements provide recurrent and accurate but geographically localized information of the ionospheric electron density profile. In mesoscale tomographic analysis, it is often the best information available, even if the analyzed region is somewhat displaced from the ionosonde site.

Inclusion of ionosonde measurements in ionospheric tomography has been studied by (Kersley et al., 1993), where ionosonde measurements were used to form the background profile for an iterative reconstruction algorithm. The study had mixed results on the impact of ionosonde measurement inclusion. They also observed up to 70 % differences between the ionosonde and incoherent scatter radar-derived electron density profiles. More recently (Chartier et al., 2012) used ionosonde measurements to set vertical basis functions for the inversion, as well as using them as local measurements of peak density and bottom-side profile gradients. The inclusion improved the tomographic results significantly, but the sensitivity to ionosonde measurement bias was also underlined. (Chiang and Psiaki, 2014) also combined ionosonde data with GPS measurements for ionospheric tomography. The presented method concentrates on estimating parameterized local electron density profile at the location of the ionosonde. For latitudinal and longitudinal changes, only

the first-order dependence of vertical total electron content was considered.

In this article we continue the work presented in (Norberg et al., 2015) and include the ionosonde measurements in the Bayesian statistical inversion approach for ionospheric tomography. For comparison, we analyze the data also with the prior mean taken from the International Reference Ionosphere (IRI) model, and with a zero-mean prior. The IRI model is chosen as it is a well-known ionospheric model, and unlike the ionosonde, it provides information also on horizontal electron density gradients. The zero-mean prior is included to demonstrate the performance with simpler and more general prior information. The zero-mean prior carries essentially similar information to the prior model used in (Markkanen et al., 1995). We construct the prior mean electron density profile for the entire ionospheric tomography domain according to the chosen information source. This assumption is then controlled with the prior covariance, as it states how strictly the reconstruction should follow the prior mean. As the prior distribution is parameterized with physical units, the method provides clear understanding on information used for the tomographic reconstruction. Hence the approach makes the inversion possible with less ad hoc adjustment. This is a very important aspect for achieving reliable operational near-real-time tomography results.

The approach is applied to Scandinavian sector with tomographic measurements from the TomoScand receiver chain (Vierinen et al., 2014) and ionosonde data from the European Incoherent Scatter Scientific Association (EISCAT) dynamometer in Tromsø, Norway. The IRI model used for the comparison is the International Reference Ionosphere 2007 (IRI-2007) (Bilitza and Reinisch, 2008). We validate the results with EISCAT ultra-high-frequency (UHF) incoherent scatter radar measurements carried out on 20 and 21 November 2014 and 11 and 14 March 2015 in Tromsø.

2 Methodology

The dual-frequency signal transmitted from low Earth orbit (LEO) satellites consists of frequencies of 150 and 400 MHz. The ionospheric refraction causes a phase shift to propagating electromagnetic waves. This phase shift is proportional to density of electrons along the signal path (Davies, 1990) and can be modeled as

$$m_{t, \text{sat}, \text{rec}} = \gamma_{\text{sat}, \text{rec}} + \int_{L_{t, \text{sat}, \text{rec}}} N_e(z) dl + \varepsilon_{t, \text{sat}, \text{rec}}, \quad (1)$$

where $m_{t, \text{sat}, \text{rec}}$ is the measured relative total electron content at time t between the satellite sat and receiver rec, and $\varepsilon_{t, \text{sat}, \text{rec}}$ the corresponding measurement error. $N_e(z)$ is the two-dimensional continuous field of electron densities with coordinates $z = (z_1, z_2) \in \mathbb{R}^2$. The integral is defined over the measurement signal path $L_{t, \text{sat}, \text{rec}}$. The receiver–satellite

specific constant $\gamma_{\text{sat, rec}}$ is due to the unknown amount of electron content when the satellite is first observed.

For practical computations, we discretize Eq. (1) for all measurements. The discretized measurement model for the ionospheric tomography is given as

$$\mathbf{M} = \mathbf{A}\mathbf{X} + \mathbf{E}. \quad (2)$$

The measurement vector is $\mathbf{M} \in \mathbb{R}^{n_m}$. Theory matrix $\mathbf{A} \in \mathbb{R}^{n_m \times n_x}$ gives the measurement geometry between the satellite measurement points and receiver locations. The vector of unknown parameters $\mathbf{X} \in \mathbb{R}^{n_x}$ includes both electron densities and the 2π -ambiguity constants $\boldsymbol{\gamma}$. The measurement error vector is $\mathbf{E} \in \mathbb{R}^{n_m}$. The number of measurements is given as n_m and the number of unknown parameters as n_x .

Let us denote by \mathbf{x} and \mathbf{m} the realizations of the random variables \mathbf{X} and \mathbf{M} , respectively. We can then write the likelihood function for unknown parameters, given the measurements as

$$\mathcal{L}(\mathbf{x}|\mathbf{m}) = D_E(\mathbf{A}\mathbf{x} - \mathbf{m}), \quad (3)$$

where D_E is the probability density function of measurement errors. From here on we assume that $\mathbf{E} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_E)$: the measurement errors follow a multivariate normal distribution with zero mean and covariance $\boldsymbol{\Sigma}_E \in \mathbb{R}^{(n_m) \times (n_m)}$.

As the ionospheric tomography is an ill-posed problem, the maximum likelihood estimate for Eq. (3) cannot be solved without including some additional information regarding the unknown parameters. Here we use Bayesian statistical inversion (Markkanen et al., 1995; Kaipio and Somersalo, 2005) to give this information as a prior distribution. We assume that the unknown \mathbf{X} follows a multivariate normal distribution $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}_{\text{pr}})$, where vector $\boldsymbol{\mu} \in \mathbb{R}^{n_x}$ is the mean value and the matrix $\boldsymbol{\Sigma}_{\text{pr}} \in \mathbb{R}^{(n_x) \times (n_x)}$ the covariance. Again, the vector $\boldsymbol{\mu}$ as well as the matrix $\boldsymbol{\Sigma}_{\text{pr}}$ consists of parts for both the unknown electron densities and the unknown $\boldsymbol{\gamma}$ parameters. We denote the prior probability density function with $D_{\text{pr}}(\mathbf{x})$. Following the Bayes' theorem, we then obtain the posterior distribution for \mathbf{X} as

$$D_{\text{post}}(\mathbf{x}|\mathbf{m}) = \frac{D_E(\mathbf{A}\mathbf{x} - \mathbf{m})D_{\text{pr}}(\mathbf{x})}{\int_{\mathbb{R}^{n_x}} D_E(\mathbf{A}\mathbf{x} - \mathbf{m})D_{\text{pr}}(\mathbf{x})d\mathbf{x}}, \quad (4)$$

where the denominator is a normalization constant and we can write

$$D_{\text{post}}(\mathbf{x}|\mathbf{m}) \propto D_E(\mathbf{A}\mathbf{x} - \mathbf{m})D_{\text{pr}}(\mathbf{x}). \quad (5)$$

From the posterior distribution we can then derive the most probable value for the unknown parameters based on the prior distribution and observed measurements, namely, the maximum a posteriori estimator (MAP)

$$\mathbf{x}_{\text{MAP}} = \boldsymbol{\Sigma}_{\text{post}} \left(\mathbf{A}^T \boldsymbol{\Sigma}_E^{-1} \mathbf{m} + \boldsymbol{\Sigma}_{\text{pr}}^{-1} \boldsymbol{\mu} \right), \quad (6)$$

where

$$\boldsymbol{\Sigma}_{\text{post}} = \left(\mathbf{A}^T \boldsymbol{\Sigma}_E^{-1} \mathbf{A} + \boldsymbol{\Sigma}_{\text{pr}}^{-1} \right)^{-1} \quad (7)$$

is called posterior covariance.

As we assume that the unknown parameters follow multivariate normal distribution, the prior density function $D_{\text{pr}}(\mathbf{x})$ is defined with its mean and covariance. In Bayesian statistical approach for ionospheric tomography, the prior mean can be understood as the most probable state of the ionosphere before the actual satellite measurements. With the covariance we can express how reliable the information of prior mean is and how correlated the ionospheric electron densities are. Actual values of these parameters should be based on all information we have at our disposal, i.e., on other measurements, models, statistical data and the physics of ionosphere. In the performed experiments, we use three different schemes to compose the prior: IRI-2007 ionospheric model, zero mean and, most importantly, the ionosonde measurements. The prior covariance is given as a squared exponential, i.e., as a Gaussian-shaped function that is defined with its amplitude (variance or standard deviation) and correlation length. The correlation length is given separately for horizontal and vertical directions and is defined here as the distance where the covariance drops to 10 % of variance.

It is very natural to represent the prior information as a probability distribution. However, for the MAP estimator Eq. (6) only the precision matrix $\boldsymbol{\Sigma}_{\text{pr}}^{-1}$ (i.e., the inverse of the prior covariance) is required besides the prior mean. In (Norberg et al., 2015) it is shown how the precision matrix of a known covariance can be constructed with a sparse matrix representation with Gaussian Markov random fields. The approach provides us with the interpretation of a probability distribution, yet it keeps the approach computationally feasible, in comparison to operating with full covariance matrices.

Unfortunately, the linear system allows also negative values in the solution. A large proportion of negative values would suggest that the prior distribution differs drastically from the actual ionospheric conditions and needs to be reconsidered. Then again, small areas of negative values indicate that the model accuracy is less than the corresponding absolute values. Here, if some negative values are found, we add them as new measurements into the linear system. We then set these new measurements to zero with a small variance (10^{-20}) and solve the system again. We note that here this positivity constraint is mostly a cosmetic ad hoc method which will be reconsidered in future studies.

3 Experiments

Two EISCAT UHF incoherent scatter radar measurement campaigns were performed in November 2014 and March 2015. Three daytime and one nighttime COSMOS satellite overflights, suitable for two-dimensional tomography, were measured with TomoScand receivers starting approximately on 20 November 2014 at 12:50, 3 November 2015 at 13:50, 14 March 2015 at 13:20 and 21 November 2014 at 02:50 UTC. The magnetic local time is approximately UTC + 2.5 h. The altitude of COSMOS satellites is

approximately 1000 km and the duration of measurements from an overflight is roughly 10 min. For the ionosonde prior mean the NeXtYZ (Zabotin et al., 2006) analyzed EISCAT dynasonde results from Tromsø (see Sect. Data availability) were collected. The ionosonde electron density profiles that were measured during each satellite overflight were averaged together to form one profile. We denote the resulting profile with μ_{NeXtYZ} . The NeXtYZ provides also a modeled profile for the top-side ionosphere, but to gain better control over the prior, we here give the top side as an exponential profile. The complete altitude profile for the prior mean based on ionosonde measurement can be written as

$$\mu_{\text{ionosonde}}(z) = \begin{cases} \mu_{\text{NeXtYZ}}(z_{\text{peak}}) \exp\left(-\frac{z-z_{\text{peak}}}{s}\right), & z_{\text{peak}} < z \leq z_{\text{max}} \\ \mu_{\text{NeXtYZ}}(z), & 0 \leq z \leq z_{\text{peak}}, \end{cases} \quad (8)$$

where z is the altitude with the maximum $z_{\text{max}} = 1000$ (km) and $z_{\text{peak}} = \underset{z}{\operatorname{argmax}}(\mu_{\text{NeXtYZ}}(z))$, i.e., the altitude of the maximum electron density. The parameter s defines how rapidly the electron density decreases at the higher altitudes.

The IRI-2007 electron density profiles were taken for the reconstruction times with longitude parameter 26° . With the IRI-2007 we obtain a two-dimensional profile with latitudinal variation for the complete domain where the ionospheric tomography takes place.

To validate the resulting tomographic reconstructions, for each satellite overflight, the EISCAT UHF was set to perform a scan of four measurements along the corresponding satellite track. The altitude resolution used for EISCAT data analysis was 10 km. The UHF data were calibrated against the EISCAT Tromsø dynasonde. The calibration data were taken from periods when the radar was not scanning and the ionosphere was reasonably stable. Each few-hours-long continuous radar run was calibrated separately.

In the following three subsections we compare the EISCAT UHF measurements to corresponding electron density profiles from the obtained tomographic reconstructions. With the Overflight I the reconstruction was made multiple times to choose the measurement domain and prior parameters other than the prior mean. Based on these trials the measurements used for the tomography were limited between the latitudes of 55 and 75° and the elevation angles over 20° . The chosen sampling rate of 0.5 Hz then produces between 100 and 200 suitable measurements from each receiver station. The corresponding measurement errors are estimated from the original 20 Hz sampling rate data. The measurement errors are assumed to be independent, resulting in a diagonal measurement error covariance matrix.

The prior standard deviation (SD) is given as a Chapman function for the vertical profile, with approximately the same peak altitude as the prior mean, and the maximum electron density approximately 40 % of the corresponding NeXtYZ maximum. The Chapman profile was modified to have dif-

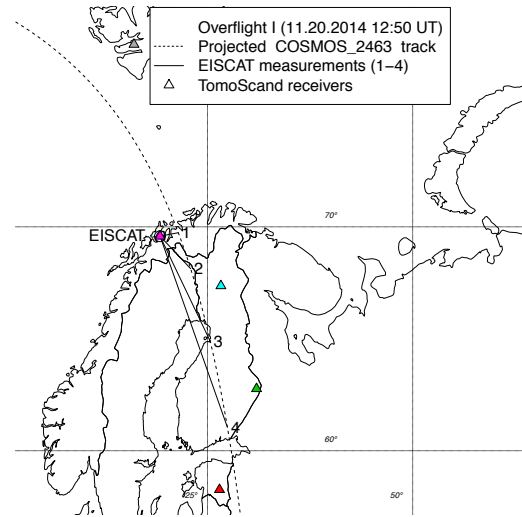


Figure 1. TomoScan receiver network and the satellite overflight ground track with four EISCAT UHF scan paths.

ferent scale heights for above and below the maximum. The chosen values used here are 200 and 60 km correspondingly. In vertical direction the prior correlation length was chosen to be 200 km and in the horizontal 2° . The s parameter for the upper profile of the prior was chosen to be 140 km. This results as a slightly steeper gradient for the top-side ionosphere than provided by NeXtYZ. With the zero-mean prior we use the same prior standard deviation as with the ionosonde case but, to allow larger changes in electron density, the maximum is set to 80 % of the NeXtYZ maximum. For all of the experiments, the prior mean values for the γ parameters are set to zero, and the prior standard deviations as large as 10^{20} m^{-3} , nearing an uninformative prior. The resolution for the domain is 200×200 , resulting in pixel size of $5 \text{ km} \times 0.1^\circ$.

For numerical reasons, the prior distribution is built to have periodic boundary conditions (Norberg et al., 2015). Here, the given vertical prior profile constrains the values of highest and lowest altitudes so strongly that boundary effects in that direction are prevented. To avoid boundary effects in horizontal direction, the correlation lengths at the boundaries are decreased to 10 % of the initial values and the actual inversion is carried out in a larger domain than is our actual interest.

After calibrating the parameters with the Overflight I, for the Overflights II and III the parameter values are adjusted only according to corresponding ionosonde measurements without additional tuning. For the Overflight IV the ionosonde profiles differ significantly from the previous ones. Hence also the prior standard deviation shape is adjusted to correspond to these conditions.

In each of the following cases we first visualize the general measurement setup on a map in Figs. 1, 3, 5 and 7. The results are presented in Figs. 2, 4, 6 and 8, first as

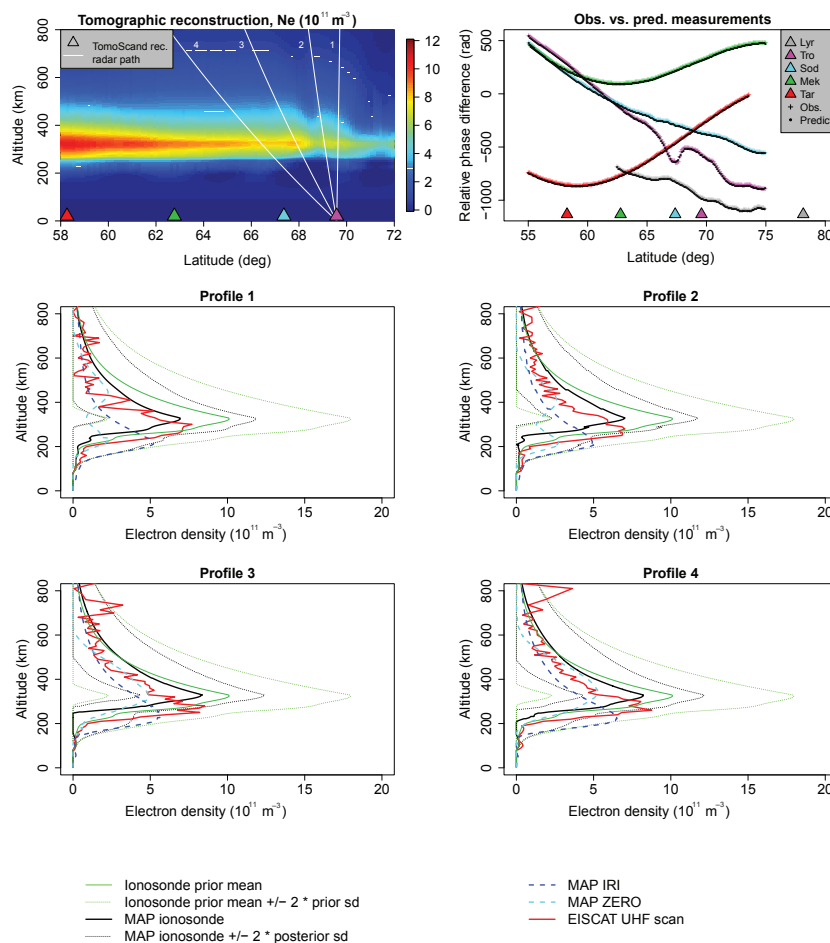


Figure 2. Reconstruction, phase curves and profile comparisons for Overflight I starting on 20 November 2014 at 12:50 UTC.

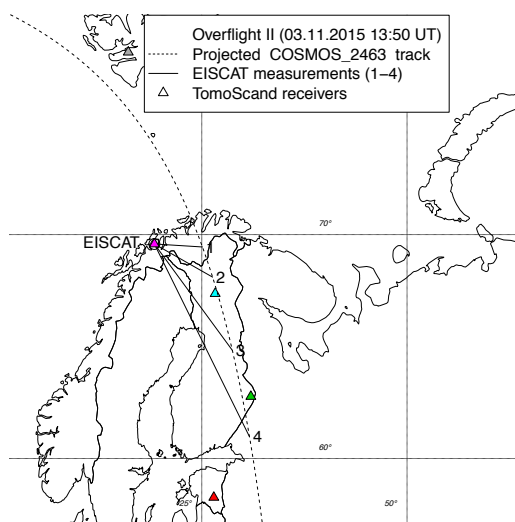


Figure 3. TomoScand receiver network and the satellite overflight ground track with four EISCAT UHF scan paths.

two-dimensional altitude–latitude reconstructions of electron densities, i.e., the MAP estimates where the ionosonde prior is used. On top of the reconstruction the EISCAT UHF scans are shown with white paths. We then compare the prior and posterior distribution parameters to corresponding EISCAT UHF scan locations by assuming a longitudinally uniform ionosphere. The ionosonde prior means are plotted with solid green lines and the 95 % prior credible intervals (Ionosonde prior mean $\pm 2 \times$ priorSD) with dashed green lines. The profiles taken from the reconstruction with ionosonde prior are plotted with solid black lines (MAP ionosonde) and the corresponding 95 % posterior credible intervals with dashed black lines (MAP ionosonde $\pm 2 \times$ posteriorSD). The electron density profiles obtained with EISCAT UHF scans are plotted with red. The blue dashed line is a profile taken from the reconstruction where the prior is based on IRI-2007 profile (MAP IRI) and the cyan dashed line from the reconstruction with zero-mean prior (MAP ZERO). In Table 1 the relative mean errors for profile peak electron densities and the mean errors for peak altitudes are given. In addition to the profile comparisons, we show the relative phase difference

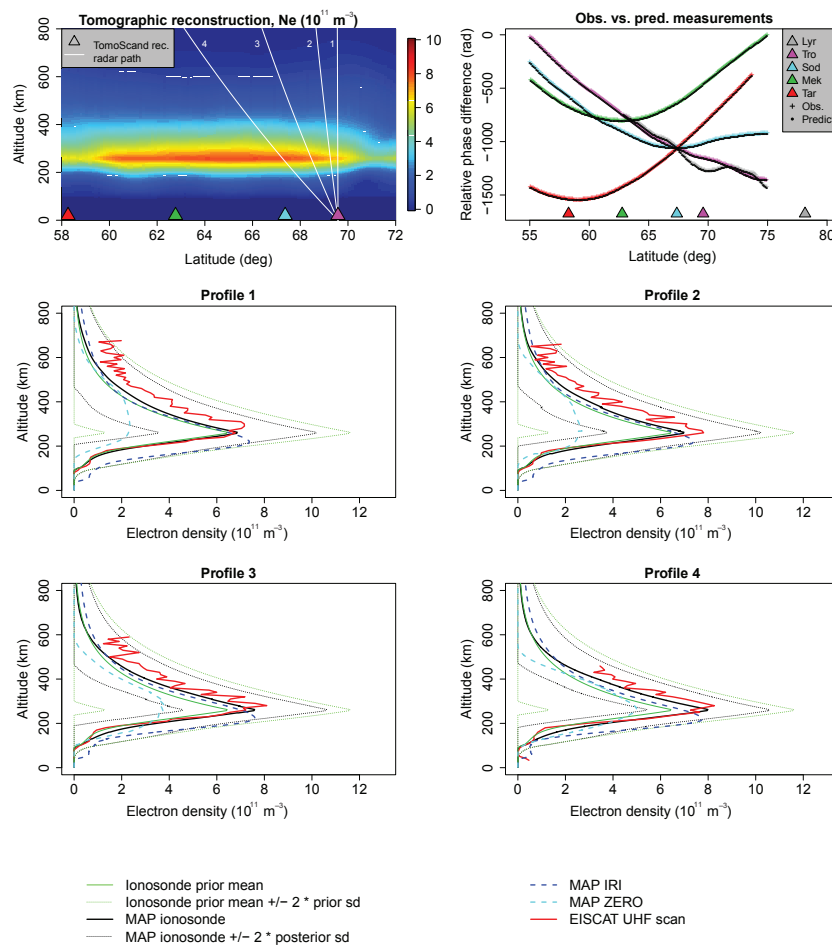


Figure 4. Reconstruction, phase curves and profile comparisons for Overflight II starting on 3 November 2015 at 13:50 UTC.

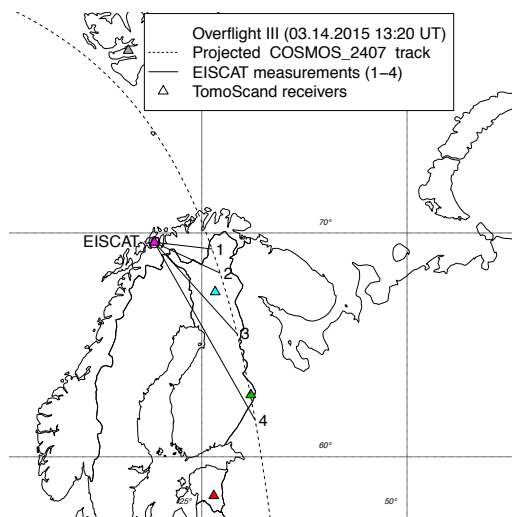


Figure 5. TOMOScand receiver network and the satellite overflight ground track with four EISCAT UHF scan paths.

measurements used for the inversion for each station, as well as the corresponding measurements predicted from the reconstruction obtained with ionosonde prior.

3.1 Overflight I

The COSMOS 2463 overflight (Fig. 1) starts on 20 November 2014 at 12:50 UTC. The direction of the satellite track is from north to south. The relative phase difference curves in the top right panel of Fig. 2 indicate smooth ionosphere, with some local structures visible in the Tromsø station curve. The ionosonde measurements used for the prior are from 12:54, 12:56, 12:58 and 13:00 UTC. The largest differences between the ionosonde profiles were at 330 km altitude, with standard deviation of $2.3 \times 10^{11} \text{ m}^{-3}$. The peak altitudes range from 320 to 340 km. In Fig. 2, the obtained tomographic reconstruction is shown in the top left panel. On top of the reconstruction are plotted the four EISCAT UHF measurements performed at (1) 12:53:00–12:54:10, (2) 12:55:03–12:56:03, (3) 12:56:20–12:57:20 and (4) 12:57:35–12:58:35 UTC. The latitude–longitude directions of the measurement can be seen in Fig. 1. Hourly aver-

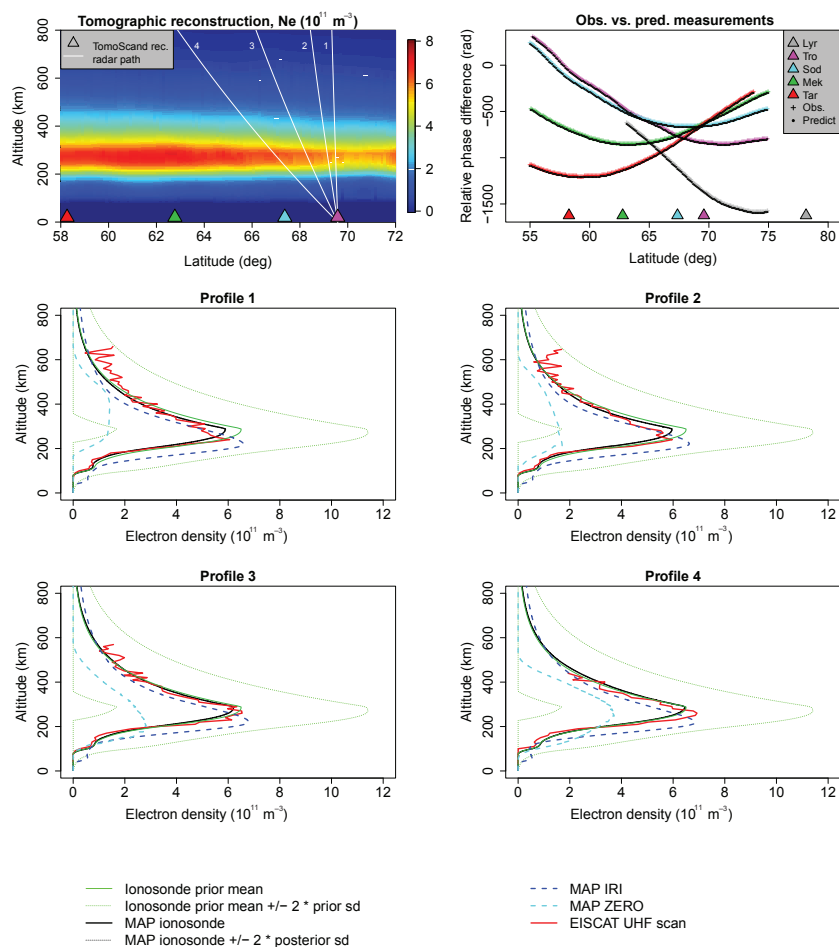


Figure 6. Reconstruction, phase curves and profile comparisons for Overflight III starting on 14 November 2015 at 13:20 UTC.

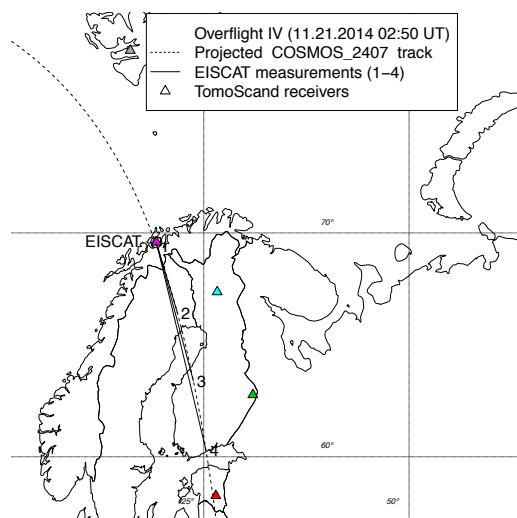


Figure 7. TomoScand receiver network and the satellite overflight ground track with four EISCAT UHF scan paths.

aged Kp and F10.7 indices at 13:00 UTC were 1.3 and 164.1, respectively.

The profile comparisons 1–4 in Fig. 2 show that the southward increment of electron density is captured by all three reconstructions. In the profiles based on the IRI-2007 and zero-mean prior reconstructions the maximum electron density is significantly lower than in the EISCAT UHF profiles and shape of the profiles clearly disagrees with the UHF measurements in comparisons 1 and 2. With IRI-2007 the peak altitude is underestimated in all of the profiles. The ionosonde prior shows a good agreement between shapes of the corresponding profiles. Although the satellite rises almost to zenith above Tromsø, F-region peak density estimates from the ionosonde are about 30 % higher than the calibrated UHF measurements. However, the prior standard deviation enables large enough changes to capture the correct level in the MAP estimate. With the ionosonde prior the most glaring difference between the UHF and tomographic profiles is in the altitude of the peak electron densities.

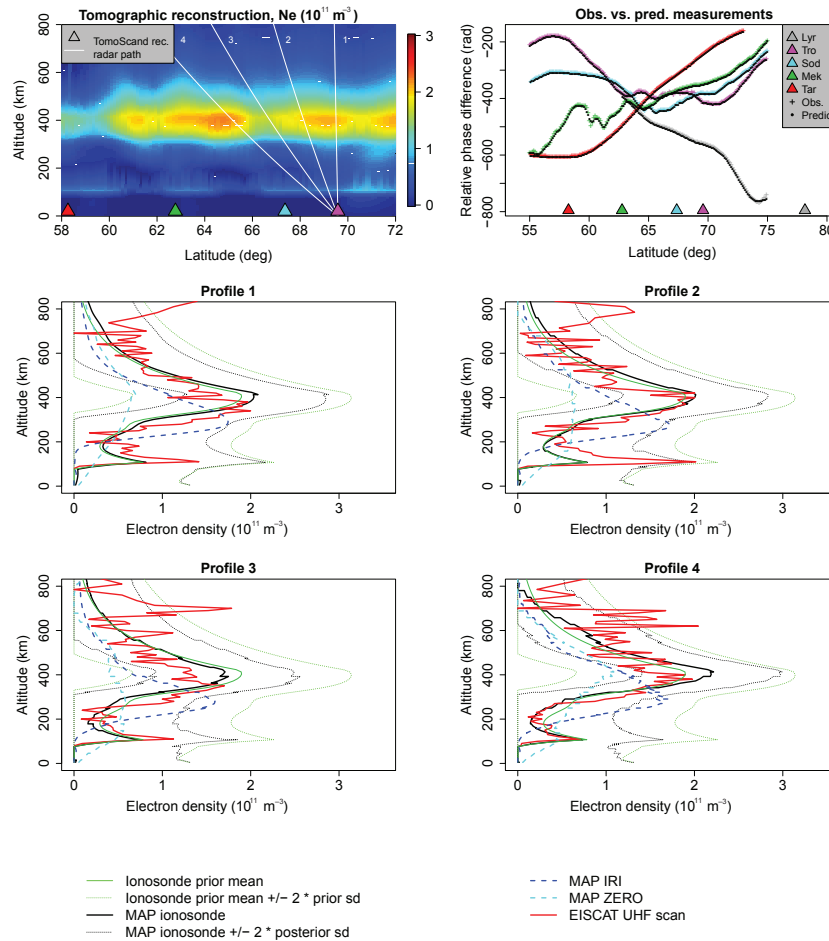


Figure 8. Reconstruction, phase curves and profile comparisons for Overflight IV starting on 20 November 2014 at 02:50 UTC.

3.2 Overflight II

The COSMOS 2407 overflight starts approximately on 3 November 2015 at 13:50 UTC (Fig. 3). The direction of the satellite track is from north to south. The relative phase difference curves in Fig. 4 indicate a smooth ionosphere. Based on the ionosonde measurements collected at 13:54, 13:56, 13:58 and 14:00 UTC the electron density level is expected to be lower than in Overflight I. The largest differences between the ionosonde profiles were at 260 km altitude, with standard deviation of $0.6 \times 10^{11} \text{ m}^{-3}$. The peak altitudes range between 260 and 280 km. The new prior profiles for this overflight are shown in the lower four panels of Fig. 4. Besides the altitude profiles for prior mean and standard deviation, the other parameters remain unchanged. In the top left panel of Fig. 4 the reconstruction and the EISCAT UHF measurement projections from (1) 13:54:28–13:55:28, (2) 13:55:50–13:56:50, (3) 13:57:11–13:58:11, and (4) 13:58:26–13:59:30 UTC are shown. Hourly averaged Kp and F10.7 indices at 14:00 UTC were 2.3 and 129.9, respectively.

The IRI-based profiles have very good agreement with the maximum densities of EISCAT scans. However the peak altitude is underestimated. The profiles taken from the reconstruction with zero-mean prior clearly disagree with the UHF measurement, in terms of both profile shape and peak electron density.

With the ionosonde-based prior, in Profile comparison 1 the prior mean and the closest UHF measurement are very similar and also the tomographic reconstruction is almost unchanged from the prior profile. Again, the electron density slightly increases southwards, which is well captured in the reconstruction. Both the peak density and altitude are very close to each other between the reconstruction and UHF profiles.

3.3 Overflight III

The COSMOS 2407 overflight starts on 14 March 2015 at 13:20 UTC (Fig. 5). The direction of the satellite track is from north to south. The ionosonde measurements used for the prior were collected at 13:26, 13:28, 13:30 and 13:32 UTC. The largest differences between the ionosonde

profiles were at 310 km altitude, with standard deviation of $0.5 \times 10^{11} \text{ m}^{-3}$. The peak altitudes range from below 250 km to almost 320 km. The reconstruction and the EISCAT UHF measurement directions at (1) 13:27:45–13:28:45, (2) 13:29:01–13:30:02, (3) 13:30:20–13:31:21 and (4) 13:31:35–13:32:35 UTC are shown in the top left panel of Fig. 6. Hourly averaged Kp and F10.7 indices at 13:00 UTC were 1.7 and 114.3, respectively.

With IRI prior the maximum densities are slightly pronounced and the peak altitude remains below the UHF peak. With the zero-mean prior both the profile shapes and peak densities clearly disagree with the UHF, again. For the ionosonde case the best agreement in general profile shape is again visible, even though the errors in peak altitudes and densities are in the same level with the IRI-based reconstructions.

3.4 Overflight IV

The COSMOS 2407 overflight starts on 21 November 2014 at 02:50 UTC (Fig. 7). Direction of the satellite track is from north to south. The relative phase difference curves in Fig. 8 indicate more small-scale structures in ionosphere than in the previous measurements. The ionosonde measurements were collected at 02:56, 02:58, 03:02 and 03:04 UTC, as the data for 03:00 are missing. The largest differences between the ionosonde profiles were at 180 km altitude, with standard deviation of $0.7 \times 10^{11} \text{ m}^{-3}$. The measurements show a strong E region at 100 km altitude. As the ionosonde measurements indicate that the electron density is not concentrated on one altitude, the maximum of the prior standard deviation is here set to the lower E-region peak of the ionosonde profile and the upper-scale height is increased to 600 km to allow more variation also around the higher F-region peak. Otherwise the prior profiles are formed similarly to previous cases. The reconstruction and the EISCAT UHF measurement directions at (1) 02:57:40–02:58:50, (2) 02:59:15–03:00:30, (3) 03:00:50–03:02:05 and (4) 03:02:25–03:03:35 UTC are shown in the top left panel of Fig. 8. Hourly averaged Kp and F10.7 indices at 03:00 UTC were 3.3 and 158.6, respectively.

With IRI prior an F region is visible, although at the wrong altitude, but the E-region peak is completely missing. The zero-mean prior spreads electron density also to lower altitudes, but it cannot distinguish the two-peak structure. With ionosonde the shape of the reconstruction seems to be strongly dictated by the prior. Horizontal gradients in F-region peak density are rather well reproduced in the reconstruction, whereas the reconstructed E-region peak is almost unchanged in the profile comparisons, although the UHF radar shows significantly different peak density at each pointing direction. In the reconstruction in the upper left panel of Fig. 8 a southward decrement in E-region density is visible between the receivers, where the information provided by the measurements is higher. Directly above the receivers

Table 1. Errors of tomographic profiles compared with EISCAT UHF scans.

		Relative mean error of peak density (%)	Mean error of peak altitude (km)
Overflight I	Ionosonde	5	41
	IRI	27	55
	Zero	52	74
Overflight II	Ionosonde	5	17
	IRI	6	58
	Zero	54	15
Overflight III	Ionosonde	4	33
	IRI	6	31
	Zero	60	33
Overflight IV	Ionosonde	5	40
	IRI	12	84
	Zero	61	50

information about the vertical profile is very poor and the reconstruction relies on the prior information. Hence the lower layer remains.

4 Discussion

The presented method for ionospheric tomography includes several prior parameters, and the selection of the corresponding values might seem arbitrary. The objective of this article is not to optimize all of the prior parameters, but to concentrate on the altitude profiles of the prior mean and the standard deviation. Based on trials with the algorithm and different data, the information on the vertical structures has the most crucial effect on the reconstruction quality. This is also evident in the presented results. When zero-mean prior is used, the peak altitude can be found relatively well, but the measurements do not contain enough information to produce steep enough vertical gradients. Then again, when a vertical profile is given within the prior, the reconstruction of peak electron density is improved significantly, but the peak altitude becomes less sensitive to measurements.

In horizontal direction, the gradients can be reconstructed rather well regardless of the prior mean in use. Hence, information on horizontal electron density structures (IRI model) is less important if the trade-off is the accuracy on the vertical structure.

When accurate vertical electron density profile is provided within the prior, the selection for the values of the other prior parameters is less critical. For all prior parameters the stabilizing effect is also rather intuitive. Decreasing the correlation lengths allows more small-scale variation in the reconstructions; however, getting close to the corresponding discretization can result in artifacts. The increment of correlation lengths smoothens the reconstruction, but very long correlation lengths can again produce unexpected behavior.

With all cases in the previous section, the use of horizontal correlation length values between 1 and 10° and vertical correlation lengths between 20 and 500 km were carried

out without drastically unrealistic changes in reconstructions. The peak value of standard deviation was also altered in a range from 20 to 100% with anticipated results.

As mentioned in Sect. 3, the standard deviation profile is parameterized as a Chapman function. Hence, the ionosonde profile cannot be used explicitly, but the choice of the parameter values can be done viably based on the ionosonde measurements. For the first three overflights only the peak standard deviation altitude and density were set according to the corresponding ionosonde measurements. With Overflight IV, the ionosonde profiles are significantly different; thus also the scale heights of the prior standard deviation were changed. Altogether, the results for the overflights II, III and IV could be enhanced by optimizing the parameters through trial and error individually for each case, but the results show that already intuitive realistic choices of these parameters are enough to give reasonable solutions.

As the ionosonde measurements provide relatively accurate measurements of the ionospheric electron density, it would be straightforward to use them also as direct measurements above the instrument location. However, the satellite overflight hits rarely at the zenith of the ionosonde site, and the electron densities measured by ionosonde and tomographic receiver can vary largely. When 2-D assumption (i.e., small gradients in longitude) is used, the ionosonde measurement error should reflect this discrepancy. Hence the information for the projected ionosonde measurement points cannot be modeled as accurately as they are in their actual location, and the prior distribution provides substantially the same information. In the 3-D case the situation will be different as all of the measurements will be modeled in their actual locations.

Electron density profiles measured with the EISCAT UHF are routinely calibrated by means of comparing F-region peak electron density estimates from the UHF and the dynasonde. Thus, when the ionosonde-based prior is used, F-region peak densities above the Tromsø site are taken from the same instrument in both the tomography prior and the UHF results. Our tomography measurements and the ground truth UHF measurements are thus not completely independent. However, we anticipate that this is not a very serious problem, as the calibration data were not used for the validation. Furthermore, calibration does not affect the UHF density profile shape, but only its absolute values, and calibration is not performed for individual profiles, but the same scaling is used for a longer period of time. Especially, the actual validation measurements with beam steered far away from zenith are never used for calibration.

5 Conclusions

In this study the use of Bayesian statistical inversion with known prior distributions and with the inclusion of simultaneous ionosonde measurements for ionospheric tomogra-

phy is validated. Most importantly we show that the prior distribution can be constructed based on the ionosonde measurements, which helps in constraining the otherwise poorly defined altitude profile shape of the tomographic reconstruction.

We demonstrate the applicability of the method with four satellite overflights measured with the TomoScand receiver network, and with EISCAT dynasonde measurements from the EISCAT Tromsø site. In comparisons we used International Reference Ionosphere 2007 and zero mean in building of the prior. The validation is made against simultaneous EISCAT UHF incoherent scatter radar measurements.

The biggest issue with IRI-2007 consists in the problems with the peak altitude. With zero mean it is the significant underestimation of the electron density. From both of the reference schemes it can be seen that the measurements cannot provide enough information on the vertical gradients of the ionosphere. The use of ionosonde in the building of the prior distribution outperforms the compared alternatives. The results show better agreement between the incoherent scatter radar measurements and the corresponding electron density profiles taken from the reconstruction. The reconstructions seem reasonable even further away from the ionosonde location. However, the electron density height profiles are dictated by the prior model and could be biased further away from the ionosonde. Use of multiple ionosondes and altering the prior profile in horizontal direction would be straightforward within the method and highly recommended.

The results also indicate that when reliable prior information is provided, the required prior parameters can be pre-determined and the method used without additional tuning. This makes the operational stand-alone use feasible, at least for typical ionospheric conditions. With the lattice sizes in the reported scale and with a modern PC the required computations can be made in real time.

As in the Bayesian inference we are presenting the information as probability distributions, we also have direct access to the credible intervals. If the prior is truly realistic, the posteriori credible interval can be highly informative. However, it is important to note that when interpreting the posterior distribution and credible intervals derived from it, they are highly dependent on the given prior distribution. Posterior credible intervals should thus be used with caution.

Data availability

The data for analyzed EISCAT dynasonde results from Tromsø are available from the EISCAT Dynasonde Database (http://dynserv.eiscat.uit.no/DD/Iono_form.php). The IRI-2007 electron density profiles are available from the IRI-2007 website (http://omniweb.gsfc.nasa.gov/vitmo/iri_vitmo.html).

Ionospheric tomography measurements and analyzed data products used in this paper are freely available upon request from the Finnish Meteorological Institute.

Acknowledgements. The work of J. Norberg has been funded by Academy of Finland (decision no. 287679) and European Regional Development Fund (Regional Council of Lapland, decision no. A70179). The work of I. I. Virtanen has been funded by Academy of Finland (decision no. 285474). The work of L. Roininen, M. Orispää and M. Lehtinen has been funded by European Research Council (ERC Advanced Grant 267700 – InvProb) and Academy of Finland (Finnish Centre of Excellence in Inverse Problems Research 2012–2017, decision no. 250215).

Authors thank EISCAT staff, especially Jussi Markkanen, for kindly assisting in the EISCAT UHF radar experiments, and Yoshimasa Tanaka and Yasunobu Ogawa of NIPR for executing the EISCAT UHF radar experiments in March 2015. We also would like to thank EISCAT for providing the dynasonde data. EISCAT is an international association supported by research organizations in China (CRIRP), Finland (SA), Japan (NIPR and STEL), Norway (NFR), Sweden (VR) and the UK (NERC).

Edited by: M. Nicolls

References

- Andreeva, E. S.: Radio tomographic reconstruction of ionization dip in the plasma near the Earth, *J. Exp. Theor. Phys. Lett.*, 52, 142–148, 1990.
- Austen, J. R., Franke, S. J., and Liu, C. H.: Ionospheric imaging using computerized tomography, *Radio Sci.*, 3, 299–307, 1988.
- Bilitza, D. and Reinisch, B. W.: International Reference Ionosphere 2007: improvements and a new parameters, *Adv. Space Res.*, 42, 599–609, 2008.
- Breit, G. and Tuve, M. A.: A Test of the Existence of the Conducting Layer, *Phys. Rev.*, 28, 554–575, 1926.
- Bust, G. S. and Mitchell, C. N.: History, current state, and future directions of ionospheric imaging, *Rev. Geophys.*, 46, RG1003, doi:10.1029/2006RG000212, 2008.
- Chartier, A. T., Smith, N. D., Mitchell, C. N., Jackson, D. R., and Patilongo, P. J. C.: The use of ionosondes in GPS ionospheric tomography at low latitudes, *J. Geophys. Res.*, 117, A10326, doi:10.1029/2012JA018054, 2012.
- Chiang, K. Q. Z., and Psiaki, M. L.: GPS and Ionosonde Data Fusion for Ionospheric Tomography, *Proc. ION GNSS+ 2014*, 9–12 September 2014, Tampa, FL, USA, 1163–1172, 2014.
- Davies, K.: *Ionospheric Radio* IEE Electromagn. Waves ser., IET, London, UK, 1990.
- EISCAT Dynasonde Database: available at: http://dynserv.eiscat.uit.no/DD/Iono_form.php, last access: 27 April 2016.
- IRI-2007: International Reference Ionosphere – IRI-2007, available at: http://omniweb.gsfc.nasa.gov/vitmo/iri_vitmo.html, last access: 27 April 2016.
- Kaipio, J. and Somersalo, E.: *Statistical and Computational Inverse Problems*, Applied Mathematical Sciences, Springer, New York, USA, 2005.
- Kersley, L., Heaton, J. A. T., Pryse, S. E., and Raymund, T. D.: Experimental ionospheric tomography with ionosonde input and EISCAT verification, *Ann. Geophys.*, 11, 1064–1074, 1993.
- Markkanen, M., Lehtinen, M., Nygren, T., Pirttilä, J., Helenius, P., Vilenius, E., Tereshchenko, E. D., and Khudukon, B. Z.: Bayesian approach to satellite radiotomography with applications in the Scandinavian sector, *Ann. Geophys.*, 13, 1277–1287, 1995.
- Norberg, J., Roininen, L., Vierinen, J., Amm, O., McKay-Bukowski, D., and Lehtinen, M. S.: Ionospheric tomography in Bayesian framework with Gaussian Markov random field priors, *Radio Sci.*, 50, 138–152, doi:10.1002/2014RS005431, 2015.
- Vierinen, J., Norberg, J., Lehtinen, M. S., Amm, O., Roininen, L., Väänänen, A., and McKay-Bukowski, D.: Software defined beacon satellite receiver software for ionospheric tomography, *Radio Sci.*, 49, 1141–1152, doi:10.1002/2014RS005434, 2014.
- Zabotin, N. A., Wright, J. W., and Zhabankov, G. A.: NeXtYZ: three-dimensional electron density inversion for dynasonde ionograms, *Radio Sci.*, 41, RS6S32, doi:10.1029/2005RS003352, 2006.

Publication IV

J. Norberg, J. Vierinen, L. Roininen, M. Orispää, K. Kauristie, W. C. Rideout, A. J. Coster, and M. S. Lehtinen, Gaussian Markov Random Field Priors in Ionospheric 3-D Multi-Instrument Tomography, *IEEE Trans. Geosci. Remote Sens.*, 1–13, 2018. ISSN 1558-0644. doi: 10.1109/TGRS.2018.2847026.

Gaussian Markov Random Field Priors in Ionospheric 3-D Multi-Instrument Tomography

Johannes Norberg¹, Juha Vierinen, Lassi Roininen, Mikko Orispää, Kirsti Kauristie, William C. Rideout, Anthea J. Coster, and Markku S. Lehtinen

Abstract—In ionospheric tomography, the atmospheric electron density is reconstructed from different electron density related measurements, most often from ground-based measurements of satellite signals. Typically, ionospheric tomography suffers from two major complications. First, the information provided by measurements is insufficient and additional information is required to obtain a unique solution. Second, with necessary spatial and temporal resolutions, the problem becomes very high dimensional, and hence, computationally infeasible. With Bayesian framework, the required additional information can be given with prior probability distributions. The approach then provides physically quantifiable probabilistic interpretation for all model variables. Here, Gaussian Markov random fields (GMRFs) are used for constructing the prior electron density distribution. The use of GMRF introduces sparsity to the linear system, making the problem computationally feasible. The method is demonstrated over Fennoscandia with measurements from global navigation satellite system (GNSS) and low Earth orbit (LEO) satellite receiver networks, GNSS occultation receivers, LEO satellite Langmuir probes, and ionosonde and incoherent scatter radar measurements.

Index Terms—Bayesian, Gaussian Markov random fields (GMRFs), ionospheric tomography, multi-instrument.

I. INTRODUCTION

IONOSPHERIC tomography involves reconstruction of the atmospheric electron density within a volume, using a number of different measurements of electron density. The first studies on ionospheric tomography were made in [1]. A general introduction to the topic is provided in [2].

The electron density measurements can be divided to *indirect* and *direct* measurements. Indirect measurements refer

Manuscript received November 20, 2017; revised April 20, 2018; accepted May 23, 2018. The work of J. Norberg was supported in part by the Academy of Finland under Grant 287679 and in part by the Regional Council of Lapland through European Regional Development Fund under Grant A70179. The work of L. Roininen was supported by the Academy of Finland under Project 307741 and Project 313709. The work of K. Kauristie was supported by the Academy of Finland under Grant 287679. (Corresponding author: Johannes Norberg.)

J. Norberg is with the Finnish Meteorological Institute, 00560 Helsinki, Finland, and also with Sodankylä Geophysical Observatory, University of Oulu, 99600 Sodankylä, Finland (e-mail: johannes.norberg@fmi.fi).

J. Vierinen is with the Department of Physics and Technology, University of Tromsø, 9019 Tromsø, Norway.

L. Roininen, M. Orispää, and M. S. Lehtinen are with Sodankylä Geophysical Observatory, University of Oulu, 90014 Oulu, Finland.

K. Kauristie is with the Finnish Meteorological Institute, 00560 Helsinki, Finland.

W. C. Rideout and A. J. Coster are with Haystack Observatory, Massachusetts Institute of Technology, Cambridge, MA 02139 USA.

Digital Object Identifier 10.1109/TGRS.2018.2847026

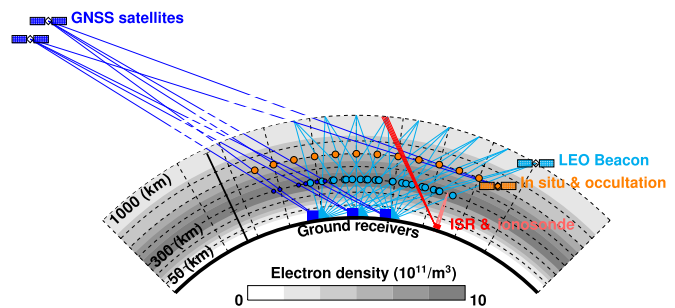


Fig. 1. Multi-instrument ionospheric electron density measurements. GNSS satellites provide line-integrated measurements all the way up to the plasmasphere, but as the satellites move relatively slow, the measurement geometry with respect to ground stations does not change significantly. The inclination angles of different GNSSs are low. LEO beacon satellites, which cover the ionosphere up to about 1000 km, can have a polar orbit and move relatively fast, allowing a 2-D slice of the ionosphere to be covered. LEO satellites with GNSS receivers provide variable look angles through the topside of the ionosphere. Langmuir probe provides the only *in situ* measurements available. Ionosondes provide localized bottomside profiles. ISRs provide localized profiles of electron density.

here to satellite-to-ground and satellite-to-satellite measurements, where the measurements are modeled as integrals over electron density. In this paper, ground-based total electron content (TEC) measurements from global navigation satellite system (GNSS) and low Earth orbit (LEO) satellites and GNSS occultation TEC measurements are employed. These measurements have a lower accuracy of structural information, but typically a large spatial coverage can be attained. Direct measurements, such as different radar and satellite *in situ* measurements, provide relatively accurate structural information, but from a restricted area. This paper utilizes a satellite on-board Langmuir probe, incoherent scatter radars (ISRs), and ionosondes as direct measurements. However, even within these two categories the different measurements have different strengths and weaknesses, and are highly complementary. A 2-D simplification of measurement geometries is shown in Fig. 1.

Typically, the available measurements are predominantly indirect satellite measurements with limited elevation angles, and the resulting information, particularly on vertical gradients is low. Consequently, the electron density cannot be solved uniquely without some additional structural information. The essential difference between different ionospheric tomography methods follows from how this information is implemented into the algorithm.

In [3], an extensive overview on different approaches and their development is provided. More recently, the topic has been studied in [4]–[7].

With Bayesian statistical approach, the required additional information can be given as a prior probability distribution. When Gaussian distributions are assumed, the prior information can be represented with the mean and covariance. In ionospheric tomography, the prior mean corresponds to most probable state of the atmospheric electron density, whereas the covariance describes its uncertainty. All the parameters included in the prior have, thus, a clear probabilistic interpretation with physical quantities. Hence, when the measurement and model errors are modeled correctly, in principle, the model has no free parameters that need to be calibrated. If additional tuning is nevertheless required, all the changes in the parameters have a clear meaning.

A comprehensive introduction to statistical inverse problems is provided in [8] and [9]. In ionospheric tomography, the Bayesian inference has been applied in [10]–[13]. The Ionospheric Data Assimilation Three-Dimensional (IDA3D) presented in [14] is based on the Three-Dimensional Variational Data Assimilation Technique (3DVAR) and uses slightly different terminology. However, with Gaussian prior and error distributions, the Bayesian statistical approach and 3DVAR are computationally identical.

The drawback of Bayesian approach is in computation, as the solution requires operations of large covariance matrices. In this paper, Gaussian Markov random field (GMRF) [15] priors for ionospheric tomography, presented in [12] and [13], are generalized to the 3-D multi-instrument case. The GMRF approach provides an extension to Bayesian and 3DVAR methods, as the prior covariance can be replaced with a sparse matrix approximation. Essentially, by implementing GMRF with sparse systems solvers, the computational cost in the high-dimensional matrix operations is reduced significantly.

This paper is organized as follows. In Section II, the models for different measurements are described. In Section III, the Bayesian statistical approach for linear tomography is revisited. In Section IV, the GMRF approximation of prior precision matrix is described, with a short summary of alternative ways to overcome the computational issues. In Section V, the performance of the considered method is presented. First, with a simulated example using the International Reference Ionosphere 2012 (IRI-2012) model as the unknown, and then with real multi-instrument data. The EISCAT ISR measurements are included for validation. Discussion and conclusion of the study are given in Sections VI and VII. The future plans are considered in Section VIII.

II. MEASUREMENTS

The approach presented in this paper can exploit any measurement depending linearly on the ionospheric electron density with estimable measurement error. The electron density is given here as function $N_e(t, z)$, where t is the time and $z = \{z_{\text{lat}}, z_{\text{long}}, z_{\text{alt}}\} \in \mathbf{R}^3$ is the spatial coordinates.

A. Indirect Measurements

A GNSS TEC measurement along signal path $L_{\text{sat},\text{rec}}(t)$ between satellite sat and receiver rec at time t can be modeled as a line integral

$$m_{\text{sat},\text{rec}}(t) = \int_{L_{\text{sat},\text{rec}}(t)} N_e(t, z) dl + a_{\text{sat}}(t) + b_{\text{rec}}(t) + \varepsilon_{\text{sat},\text{rec}}(t) \quad (1)$$

where $a_{\text{sat}}(t)$ and $b_{\text{rec}}(t)$ are the receiver and satellite instrument biases. Discretization and assumption of time homogeneous ionosphere cause errors that need to be taken into account in the model. Here, the modeling errors are assumed independent of the unknown electron densities and added all together in $\varepsilon_{\text{sat},\text{rec}}(t)$.

The forward model for GNSS occultation measurements is the same as for the ground-based measurements given in (1). The practical difference is that the on-board receiver is also in motion.

The relative TEC (RTEC) measurement between a LEO beacon satellite and a ground-based receiver can be modeled as

$$m_{\text{sat},\text{rec}}(t) = \int_{L_{\text{sat},\text{rec}}(t)} N_e(t, z) dl + \gamma_{\text{sat},\text{rec}} + \varepsilon_{\text{sat},\text{rec}}(t) \quad (2)$$

where the phase ambiguity is given with $\gamma_{\text{sat},\text{rec}}$ and it is different for each signal lock [16].

Despite the similarity of the TEC measurement models, there are significant practical differences. By combining the different observables, the GNSS TEC does not suffer from the phase ambiguity [17]. The remaining GNSS satellite and receiver biases can have a significant effect on the TEC measurement. However, the changes in these biases are relatively slow and they can be estimated before the tomographic analysis [18], [19], thus the bias-corrected GNSS measurements can be assumed to be close to absolute TEC. On the other hand, as the bias estimation requires several simplifying assumptions, it is beneficial to keep the parameters a_{sat} and b_{rec} in the model and assume that after the bias correction, they are closer to zero, but not completely eliminated.

Due to lower signal frequencies (150 and 400 MHz), the LEO TEC measurements are more sensitive for detailed ionospheric structures, however, as the phase ambiguity is typically in the magnitude of current local maximum TEC, the measurement is relative and $\gamma_{\text{sat},\text{rec}}$ needs to be solved as an additional unknown in the tomographic analysis. The orbital inclinations of GNSS satellites are relatively low [GPS 55°, Global Navigation Satellite System (GLONASS) 64.13°, GALILEO 56°]. For the polar regions, the low inclination induces low elevation angles making the measurements susceptible for larger errors. LEO beacon satellites can have orbital inclinations close to polar orbits (CASSIOPE/e-POP 80.99°, COSMOS 2407 and 2463 83°), thus providing high-elevation measurements also from high-latitude receivers. Following from the higher orbital altitudes of GNSS satellites (GLONASS 19140 km, GPS 20180 km, and GALILEO 23222 km), the plasmaspheric contribution in the TEC measurement can be significant. With orbital altitudes around 1000 km, there is no plasmaspheric contribution in the LEO

RTEC measurements. The lower altitude also results in a higher relative satellite velocity, and hence better spatial coverage for the individual satellite when present. Multiple GNSS satellites can be viewed from all locations and at all times, especially now there are several satellite constellations. GNSS observations are also easily obtained as large receiver networks exist. This differs from the case of LEO satellites as there are fewer receivers (LEO receivers are built specifically for this purpose) and there are only few suitable satellites. Hence, in general, the GNSS TEC measurements are the most important data component, providing constant spatiotemporal coverage with absolute measurements. LEO RTEC measurements provide more detailed information and provide higher elevation data in polar areas, but the spatiotemporal availability is inferior to GNSS.

B. Direct Measurements

When the electron density is detected in a specific location, as is the case with different radar and satellite *in situ* measurements, the direct measurement from instrument \mathcal{I} can be modeled simply as

$$m_{\mathcal{I}}(t, z) = N_e(t, z) + \varepsilon_{\mathcal{I}}(t, z). \quad (3)$$

With satellite *in situ* measurements $z = (z_{\text{lat}}, z_{\text{lon}}, z_{\text{alt}})$ is the location of the Langmuir probe, and with ISR, it is the location of measurement integration. With inverted, real-height ionosonde electron density profiles $(z_{\text{lat}}, z_{\text{long}})$ is the location of the instrument and z_{alt} is the real reflection height.

In comparison to any indirect measurement, these measurements provide significantly more accurate and detailed information on ionospheric structures. As a downside, the spatial coverage of measurements is typically limited.

III. BAYESIAN STATISTICAL APPROACH

By assuming stationary electron density for a given time interval and discretizing the measurement models (1)–(3), all the ionospheric measurements can be combined and written as

$$\mathbf{m} = \mathbf{A}\mathbf{X} + \boldsymbol{\varepsilon} \quad (4)$$

where $\mathbf{m} \in \mathbf{R}^M$ is a vector of all measurements. Geometry matrix $\mathbf{A} \in \mathbf{R}^{M \times N}$ is a linear mapping from discretized spatial domain to measurement space. Vector

$$\mathbf{X} = \begin{pmatrix} \mathbf{x} \\ \boldsymbol{\theta} \end{pmatrix} \in \mathbf{R}^N$$

consists of unknown electron densities $\mathbf{x} \in \mathbf{R}^n$, as well as all the additional unknown error parameters $\boldsymbol{\theta} = (a_{\text{sat}_1} \dots a_{\text{sat}_A}, b_{\text{rec}_1}, \dots, b_{\text{rec}_B}, \gamma_{1,1} \dots \gamma_{C,D}, \rho)^T$, where ρ is the plasmaspheric electron density contribution per meter, A , B , C , and D are the counts of GNSS satellites, GNSS receivers, LEO satellites, and LEO receivers, respectively, and $n + A + B + C \times D + 1 = N$. Vector $\boldsymbol{\varepsilon} \in \mathbf{R}^M$ contains all the measurement and modeling errors. Here, it is assumed that the error vector follows a multivariate normal distribution

$$\boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\varepsilon})$$

and that the realistic states of the ionosphere and additional parameters can be described satisfactorily with a prior distribution, which is also multivariate normal

$$\mathbf{X} \sim \mathcal{N}(\mathbf{X}_{\text{pr}}, \boldsymbol{\Sigma}_{\text{pr}}) \quad (5)$$

where \mathbf{X}_{pr} is the prior mean of the ionosphere and additional parameters. Covariance $\boldsymbol{\Sigma}_{\text{pr}}$ describes the related prior uncertainties. Following the Bayes theorem [9], the posterior distribution for \mathbf{X} is then also a multivariate normal distribution

$$\mathbf{X}|\mathbf{m} \sim \mathcal{N}(\mathbf{X}_{\text{MAP}}, \boldsymbol{\Sigma}_{\text{post}})$$

where *maximum a posteriori* (MAP) estimator

$$\mathbf{X}_{\text{MAP}} = \boldsymbol{\Sigma}_{\text{post}}(\mathbf{A}^T \boldsymbol{\Sigma}_{\varepsilon}^{-1} \mathbf{m} + \boldsymbol{\Sigma}_{\text{pr}}^{-1} \mathbf{X}_{\text{pr}}) \quad (6)$$

and posterior covariance estimator

$$\boldsymbol{\Sigma}_{\text{post}} = (\mathbf{A}^T \boldsymbol{\Sigma}_{\varepsilon}^{-1} \mathbf{A} + \boldsymbol{\Sigma}_{\text{pr}}^{-1})^{-1}. \quad (7)$$

In an application to ionospheric tomography, the MAP estimator \mathbf{X}_{MAP} can be understood as the most probable state of the ionospheric electron density and other unknown parameters, whereas the remaining uncertainty is given with the error covariance $\boldsymbol{\Sigma}_{\text{post}}$.

IV. COMPUTATION

The downside of Bayesian statistical approach for inverse problems is that the assumption of a proper prior distribution (5) results with a dense $N \times N$ covariance matrix $\boldsymbol{\Sigma}_{\text{pr}}$. The estimators (6) and (7) contain inverted covariance matrices and the posterior covariance estimator involves one more matrix inversion. Hence, the solution becomes exceedingly demanding computationally. To ease the computational burden, the dimensions would need to be reduced or sparsity would need to be introduced into the linear system. Two previously applied approaches are first revisited, before introducing the GMRF approach for sparse approximation of the covariance matrix.

A. Generalized Tikhonov Regularization

When independent measurement errors are assumed, the measurement covariance matrix is diagonal, $\boldsymbol{\Sigma}_{\varepsilon} = \sigma_{\varepsilon}^2 \mathbf{I}$, and due to the nature of the measurements, the geometry matrix \mathbf{A} is a sparse matrix. If the inverted covariance matrix, i.e., the prior precision matrix $\boldsymbol{\Sigma}_{\text{pr}}^{-1}$ could be given also as a sparse matrix, the memory required for storage would decrease and optimized solvers for sparse linear systems could be utilized for the computation.

In more general sense, the MAP estimator (6) is a regularized least squares solution. Selecting $\boldsymbol{\Sigma}_{\text{pr}}^{-1} = \mathbf{0}$ reduces it to ordinary least squares estimator. With $\boldsymbol{\Sigma}_{\text{pr}}^{-1} = \delta \mathbf{I}$, the estimator corresponds to Ridge regression and Tikhonov regularized solution. When $\boldsymbol{\Sigma}_{\text{pr}}^{-1}$ is any suitably chosen matrix, the estimator is called the generalized Tikhonov regularization [9]. Typically, $\boldsymbol{\Sigma}_{\text{pr}}^{-1} = \delta \boldsymbol{\Gamma}^T \boldsymbol{\Gamma}$, where $\boldsymbol{\Gamma}$ is a difference matrix. These can be extremely sparse systems and the inversion in (7) is possible even for relatively high-dimensional problems. With these approaches, it is possible to add structural constraints for

the solution. For ionospheric tomography, in [4] and [10], a vertical weight profile is included in matrices $\mathbf{\Gamma}$ to regularize the problem more strictly at the very low and high altitudes and to allow more variability to electron densities at altitudes where the highest values and dynamics are expected. The constraint ends up to provide essentially similar information than a full prior covariance would, but as it lacks some of the interpretation, it requires *ad hoc* calibration.

B. Data Space Solution

With matrix inversion lemma, the MAP estimator (6) can be written equivalently as

$$\mathbf{X}_{\text{MAP}} = \mathbf{X}_{\text{pr}} + \mathbf{\Sigma}_{\text{pr}} \mathbf{A}^T (\mathbf{A} \mathbf{\Sigma}_{\text{pr}} \mathbf{A}^T + \mathbf{\Sigma}_{\varepsilon})^{-1} (\mathbf{m} - \mathbf{A} \mathbf{X}_{\text{pr}}) \quad (8)$$

and

$$\mathbf{\Sigma}_{\text{post}} = \mathbf{\Sigma}_{\text{pr}} - \mathbf{\Sigma}_{\text{pr}} \mathbf{A}^T (\mathbf{A} \mathbf{\Sigma}_{\text{pr}} \mathbf{A}^T + \mathbf{\Sigma}_{\varepsilon})^{-1} \mathbf{A} \mathbf{\Sigma}_{\text{pr}}. \quad (9)$$

This form is preferred by IDA3D [14] and in the standard formulation of Kalman filter. The advantage is that the inversion is done for $M \times M$ matrix, whereas in (6), an $N \times N$ matrix is inverted. Typically, in ionospheric tomography, $M \ll N$ which can unburden the inversion. Here, the computational bottleneck is in representing and storing the covariance matrix $\mathbf{\Sigma}_{\text{pr}}$ and performing the matrix multiplications. In case of global ionospheric tomography, it can be assumed that locations with long enough distance have a zero correlation. This covariance tapering [20] introduces sparsity to the system. However, in local tomography, the correlation lengths extend over the domain, hence, the prior covariance is a dense matrix.

C. Gaussian Markov Random Field Prior

The idea here is the following. As the estimator (6) does not require actual covariance matrices, but rather precision matrices, a sparse precision matrix \mathbf{Q} is constructed so that

$$\mathbf{Q}^{-1} \approx \mathbf{\Sigma}_{\text{pr}} \quad (10)$$

where the matrix \mathbf{Q} is built with the formalism of GMRFs. When parameterized correctly, a GMRF gives rise to a known spatial covariance. So, even though the actual covariance matrix is never constructed, the precision matrix can be determined by its covariance properties.

The detailed theoretical background for these types of GMRFs is given in [21]–[24]. Here, an example is given, where the target precision $\mathbf{\Sigma}_{\text{pr}}^{-1}$ is determined with a squared exponential covariance function

$$\begin{aligned} \text{Cov}(z_i - z_j) = \alpha(z_i, z_j) \exp \left[-\frac{1}{2} \left(\left(\frac{z_{i,\text{lat}} - z_{j,\text{lat}}}{\ell_{\text{lat}}} \right)^2 \right. \right. \\ \left. \left. + \left(\frac{z_{i,\text{long}} - z_{j,\text{long}}}{\ell_{\text{long}}} \right)^2 + \left(\frac{z_{i,\text{alt}} - z_{j,\text{alt}}}{\ell_{\text{alt}}} \right)^2 \right) \right] \end{aligned} \quad (11)$$

where $\ell_{\text{lat}}, \ell_{\text{long}},$ and $\ell_{\text{alt}} > 0$ are the correlation lengths homogeneous in each coordinate direction and $\alpha(z_i, z_j) > 0$ is a location-dependent variance mask.

Following [24], an anisotropic continuous GMRF with a covariance approximating (11) can be formed as a solution to a stochastic partial differential equation

$$\begin{pmatrix} \sqrt{c_0} I \\ \sqrt{c_1} \ell_{\text{lat}} \partial_{\text{lat}} \\ \sqrt{c_1} \ell_{\text{alt}} \partial_{\text{alt}} \\ \sqrt{c_1} \ell_{\text{long}} \partial_{\text{long}} \\ \sqrt{c_2} (\ell_{\text{lat}}^2 \partial_{\text{lat}}^2 + \ell_{\text{alt}}^2 \partial_{\text{alt}}^2 + \ell_{\text{long}}^2 \partial_{\text{long}}^2) \end{pmatrix} X(z) = \sqrt{\alpha(z) \ell_{\text{lat}} \ell_{\text{alt}} \ell_{\text{long}}} \begin{pmatrix} \mathcal{W}^{(0)}(z) \\ \mathcal{W}^{(1,\text{lat})}(z) \\ \mathcal{W}^{(1,\text{alt})}(z) \\ \mathcal{W}^{(1,\text{long})}(z) \\ \mathcal{W}^{(2)}(z) \end{pmatrix} \quad (12)$$

where $X(z)$ and $\mathcal{W}^{(\cdot)}(z)$, $z \in \mathbf{R}^3$ are the continuous random fields and independent continuous white noise fields, respectively. For squared exponential covariance function (11), the shape parameters $c_k = 2^{-k}/k!$, with k corresponding to the order of the derivative.

To make (12) applicable for computations, there are two objects to discretize the differential operator on the left-hand side and the white noise on the right-hand side. By using finite difference methods, a 3-D domain is defined with dimensions $n_{\text{lat}} \times n_{\text{alt}} \times n_{\text{long}} = n$ and voxel widths $h_{\text{lat}}, h_{\text{alt}}$ and h_{long} in the different coordinate directions. To simplify the following formulation, new parameters $s_{\text{lat}} := (h_{\text{lat}}/\ell_{\text{lat}})$, $s_{\text{alt}} := (h_{\text{alt}}/\ell_{\text{alt}})$ and $s_{\text{long}} := (h_{\text{long}}/\ell_{\text{long}})$ are defined. Here, only the unknown electron densities $\mathbf{x} = (x_1, \dots, x_j, \dots, x_n)^T$ are considered. The additional error parameters $\boldsymbol{\theta}$ are independent of \mathbf{x} and are included after the field is discretized at the end of the section.

The discrete approximation for white noise is

$$W_j^{(k)} \sim \mathcal{N} \left(0, \frac{\alpha_j}{c_k s_{\text{lat}} s_{\text{alt}} s_{\text{long}}} \right)$$

where the superscript k is an index for the order of difference. The discrete white noise process is then set as

$$x_j = W_j^{(0)}. \quad (13)$$

The first-order differences are

$$\begin{aligned} \Delta_{\text{lat}} x_j &:= \frac{1}{s_{\text{lat}}} (x_{j+1} - x_j) = W_j^{(1,\text{lat})} \\ \Delta_{\text{alt}} x_j &:= \frac{1}{s_{\text{alt}}} (x_{j+n_{\text{lat}}} - x_j) = W_j^{(1,\text{alt})} \\ \Delta_{\text{long}} x_j &:= \frac{1}{s_{\text{long}}} (x_{j+n_{\text{lat}} n_{\text{alt}}} - x_j) = W_j^{(1,\text{long})} \end{aligned} \quad (14)$$

where the first superscript on the right-hand side is the order of difference and the second superscript is the effective coordinate direction. The second-order differences are given as

$$\Delta^2 x_j := \Delta_{\text{lat}}^2 x_j + \Delta_{\text{alt}}^2 x_j + \Delta_{\text{long}}^2 x_j = W_j^{(2)} \quad (15)$$

where

$$\begin{aligned}\Delta_{\text{lat}}^2 x_j &:= \left(\frac{x_{j+1} - x_j}{s_{\text{lat}}} - \frac{x_j - x_{j-1}}{s_{\text{lat}}} \right) \frac{1}{s_{\text{lat}}} \\ \Delta_{\text{alt}}^2 x_j &:= \left(\frac{x_{j+n_{\text{lat}}} - x_j}{s_{\text{alt}}} - \frac{x_j - x_{j-n_{\text{lat}}}}{s_{\text{alt}}} \right) \frac{1}{s_{\text{alt}}} \\ \Delta_{\text{long}}^2 x_j &:= \left(\frac{x_{j+n_{\text{lat}}n_{\text{alt}}} - x_j}{s_{\text{long}}} - \frac{x_j - x_{j-n_{\text{lat}}n_{\text{alt}}}}{s_{\text{long}}} \right) \frac{1}{s_{\text{long}}}.\end{aligned}$$

The white noise $W_j^{(k)}$ can be standardized as

$$\sqrt{\frac{c_k}{\alpha_j s_{\text{lat}} s_{\text{alt}} s_{\text{long}}}} W_j^{(k)} = \zeta_j^{(k)} \sim \mathcal{N}(0, 1)$$

and for all indices j and k as

$$\begin{aligned}\xi_0 &= (\zeta_1^{(0)}, \dots, \zeta_n^{(0)}) \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_n), \\ \xi_1 &= (\zeta_1^{(1, \text{lat})}, \dots, \zeta_n^{(1, \text{long})}) \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{3n}), \\ \xi_2 &= (\zeta_1^{(2)}, \dots, \zeta_n^{(2)}) \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_n).\end{aligned}$$

Now the differences (13)–(15) can be written in matrix form as follows.

For the standardized white noise process, the matrix form is

$$\sqrt{\frac{c_0}{\alpha}} s_{\text{lat}} s_{\text{alt}} s_{\text{long}} \circ [\mathbf{I}_{n_{\text{long}}} \otimes \mathbf{I}_{n_{\text{alt}}} \otimes \mathbf{I}_{n_{\text{lat}}}] \mathbf{x} =: \mathbf{L}_0 \mathbf{x} = \xi_0 \quad (16)$$

where \mathbf{I}_s are identity matrices with diagonal length corresponding to the subscript, the variance mask is given as

$$\frac{1}{\alpha} := \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \otimes \left[\frac{1}{\alpha_1}, \dots, \frac{1}{\alpha_n} \right] \in \mathbf{R}^{n \times n}$$

the symbol “ \circ ” is the Hadamard product and the symbol “ \otimes ” is the Kronecker product.

For the first-order differences, the different coordinate directions are given in matrix form separately at their own rows

$$\begin{pmatrix} \sqrt{\frac{c_1}{\alpha}} \frac{s_{\text{alt}} s_{\text{long}}}{s_{\text{lat}}} \circ [\mathbf{I}_{n_{\text{long}}} \otimes \mathbf{I}_{n_{\text{alt}}} \otimes \mathbf{L}_{n_{\text{lat}}}] \\ \sqrt{\frac{c_1}{\alpha}} \frac{s_{\text{lat}} s_{\text{long}}}{s_{\text{alt}}} \circ [\mathbf{I}_{n_{\text{long}}} \otimes \mathbf{L}_{n_{\text{alt}}} \otimes \mathbf{I}_{n_{\text{lat}}}] \\ \sqrt{\frac{c_1}{\alpha}} \frac{s_{\text{lat}} s_{\text{alt}}}{s_{\text{long}}} \circ [\mathbf{L}_{n_{\text{long}}} \otimes \mathbf{I}_{n_{\text{alt}}} \otimes \mathbf{I}_{n_{\text{lat}}}] \end{pmatrix} \mathbf{x} =: \mathbf{L}_1 \mathbf{x} = \xi_1 \quad (17)$$

where $\mathbf{L}_{n_{\text{lat}}}$, $\mathbf{L}_{n_{\text{alt}}}$, and $\mathbf{L}_{n_{\text{long}}}$ are difference matrices with structure

$$\begin{pmatrix} -1 & 1 & & \\ & -1 & 1 & \\ & & \ddots & \ddots \\ 1 & & & -1 \end{pmatrix}$$

and dimensions $n_{\text{lat}} \times n_{\text{lat}}$, $n_{\text{alt}} \times n_{\text{alt}}$, and $n_{\text{long}} \times n_{\text{long}}$, correspondingly.

The second-order differences are given in matrix form as

$$\begin{aligned}& \left(\sqrt{\frac{c_2}{\alpha}} \frac{s_{\text{alt}} s_{\text{long}}}{s_{\text{lat}}^3} \circ [\mathbf{I}_{n_{\text{long}}} \otimes \mathbf{I}_{n_{\text{alt}}} \otimes (\mathbf{L}_{n_{\text{lat}}}^T \mathbf{L}_{n_{\text{lat}}})] \right. \\ & + \sqrt{\frac{c_2}{\alpha}} \frac{s_{\text{lat}} s_{\text{long}}}{s_{\text{alt}}^3} \circ [\mathbf{I}_{n_{\text{long}}} \otimes (\mathbf{L}_{n_{\text{alt}}}^T \mathbf{L}_{n_{\text{alt}}}) \otimes \mathbf{I}_{n_{\text{lat}}}] \\ & \left. + \sqrt{\frac{c_2}{\alpha}} \frac{s_{\text{lat}} s_{\text{alt}}}{s_{\text{long}}^3} \circ [(\mathbf{L}_{n_{\text{long}}}^T \mathbf{L}_{n_{\text{long}}}) \otimes \mathbf{I}_{n_{\text{alt}}} \otimes \mathbf{I}_{n_{\text{lat}}}] \right) \mathbf{x} =: \mathbf{L}_2 \mathbf{x} = \xi_2.\end{aligned} \quad (18)$$

Finally, the matrix equations (16)–(18) can be stacked as

$$\begin{pmatrix} \mathbf{L}_0 \\ \mathbf{L}_1 \\ \mathbf{L}_2 \end{pmatrix} \mathbf{x} = \begin{pmatrix} \xi_0 \\ \xi_1 \\ \xi_2 \end{pmatrix} =: \mathbf{L} \mathbf{x} = \xi \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{5n}). \quad (19)$$

The solution for \mathbf{x} is the desired prior GMRF with an $n \times n$ precision matrix $\mathbf{L}^T \mathbf{L}$. The precision matrix can then be completed to $N \times N$ dimensions, to also take into account the additional error parameters θ as

$$\begin{pmatrix} \mathbf{L}^T \mathbf{L} & 0 \\ 0 & \text{diag}(\sigma_{\theta_{\text{pr}}}^{-2}) \end{pmatrix} = \mathbf{Q} \quad (20)$$

where a prior distribution $\theta \sim \mathcal{N}(\theta_{\text{pr}}, \text{diag}(\sigma_{\theta_{\text{pr}}}^2))$ is assumed, with prior mean $\theta_{\text{pr}} \in \mathbf{R}^{(N-n)}$ and diagonal prior covariance matrix $\text{diag}(\sigma_{\theta_{\text{pr}}}^2) \in \mathbf{R}^{(N-n) \times (N-n)}$.

Now, the matrix inversion \mathbf{Q}^{-1} is a close approximation for Σ_{pr} , with the covariance structure for \mathbf{x} given in (11). Most importantly, the matrix $\mathbf{L}^T \mathbf{L}$ in \mathbf{Q} consists only of $25 \times n$ nonzero elements, whereas the corresponding covariance would be a dense $n \times n$ matrix. Obviously, in 3-D cases, n is orders of magnitude greater than 25.

With the given discretization, the methodology provides inhomogeneous GMRF priors, which take into account the discretization of the unknown, hence providing a discretization-invariant reconstruction method (for references on discretization invariance, see [22]). From the practical point of view, this means that the posterior distributions and reconstructions on different computational meshes are essentially the same, given dense enough mesh. In this section, in (13)–(19), a regular discretization was used. However, an irregular grid could be used as well, and in the example of Section V, this is the case. The use of irregular discretization is straightforward, but, as the parameters h and s become vectors and for one index, the discretization step can be different depending on which side the difference is taken, the indexing of the previous equations would become more difficult to follow.

V. RESULTS

A time window on November 8, 2015 from 10:18:00 to 10:38:00 UTC is chosen for analysis as all the observation types described in Section II are available at that time. The corresponding magnetic local time interval at EISCAT site, Tromsø, is approximately from 12:35 to 12:55.

The analyzed period was geomagnetically quiet with the Auroral Electrojet index [25] being clearly below 100 nT. However, a major geomagnetic storm took place during the previous day and during the prior 12 h, a couple of minor (B and C-level) solar flares were apparent. During the afternoon and evening hours of November 8, auroral electrojet activity index activity enhanced again to levels above 1000 nT.

The receiver and measurement locations are presented in Figs. 2 and 3. The observations include: 1) seven in-sight GPS satellites measured with 81 Geotrim,¹ 324 SWEPOS,² and 123 International GNSS Service receivers, all first collected with 30 s and then averaged to 300 s time resolution; 2) simultaneous COSMOS 2407 and 2463 LEO beacon

¹<http://www.geotrim.fi>

²<https://swepos.lantmateriet.se>

2015.11.08. 10:18:00–10:38:00 UTC

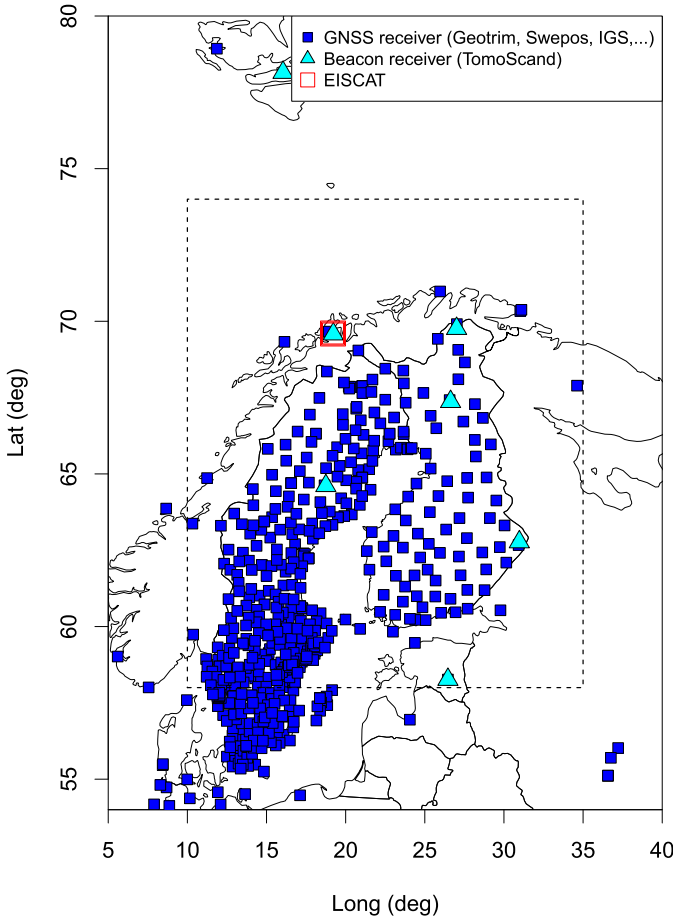


Fig. 2. Locations of available ground-based instruments. Only the area of the dashed rectangle is shown in the following 3-D reconstructions.

satellite overflights, both measured with seven ground-based TomoScand receivers [16]; 3) Swarm B satellite overflight providing Langmuir probe *in situ* measurements, as well as satellite occultation TEC measurements from one GPS satellite [26]; 4) EISCAT Dynasonde electron density profiles with 120 s time resolution³ [27]; 5) EISCAT ultra high frequency (UHF) ISR measurements with elevation = 35° and azimuth = 145°; and 6) EISCAT very high frequency (VHF) ISR measurements with elevation = 90°.⁴

Two prior mean options were used. First, a simple zero profile and then an altitude dependent ionosonde profile. With the ionosonde prior mean, the lower profile is taken as the altitude medians from 10 EISCAT Dynasonde profiles from the given time interval. Above the peak electron density altitude, an exponential profile is used with scale height of 180 km. For covariance, a squared exponential spatial function was chosen with correlation lengths: $\ell_{\text{lat}} = 20^\circ$, $\ell_{\text{long}} = 25^\circ$, and $\ell_{\text{alt}} = 400$ km. The correlation length is defined here as the distance where the covariance drops to 10% of variance. The altitude-dependent variance scaling profile α is determined here with standard deviation (SD). The SD is given as a Chapman profile with peak altitude taken from EISCAT Dynasonde

³<http://dynserv.eiscat.uit.no/DD/login.php>

⁴<https://www.eiscat.se/madrigal/>

2015.11.08. 10:18:00–10:38:00 UTC

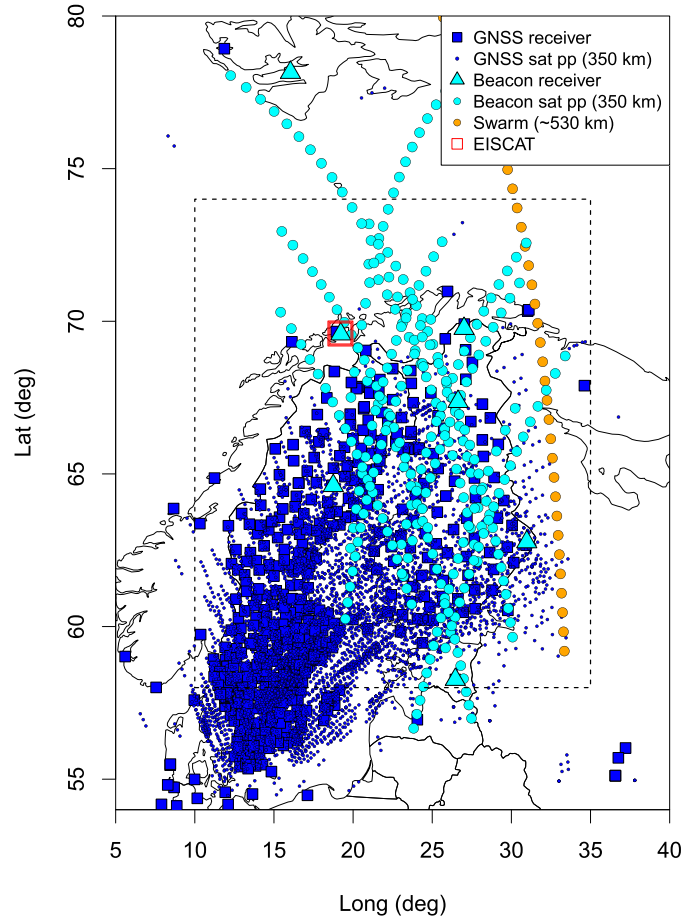


Fig. 3. Locations of available ground-based instruments, ionospheric pierce points of satellite measurements, and Swarm satellite measurements. Only the area of the dashed rectangle is shown in the following 3-D reconstructions.

real-height peak and electron density corresponds to 50% of Dynasonde's peak electron density. The scale height is set to 140 km.

Approximations for the measurement error SDs are provided with the data. SD of 2 total electron content units (TECU) is assumed for modeling errors. For preprocessed GPS station biases, zero mean with 1 TECU SD and for GPS satellite biases zero mean with 0.1 TECU SD is assumed. The phase constants of LEO measurements are given an SD of 10 TECU. The plasmaspheric contribution above 1250 km altitude is assumed to be uniformly 0.1 TECU for 2×10^4 km with SD of 0.1 TECU.

The 3-D spatial domain chosen for analysis covers latitudes from 54° to 80°, with 2° resolution at boundaries and 0.25° resolution between the latitudes 58° and 74°; longitudes from 5° to 40° with 2° resolution at boundaries and 0.25° resolution between the longitudes 9° and 36°; and altitudes from ground level to 750 km with 25 km resolution and then up to 1250 km with 50 km resolution. This results as a grid of $n = 309\,120$ voxels. Combined with phase constant, bias, and plasma parameters, the total number of unknown variables in this case is $N = 309\,648$.

Now, when using the GMRF approach of Section IV, the resulting prior precision matrix has only 0.008% nonzero

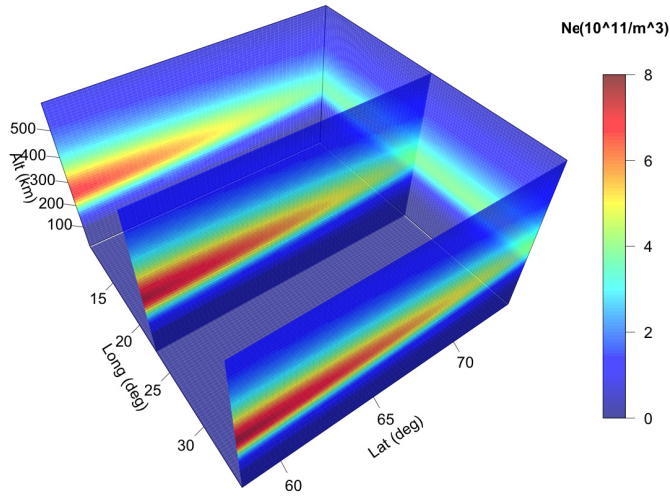


Fig. 4. IRI-2012 model electron density with default parameters for November 8, 2015 10:30:00 UTC.

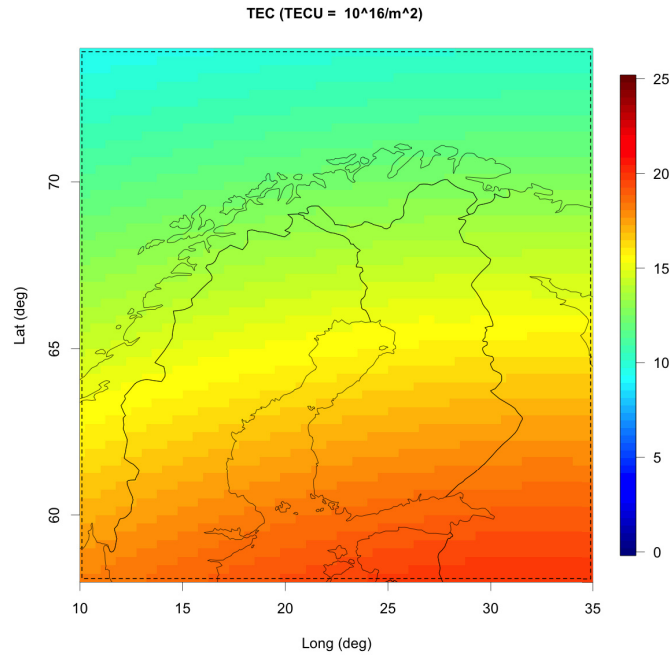


Fig. 5. Vertically integrated TEC (0–1250 km) from IRI-2012 model electron density with default parameters for November 8, 2015 10:30:00 UTC.

elements. When the prior precision is added together with the measurement information, in case, where all available instruments are used, the posterior precision that needs to be inverted in (6) still only has 0.034% nonzero elements.

Next, a simulation study is carried out, where the performance of the method is demonstrated with a known ionosphere taken from the IRI-2012 model. After simulation, the tomography is performed for the real measurements.

A. Simulation

The IRI-2012 model [28] electron densities are shown in Fig. 4, and the altitude integrated electron densities in Fig. 5. The IRI-2012 model was used with its default parameter values.⁵ The previously described electron density measurements

⁵https://omniweb.gsfc.nasa.gov/vitmo/iri2012_vitmo.html

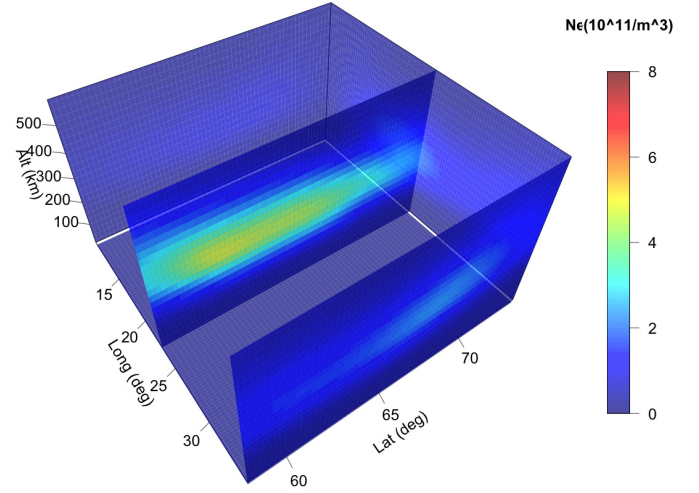


Fig. 6. Simulation case: reconstruction from simulated LEO beacon measurements with zero-mean GMRF prior.

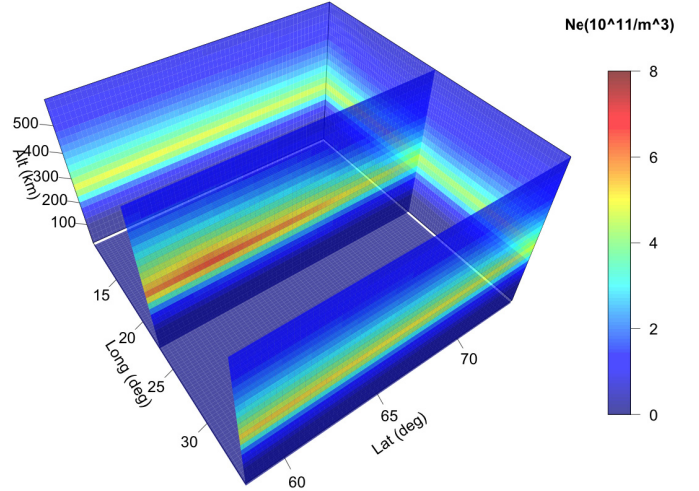


Fig. 7. Simulation case: reconstruction from simulated ground-based LEO satellite measurements. Simulated ionosonde measurements are used in the scaling of GMRF prior mean and variance mask.

are simulated with coordinates corresponding to real measurements. The measurement errors are simulated according to the estimated and assumed error distributions given earlier.

The tomographic inversion is then performed several times by adding the measurement sets one by one. The results are shown in Figs. 6–12.

In Fig. 6, only the LEO satellite measurements are used with a zero prior mean profile $\mathbf{x}_{pr} = \mathbf{0}$. The reconstructed electron density differs from zero only in the vicinity of the areas where the LEO beacon measurements are made. Even in the area of measurements, the electron density is underestimated.

In Fig. 7, again only the LEO satellite measurements are used, but the prior mean for electron densities \mathbf{x}_{pr} are taken from the simulated ionosonde profile corresponding to the location of EISCAT Dynasonde. The prior mean profile is used as such for the whole domain. Areas far from the measurements remain unchanged from the given prior value.

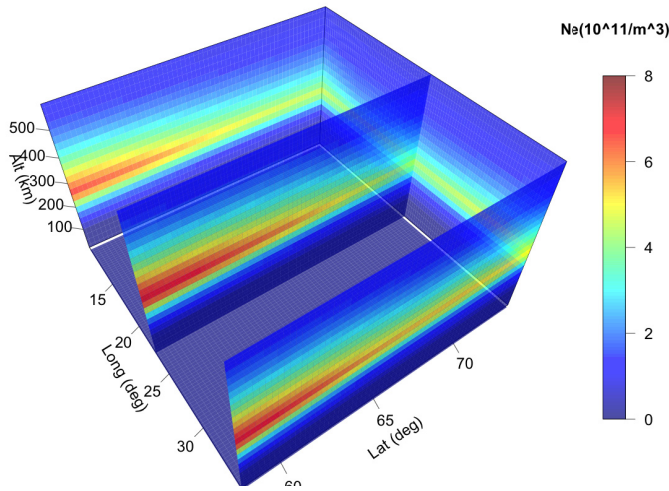


Fig. 8. Simulation case: reconstruction from simulated ground-based LEO and GPS satellite measurements. Simulated ionosonde measurements are used in the scaling of GMRF prior mean and variance mask.

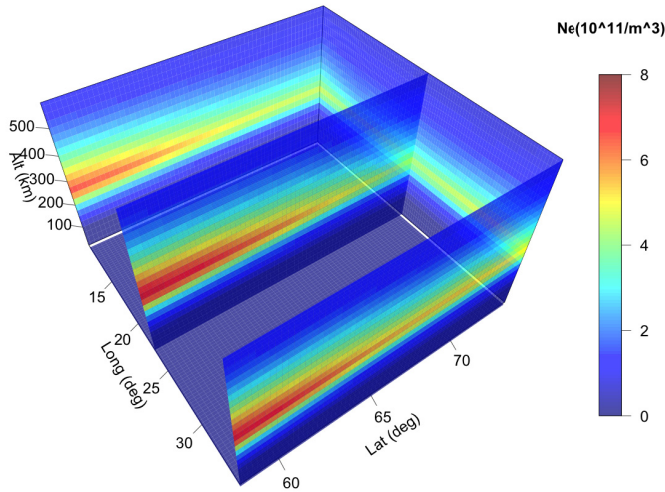


Fig. 9. Simulation case: reconstruction from simulations of ground-based LEO and GPS satellite measurements, ionosonde, satellite *in situ*, and satellite occultation measurements. Simulated ionosonde measurements are used in the scaling of GMRF prior mean and variance mask.

In Fig. 8, the simulated GPS measurements are added to the inversion. The reconstructed electron densities are changed from prior in a much wider area. However, due to the low inclination of GPS satellites, at the higher latitudes, the solution is somewhat dictated by the prior, especially in Northwest corner of the domain.

In Fig. 9, all simulated measurements are added to the inversion. When the ground-based GPS satellite observations are already included, the effect of satellite occultation and *in situ* measurements is mostly invisible. In Fig. 10, the reconstructed electron density is integrated along the altitude similar to Fig. 5. The relative differences between the vertically integrated TEC of IRI-2012 and the last reconstruction with all simulated measurements are shown in Fig. 11.

To demonstrate another aspect of Bayesian approach, the prior and posterior variances, $\sigma_{pr}^2 = \text{diag}(\Sigma_{pr})$ and

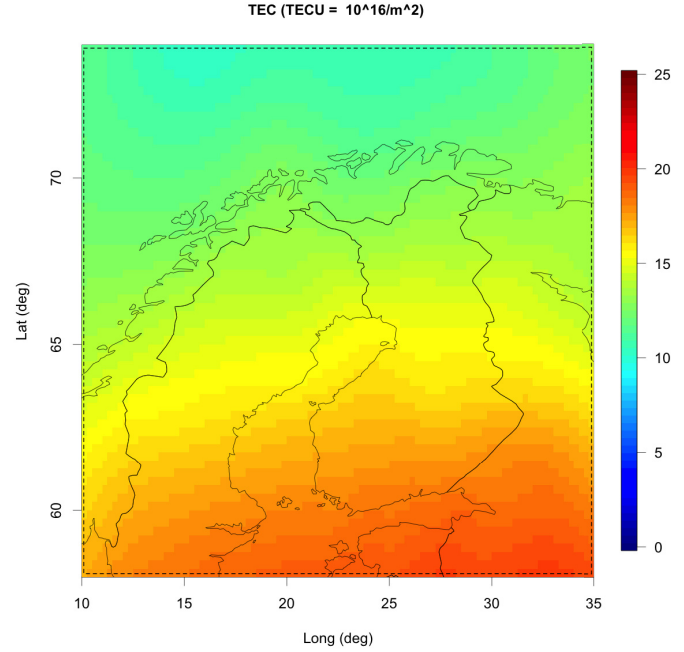


Fig. 10. Simulation case: vertically integrated TEC (0–1250 km) from tomographic reconstruction in Fig. 9.

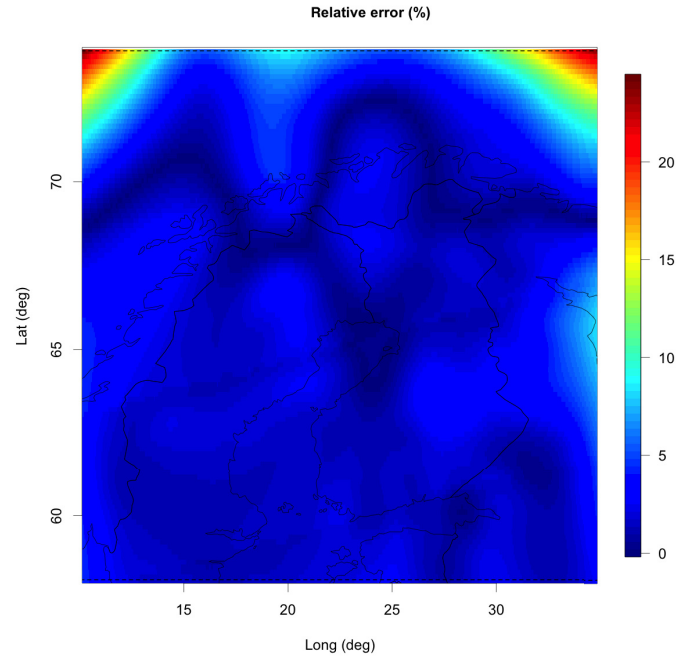


Fig. 11. Simulation case: relative error in vertically integrated TEC (0–1250 km) between IRI-2012 model in Fig. 5 and the tomographic reconstruction in Fig. 9.

$\sigma_{post}^2 = \text{diag}(\Sigma_{post})$, respectively, are computed to derive $\sigma_{expl}^2 = (1 - (\sigma_{post}^2)/(\sigma_{pr}^2)) \times 100\%$ to describe how much of the prior variance is explained with the measurements. If σ_{pr}^2 is not affected by the measurements at all, σ_{expl}^2 will be zero and if the information of the measurements is high, σ_{expl}^2 will be close to hundred.

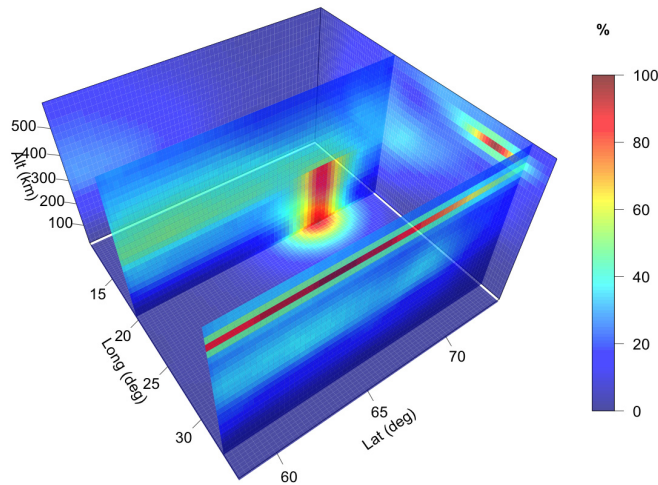


Fig. 12. Explained variance σ_{expl}^2 presents the relative difference of prior and posterior variances. When $\sigma_{\text{expl}}^2 = 0\%$, the prior assumption is not improved by the measurements at all. When $\sigma_{\text{expl}}^2 = 100\%$, the prior uncertainty is explained completely by the measurements. The explained variance depends only on measurement geometry, error and prior covariance. Here, the planes are shifted to longitudes intersecting the ionosonde and Swarm overflight locations.

In Fig. 12, σ_{expl}^2 is given for the case where all measurements are included in the computation of σ_{post}^2 . As σ_{expl}^2 can be visualized in the original grid for the electron density parameters, the uncertainty of the solution can be assessed at different locations. At areas with best measurement coverage, σ_{expl}^2 is high. The planes in Fig. 12 are moved to intersect the longitudes of EISCAT Dynasonde and Swarm *in situ* measurements to demonstrate the superior accuracy of the direct measurements. Fig. 12 shows that the satellite *in situ* measurements add information to system, but in this case, the measurements are in such a good agreement that the actual reconstruction is almost unchanged between Figs. 8 and 9.

B. Real Data

The actual GPS measurements are first presented with a single-layer model [17] in Fig. 13, where slant TEC measurements are mapped to vertical TEC, projected to location of 350 km ionospheric pierce points and bilinearly interpolated. The actual pierce points are also shown. Fig. 13 indicates lower TEC compared to IRI model from the same time in Fig. 5; hence, a lower scale is selected to the following visualizations of real data analysis.

In the inversion with real measurements the prior parameter values are unchanged, except for the mean and variance parameters that are scaled according to ionosonde. The 3-D electron density reconstruction is presented in Fig. 14, and the corresponding vertically integrated TEC in Fig. 15. The EISCAT UHF radar beam is projected in Fig. 15, and both ISR profiles are shown in Fig. 16. On top of the measured ISR electron densities are the corresponding profiles from the tomographic reconstruction.

VI. DISCUSSION

The EISCAT ISR validation presented in Fig. 16 is somewhat compromised with the proximity of EISCAT Dynasonde

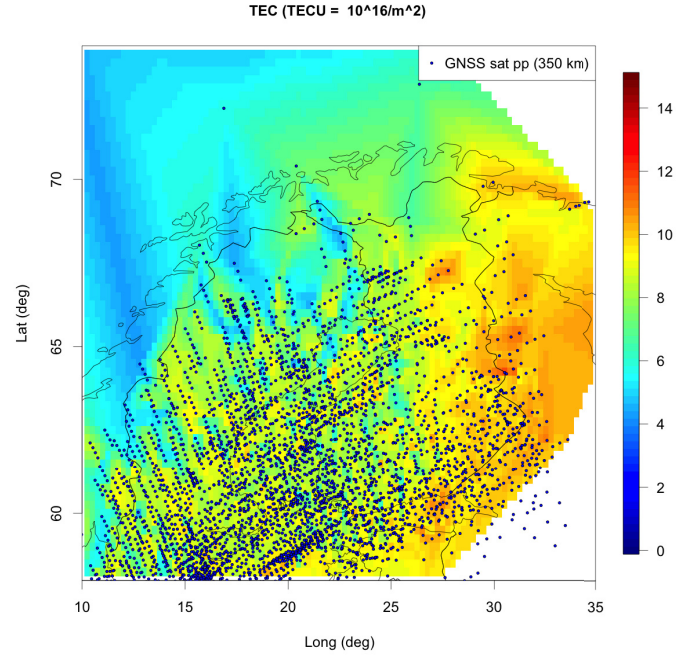


Fig. 13. Real data case: GPS measurements on November 8, 2015 from 10:18:00 UTC to 10:38:00 UTC mapped to vertical TEC with the single-layer model and bilinear interpolation.

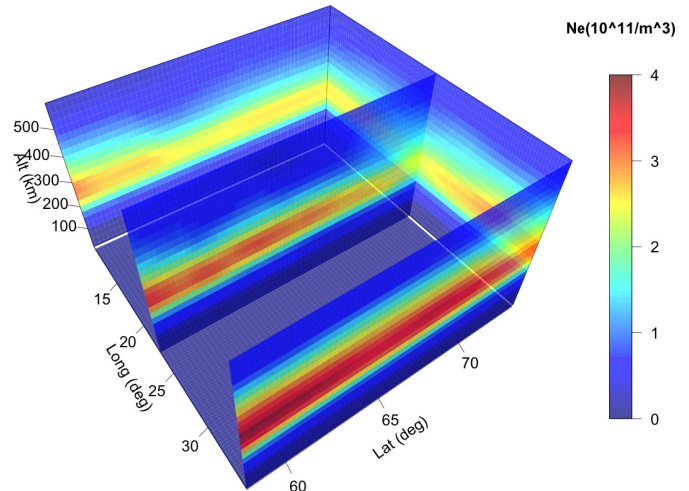


Fig. 14. Real data case: tomographic reconstruction on November 8, 2015 from 10:18:00 UTC to 10:38:00 UTC. Data consist of ground-based LEO and GPS satellite measurements, ionosonde, satellite *in situ*, and satellite occultation measurements. Ionosonde measurements are used in the scaling of GMRF prior mean and variance mask.

and ISR. Within the model resolution, the VHF radar and Dynasonde are measuring from the very same location. However, as can be seen from Fig. 15, with the low-elevation UHF profile, the location of the measured F-region maximum is several hundred kilometers Southeast from the EISCAT base. In comparison of vertical TEC mapped from GPS measurements in Fig. 13 and TEC integrated from reconstruction in Fig. 15, there is also an agreement between the main features.

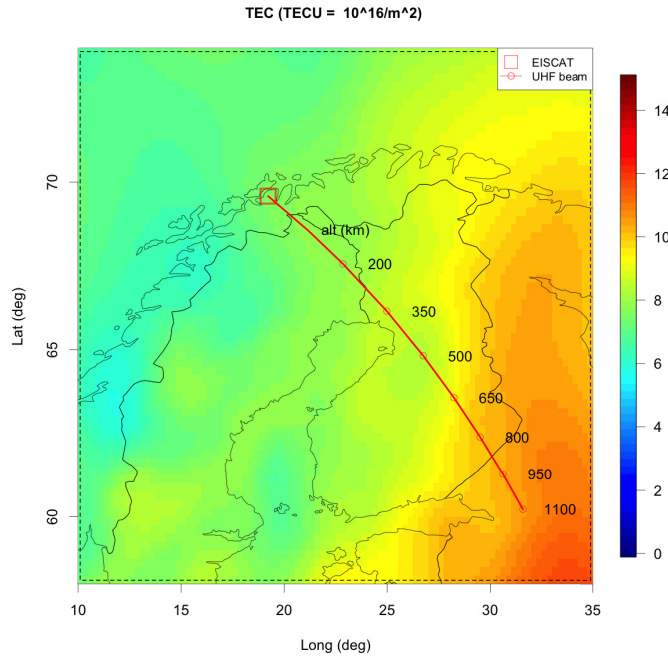


Fig. 15. Real data case: vertically integrated TEC (0–1250 km) from tomographic reconstruction in Fig. 14 with projected EISCAT UHF ISR beam that is used for validation.

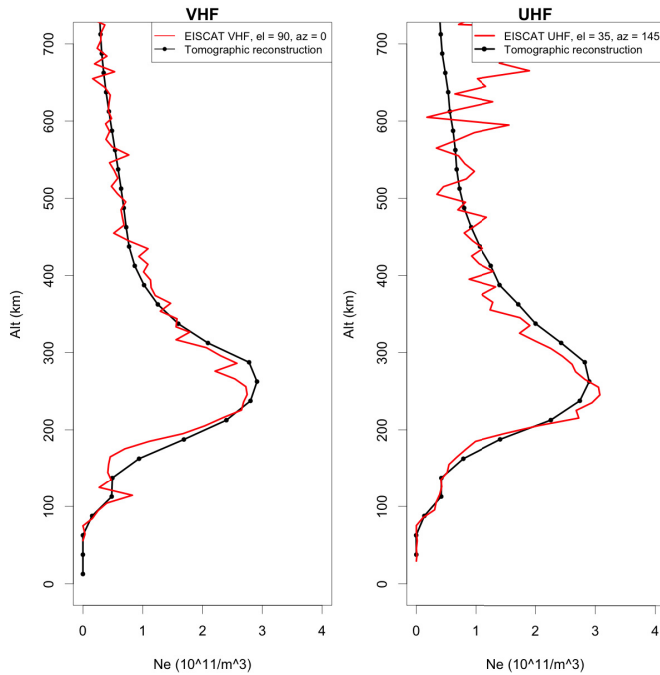


Fig. 16. Real data case: comparison between EISCAT ISR profiles and corresponding profiles from the tomographic reconstruction.

Obviously, for an individual case the parameters could be tuned endlessly; hence, more precise analysis on these results would not give a realistic concept on the overall performance from case to case. In the presented results, the prior parameters were not tuned between the cases, except for the different ionosonde profiles and the first trial with a zero prior. This suggests consistency of the model, but on the other hand, the robustness with respect to changes in the prior parameters

was not demonstrated. However, based on the development and work with the model, it can be said that in general, when a realistic prior is chosen, the changes in different parameters produce anticipated changes in the results. Some particular parameters are further discussed in the following.

In comparison of Figs. 13 and 15, the TEC map integrated from the reconstruction might seem slightly over smoothed. This is intuitive as the correlation length parameters were not changed from the simulation case and are relatively long considering the situation in Fig. 13. With shorter correlation lengths, the prior allows more small scale structures in the reconstruction, however, the prior mean dominates more strongly the areas with no measurement, as the effect of the nearby measurements is decreased.

When only LEO RTEC data are used, the distribution of the phase constants γ can have a significant effect. If the phase constant can be approximated beforehand, it improves the accuracy of the absolute level of the solution. When GNSS absolute TEC measurements and direct measurements are in use, the absolute level of LEO TEC can be estimated more accurately within the tomographic inversion.

The GPS measurements used here were already bias corrected. Therefore, a relatively small prior SD could be given for the GPS biases. A trial run was carried out also without the GPS receiver bias correction. The resulting reconstructions with larger bias SD were practically unchanged and the solved biases were very close to preprocessed ones. However, as the model was originally calibrated with bias-corrected data, the performance might be exaggerated with this respect. Then again, it is fair to assume that GNSS biases could be solved to some degree within the inversion, but it would add to general uncertainty of the solutions.

One of the last additions in the presented model was the modeling error. The satellite phase measurements are very accurate and can detect details much finer than a discrete model of this scale can represent. Also, in the time scale of tens of minutes, the changes in ionosphere can be significant. It is then intuitive to relax the model for these factors. In the presented case, the same model without any added modeling error will overestimate the electron density parameters to include unrealistic details.

Here, the nonzero prior mean was extrapolated from the ionosonde measurements horizontally uniformly for the whole domain. This is a rather simple approach and more detailed prior information could also be included. Similar to [14], it is possible to use other sources as the prior mean, for example, ionospheric models such as IRI. However, in regional scale, the models can be significantly off, hence, even a simple model that is based on direct ionospheric measurements is often preferable.

In [14], three methods for taking into account the time propagation were discussed. The presented GMRF prior can be used directly in the two first, where updates are not used at all, or only the background, i.e., the previous MAP-estimate is used without updating the covariance. In addition, instead of full covariance matrix its diagonal, the posterior variance, can be solved and used for scaling of the prediction covariance. Once implemented, the formation of the GMRF precision

matrix is so fast, that if the parameters of covariance function, e.g., the covariance lengths, can be estimated meaningfully, the change of covariance from one time to another does not increase the computation time in practice. For the general case of time propagation, the solution of posterior covariance is a major computational issue. The prediction step is additive for covariance, and the posterior covariance is needed for each time step. Even with the GMRFs, the solution for posterior covariance results a dense matrix and the parametric form is not known. Hence, in the general case, the GMRF contributes only to the first time step, and if the full posterior covariance is required for the subsequent time steps, the sparseness is lost.

VII. CONCLUSION

It is well known that ionospheric tomography is a very ill-posed problem and the atmospheric electron density cannot be reconstructed without including additional information into the system. It is mostly this information that separates the different tomographic approaches from another. Therefore, it is important to understand the nature of the constraining information in use.

In this paper, first the Bayesian statistical approach for multi-instrument ionospheric tomography is demonstrated. When Gaussian likelihood and prior distributions are assumed, the Bayesian method corresponds computationally to the widely used 3DVAR method. The method provides a clear physical interpretation for the required prior/background information. However, the problem with n unknowns is that it requires representation and operations of an $n \times n$ covariance matrix, and thus becomes computationally complex with large n .

Here, the approach is expanded with the use of GMRF priors to make the problem computationally feasible. In the presented example, a Gaussian prior field with a squared exponential covariance function is approximated with a GMRF. The resulting GMRF has a sparse prior precision matrix with only $25 \times n$ nonzero elements. In comparison to operating with $n \times n$ covariance matrix, this results in a significant decrease in the computational memory and time consumption. In the presented numerical example, the sparse matrix approximation for the $309\,648^2$ posterior precision matrix has only 0.034% nonzero elements, making the computation possible with a modern PC.

It is here shown how the GMRF model is constructed with physically interpretable covariance structure, parameterized with correlation lengths and variance mask, without forming the actual covariance matrix. The performance is validated with results from simulated and real multi-instrument data with comparisons to EISCAT ISR and vertical TEC mapped from original data.

The operative performance depends on how the prior parameters can be fixed or selected dynamically to different ionospheric conditions. Based on the presented results, it is reasonable to expect that this can be done at the level of any present ionospheric tomography system. Further validation of parameter selection requires its own study, which should be performed for consecutive reconstructions for longer time intervals.

VIII. FUTURE WORK

The authors are currently working on a 4-D Bayesian filtering for ionospheric tomography, where 3-D reconstructions are made dynamically to consecutive time instants.

ACKNOWLEDGMENT

The GPS data used in this paper are provided by Geotrim (Finland), Swepos (Sweden), and International GNSS service. The GPS data are preprocessed and bias corrected by Massachusetts Institute of Technology Haystack Observatory. The authors are grateful for all the GPS data providers. The authors would like to thank EISCAT scientific association for the ISR and Dynasonde data [EISCAT is an international association supported by research organizations in China (CRIRP), Finland (SA), Japan (NIPR and STEL), Norway (NFR), Sweden (VR), and the United Kingdom (NERC)]. They would also like to thank the International Reference Ionosphere working group for providing the model and data used in this paper, as well as ESA for the Swarm data products.

The authors are grateful for the developers of Multifrontal Massively Parallel sparse direct Solver (MUMPS) that was used in this paper for the computations with an R language interface RMUMPS (<https://github.com/morispa/rmumps>).

REFERENCES

- [1] J. R. Austen, S. J. Franke, and C. H. Liu, "Ionospheric imaging using computerized tomography," *Radio Sci.*, vol. 23, no. 3, pp. 299–307, May/Jun. 1988.
- [2] V. E. Kunitsyn, E. S. Andreeva, S. J. Franke, and K. C. Yeh, "Tomographic investigations of temporal variations of the ionospheric electron density and the implied fluxes," *Geophys. Res. Lett.*, vol. 30, no. 16, p. 1851, 2003, doi: [10.1029/2003GL016908](https://doi.org/10.1029/2003GL016908).
- [3] G. S. Bust and C. N. Mitchell, "History, current state, and future directions of ionospheric imaging," *Rev. Geophys.*, vol. 46, no. 1, pp. 1–23, Mar. 2008.
- [4] G. K. Seemala, M. Yamamoto, A. Saito, and C.-H. Chen, "Three-dimensional GPS ionospheric tomography over Japan using constrained least squares," *J. Geophys. Res., Space Phys.*, vol. 119, no. 4, pp. 3044–3052, Apr. 2014, doi: [10.1002/2013JA019582](https://doi.org/10.1002/2013JA019582).
- [5] Y. Yao, J. Tang, P. Chen, S. Zhang, and J. Chen, "An improved iterative algorithm for 3-D ionospheric tomography reconstruction," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 8, pp. 4696–4706, Aug. 2014.
- [6] C. H. Chen, A. Saito, C. H. Lin, M. Yamamoto, S. Suzuki, and G. K. Seemala, "Medium-scale traveling ionospheric disturbances by three-dimensional ionospheric GPS tomography 3. Space science," *Earth, Planets Space*, vol. 68, no. 1, pp. 1–9, 2016.
- [7] S. Saito, S. Suzuki, M. Yamamoto, A. Saito, and C.-H. Chen, "Real-time ionosphere monitoring by three-dimensional tomography over Japan," *Navigat., J. Inst. Navigat.*, vol. 64, no. 4, pp. 495–504, 2017.
- [8] A. Tarantola, *Inverse Problem Theory: Methods for Data Fitting and Model Parameter Estimation*. Amsterdam, The Netherlands: Elsevier, 1987.
- [9] J. Kaipio and E. Somersalo, *Statistical and Computational Inverse Problems* (Applied Mathematical Sciences). New York, NY, USA: Springer, 2005.
- [10] M. Markkanen *et al.*, "Bayesian approach to satellite radiotomography with applications in the Scandinavian sector," *Ann. Geophys.*, vol. 13, pp. 1277–1287, 1995.
- [11] O. Arikan, F. Arikan, and C. B. Erol, "3-D computerized ionospheric tomography with random field priors," in *Mathematical Methods in Engineering*. Dordrecht, The Netherlands: Springer, 2007, pp. 325–334.
- [12] J. Norberg, L. Roininen, J. Vierinen, O. Amm, D. McKay-Bukowski, and M. S. Lehtinen, "Ionospheric tomography in Bayesian framework with Gaussian Markov random field priors," *Radio Sci.*, vol. 50, no. 2, pp. 138–152, Feb. 2015.
- [13] J. Norberg *et al.*, "Bayesian statistical ionospheric tomography improved by incorporating ionosonde measurements," *Atmos. Meas. Techn.*, vol. 9, no. 4, pp. 1859–1869, 2016.

- [14] G. S. Bust, T. W. Garner, and T. L. Gaussiran, II, "Ionospheric Data Assimilation Three-Dimensional (IDA3D): A global, multisensor, electron density specification algorithm," *J. Geophys. Res.*, vol. 109, no. A11, pp. 1–14, 2004.
- [15] H. Rue and L. Held, *Gaussian Markov Random Fields: Theory and Applications* (Monographs on Statistics and Applied Probability). Boca Raton, FL, USA: Chapman & Hall, 2005.
- [16] J. Vierinen *et al.*, "Beacon satellite receiver for ionospheric tomography," *Radio Sci.*, vol. 49, no. 12, pp. 1141–1152, 2014.
- [17] I. Horvath and S. Crozier, "Software developed for obtaining GPS-derived total electron content values," *Radio Sci.*, vol. 42, no. 2, pp. 1–20, 2007.
- [18] W. Rideout and A. Coster, "Automated GPS processing for global total electron content data," *GPS Solutions*, vol. 10, no. 3, pp. 219–228, 2006. [Online]. Available: <https://link.springer.com/content/pdf/10.1007%2F10291-006-0029-5.pdf>
- [19] J. Vierinen, A. J. Coster, W. C. Rideout, P. J. Erickson, and J. Norberg, "Statistical framework for estimating GNSS bias," *Atmos. Meas. Techn.*, vol. 9, no. 3, pp. 1303–1312, 2016.
- [20] R. Furrer, M. G. Genton, and D. Nychka, "Covariance tapering for interpolation of large spatial datasets," *J. Comput. Graph. Statist.*, vol. 15, no. 3, pp. 502–523, 2006.
- [21] L. Roininen, M. S. Lehtinen, S. Lasanen, M. Orispää, and M. Markkanen, "Correlation priors," *Inverse Problems Imag.*, vol. 5, no. 1, pp. 167–184, 2011.
- [22] L. Roininen, P. Piironen, and M. Lehtinen, "Constructing continuous stationary covariances as limits of the second-order stochastic difference equations," *Inverse Problems Imag.*, vol. 7, no. 2, pp. 611–647, 2013.
- [23] F. Lindgren, H. Rue, and J. Lindström, "An explicit link between Gaussian fields and Gaussian Markov random fields: The stochastic partial differential equation approach," *J. Roy. Statist. Soc. Ser. B*, vol. 73, no. 4, pp. 423–498, 2011, doi: [10.1111/j.1467-9868.2011.00777.x](https://doi.org/10.1111/j.1467-9868.2011.00777.x).
- [24] L. Roininen, S. Lasanen, M. Orispää, and S. Särkkä, "Sparse approximations of fractional Matérn fields," *Scandin. J. Statist.*, vol. 45, no. 1, pp. 194–216, 2018.
- [25] T. N. Davis and M. Sugiura, "Auroral electrojet activity index AE and its universal time variations," *J. Geophys. Res.*, vol. 71, no. 3, pp. 785–801, Feb. 1966, doi: [10.1029/JZ071i003p00785](https://doi.org/10.1029/JZ071i003p00785).
- [26] N. Olsen *et al.*, "The Swarm Satellite Constellation Application and Research Facility (SCARF) and Swarm data products," *Earth, Planets Space*, vol. 65, no. 11, p. 1, Nov. 2013, doi: [10.5047/eps.2013.07.001](https://doi.org/10.5047/eps.2013.07.001).
- [27] N. A. Zlotin, J. W. Wright, and G. A. Zhabankov, "NeXTYZ: Three-dimensional electron density inversion for dynasonde ionograms," *Radio Sci.*, vol. 41, no. 6, pp. 1–12, Dec. 2006, doi: [10.1029/2005RS003352](https://doi.org/10.1029/2005RS003352).
- [28] D. Bilitza *et al.*, "The International Reference Ionosphere 2012—A model of international collaboration," *J. Space Weather Space Climate*, vol. 4, p. A07, Feb. 2014. [Online]. Available: <https://www.swsc-journal.org/articles/swsc/pdf/2014/01/swsc130043.pdf>, doi: [10.1051/swsc/2014004](https://doi.org/10.1051/swsc/2014004).
- [29] J. M. Dow, R. Neilan, and C. Rizos, "The international GNSS service in a changing landscape of global navigation satellite systems," *J. Geodesy*, vol. 83, nos. 3–4, pp. 191–198, Mar. 2009. [Online]. Available: <https://link.springer.com/content/pdf/10.1007%2F100190-008-0300-3.pdf>
- [30] P. R. Amestoy, I. S. Duff, J.-Y. L'Excellent, and J. Koster, "A fully asynchronous multifrontal solver using distributed dynamic scheduling," *SIAM J. Matrix Anal. Appl.*, vol. 23, no. 1, pp. 15–41, 2001, doi: [10.1137/S0895479899358194](https://doi.org/10.1137/S0895479899358194).
- [31] P. R. Amestoy, A. Guermouche, J.-Y. L'Excellent, and S. Pralet, "Hybrid scheduling for the parallel solution of linear systems," *Parallel Comput.*, vol. 32, no. 2, pp. 136–156, Feb. 2006. [Online]. Available: <http://linkinghub.elsevier.com/retrieve/pii/S0167819105001328>



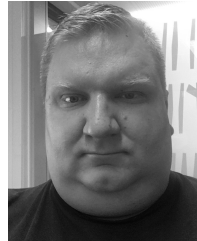
Johannes Norberg received the Ph.D. degree in applied mathematics from the University of Helsinki, Helsinki, Finland.

He is currently a Research Scientist with the Finnish Meteorological Institute, Helsinki. He is involved in inverse problems and scientific computing. He is responsible for the TomoScand Project for instrumentation and algorithm development for ionospheric tomography.



Juha Vierinen received the M.S. and Ph.D. degrees from the Helsinki University of Technology, Espoo, Finland, in 2005 and 2012, respectively.

From 2013 to 2016, he was with Massachusetts Institute of Technology Haystack Observatory, Westford, MA, USA. He is currently an Associate Professor of space physics with the University of Tromsø, Tromsø, Norway. His research interests include the development of novel radar and radio remote sensing measurement techniques to study space plasmas.



Lassi Roininen is currently an Adjunct Professor of applied mathematics with the University of Oulu, Oulu, Finland, and also a Post-Doctoral Researcher with the Academy of Finland, Sodankylä Geophysical Observatory, University of Oulu. He is involved in rigorous numerical and computational tools for inverse problems with applications in near-space remote sensing, subsurface imaging, and X-ray tomography.



Mikko Orispää received the Ph.D. degree in mathematics from the University of Oulu, Oulu, Finland, in 2002.

He is involved in inverse problems and scientific computing. He is currently an Application Designer with Sodankylä Geophysical Observatory, University of Oulu, where his responsibilities include the research and development of new measurement methods, and the development of scientific software for the measurement device operation, measurement data handling, and analysis.



Kirsti Kauristie was a Coordinator in several research projects funded by ESA and EU. She has co-authored 95 refereed articles (h-index 17). Her research interests include high-latitude ionospheric physics: Auroras, their electrodynamics, and variations in the ionospheric electron content.

Ms. Kauristie has been representing Finland in the Program Board of the ESA Space Situational Awareness Program since 2012. Since 2017, she has been a member of the Space Weather Assessment and Consolidation Working Group established by the European Science Foundation. She has served as a Topical Editor for the EGU Journal *Annales Geophysicae* from 2008 to 2012, the Council Chair for the EISCAT Association from 2009 to 2010, and the Co-Chair for the Team which composed the Cospar/ILWS Roadmap for Space Weather Research from 2015 to 2025.



William C. Rideout was born in East Orange, NJ, USA, in 1958. He received the B.A. degree in physics and English literature from Bucknell University, Lewisburg, PA, USA, in 1980, the master's degree from the Massachusetts Institute of Technology, Cambridge, MA, USA, in 1985, and the master's degree in computer engineering from the University of Massachusetts, Lowell, MA, USA, in 2001.

Since 2001, he has been with Massachusetts Institute of Technology Haystack Observatory, Westford, MA, USA.



Anthea J. Coster received the Ph.D. degree from Rice University, Houston, TX, USA, in 1983.

She was with Arecibo Observatory, Arecibo, Puerto Rico. She has been involved with GPS since 1985, and leads the Global Navigation Satellite System (GNSS) Research Program at Haystack. She is currently an Assistant Director and a Principal Research Scientist with Massachusetts Institute of Technology Haystack Observatory, Westford, MA, USA. Her research interests include space weather, ionospheric and atmospheric coupling, and GNSS positioning and measurement accuracy.



Markku S. Lehtinen is currently a Research Professor with Sodankylä Geophysical Observatory, University of Oulu, Oulu, Finland. He is currently an internationally recognized leading expert in statistical inverse methods and applications in incoherent scatter radar studies. He has developed measurement principles, as well as radar construction principles for the next-generation incoherent scatter radar EISCAT_3D.



ILMATIETEEN LAITOS
METEOROLOGISKA INSTITUTET
FINNISH METEOROLOGICAL INSTITUTE

Finnish Meteorological Institute
Erik Palménin aukio 1
P.O. Box 503
FI-00560 HELSINKI
tel. +358 29 539 1000
WWW.FMI.FI

Finnish Meteorological Institute
CONTRIBUTIONS No. 173
ISSN 0782-6117
ISBN 978-952-336-123-2 (paperback)
ISBN 978-952-336-124-9 (pdf)
<https://doi.org/10.35614/isbn.9789523361249>

Helsinki, 2020
Edita Prima Oy

