Multiresolution tomography for the ionosphere

Panicciari, Tommaso

Award date:
2016

Awarding institution:
University of Bath

Multiresolution tomography for the ionosphere

Tommaso Panicciari

A thesis submitted for the degree of Doctor of Philosophy

University of Bath

Department of Electronic and Electrical Engineering

May 2015

Acknowledgments

I owe my deepest gratitude to my supervisors for their continuous guidance and support during my Ph.D. In particular, I would like to thank Nathan Smith who devoted ideas, time and patience to my work, and Cathryn Mitchell who kept me focused, ensuring my progression toward completion. My skills and competences have greatly benefited from both of you.

It's a pleasure to thank Federico Da Dalt, Julian Rose, Mark Greco and Talini Pinto Jayawardena for their encouragement, support and for simply being good friends. This work would hardly have been completed without that.

I want to thank the members of the INVERT Group of the University of Bath for their precious suggestions and discussions. Special thanks go also to Gary Bust for his constructive comments, and Paul Spencer who opened my eyes on important aspects regarding ionospheric tomography.

Finally, I would like to thank my family for their continuous support and encouragement, and my partner Chiara Villari. This thesis would not have been possible without their great sacrifices and patience, especially during this last year.

Global Positioning System (GPS) data were provided by the Crustal dynamics data information system (CDDIS), the International GNSS Service (IGS), the Ecole Nationale des Sciences Géographiques IGS service, the University Navstar Consortium (UNAVCO), and the Garner GPS archive. CHAMP data were kindly provided by David Cooke (AF Research Lab). TRANSIT data were originally collected as part of a NSF project (ATM# 9813864) and radar observations were supported by the NSF Cooperative Agreement AGS-0836152.

This research activity was funded by the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no. 264476.

Abstract

The ionosphere is a dynamic and ionized medium. Specification of the ionospheric electron density is important for radio systems operating up to a few GHz. Such systems include communication, navigation and surveillance operations. Computerized Ionospheric Tomography (CIT) is a technique that allows specification of the electron density in the ionosphere. CIT, unlike medical tomography, has geometric limitations such as uneven and sparse distribution of ground-based receivers and limited-angle observations. The inversion is therefore underdetermined and to overcome the geometric limitations of the problem, regularization techniques need to be used.

In this thesis the horizontal variation of the ionosphere is represented using wavelet basis functions. Wavelets are chosen because the ground based ionospheric instrumentation is unevenly distributed and hence there is an expectation that the resolution of the tomographic image will change across a large region of interest. Wavelets are able to represent structures with different scale and position efficiently, which is known as Multi Resolution Analysis (MRA). The theory of sparse regularization allows the usage of a small number of basis functions with minimum loss of information. Furthermore, sparsity through wavelets can better differentiate between noise and actual information. This is advantageous because it increases the efficacy to resolve the structures of the ionosphere at different spatial horizontal scale sizes. The basis set is also extended to incorporate time dependence in the tomographic images by means of three-dimensional wavelets.

The methods have been tested using both simulated and real observations from the Global Navigation Satellite System (GNSS). The simulation was necessary in order to have a controllable environment where the ability to resolve different scale structures would be tested. The further analysis of the methods required also the use of real observations. They tested the technique under conditions of temporal dynamics that would be more difficult to reproduce with simulations, which often tend to be valid in quiet ionospheric behaviours.

Improvements in the detection and reconstruction of ionospheric structures were illustrated with sparse regularization. The comparison was performed against two standard methods. The first one was based on spherical harmonics in space, whilst the second relied on a time-dependent smoothing regularization.

In simulation, wavelets showed the possibility to resolve small-scale structures better than spherical harmonics and illustrated the potential of creating ionospheric maps at high resolution. In reality, GNSS satellite orbits allow satellite to receiver datasets that traverse the ionosphere at a few hundred km per second and hence a long time window of typically half an hour may be required to provide observations. The assumption of an unchanging ionosphere is only valid at some locations under very quiet geomagnetic conditions and at certain times of day. For this reason the theory was extended to include time dependence in the wavelet method. This was obtained by considering two approaches: a time-smooth regularization and three-dimensional wavelets. The wavelet method was illustrated on a European dataset and demonstrated some improvements in the reconstructions of the main trough.

In conclusion wavelets and sparse regularization were demonstrated to be a valid alternative to more standard methods.

Table of contents

1 Acknowledgments ii

2 Abstract iii

3 Table of contents v

4 List of tables ix

5 List of figures x

6 List of abbreviations xvi

7 Notations and symbols xix

1 Chapter 1 1

Introduction 1

2 Chapter 2 5

Ionosphere: principles and measurements 5

Introduction 5

2.1 The Ionosphere 5

2.1.1 Principles of signal propagation in the ionosphere 8

2.2 GPS, from the origins to the application 10

2.2.1 Transit, the ancestor of GPS 11

2.2.2 GPS signal propagation 12

2.2.3 TEC measurements from GPS 13

2.3 Instruments used for validation 16

2.3.1 CHAMP 17

2.3.2 Incoherent scatter radar 19

2.4 Summary 23

3 Chapter 3 24

State of the art 24

Introduction 24

3.1 Origins and evolution of ionospheric tomography 24

3.2 Early validation results and further improvements 29
3.3 State of art in CIT 31
3.3.1 IDA4D 31
3.3.2 MIDAS 32
3.3.3 GAIM (by USU) 33
3.3.4 GAIM (by USC/JPL) 34
3.4 Summary 34

4 Chapter 4 36

Ionospheric tomography 36

Introduction 36
4.1 Definition of the problem of ionospheric tomography 36
4.2 Tikhonov regularization 38
4.2.1 Performance of Tikhonov regularization by varying its tuning parameters 40
4.3 The weighting matrix 45
4.4 Resolution Matrix 45
4.5 Basis function decomposition 47
4.5.1 Vertical Basis Functions 48
4.5.2 Horizontal Basis Functions 49
4.5.2.1 Spherical harmonics 50
4.5.2.2 Wavelets 50
4.5.3 Multi-Resolution Analysis 51
4.6 Summary 54

5 Chapter 5 55

Sparse regularization techniques 55

Introduction 55
5.1 The calibration of relative measurements in CIT 56
5.2 The regularization term 58
5.2.1 ℓ_2 regularization (or Tikhonov regularization) 58
5.2.2 ℓ_0 regularization (or sparse regularization) 60
5.2.3 ℓ_1 regularization for sparse regularization 60
5.2.4 Total Variation 63
5.3 Implementation of the algorithms 64
5.4 Comparison of the algorithms 66
5.5 Summary 69

6 Chapter 6 71

Simulation and first results 71

Introduction 71

6.1 Inversion 72
6.2 Basis functions in ionospheric tomography 73
6.2.1 Two regularization techniques 75
6.3 A simulated case study 76
6.3.1 Data 77
6.3.2 Results from the simulated ionosphere 78
6.3.2.1 Robustness of sparse regularization technique 82
6.3.2.2 Offsets estimation using ionospheric tomography 85
6.3.2.3 Multi-resolution map from wavelet decomposition 86
6.3.2.4 Model-aided inversion, the best scenario 87
6.4 A real data case study 89
6.4.1 Data 89
6.4.2 Results from real data case study 90
6.5 High resolution with wavelets 94
6.6 Summary 98

7 Chapter 7 100

New time-dependent algorithm using sparse regularization 100

Introduction 100
7.1 Review of imaging the mid-latitude trough 101
7.2 Method 102
7.2.1 Sparsity meets time 103
7.2.2 Comparing TEC maps with CHAMP electron density profiles 106
7.3 Data 107
7.4 Sensitivity to the regularization parameter 108
7.5 Results and analysis 110
7.5.1 Case Study 1, 22 October 2009 110
7.5.2 Other case studies 120
7.6 Summary 131

8 Chapter 8 133

Conclusions and future work 133

9 Appendix A 137

A.1 Regularized Least Squares (RLS) Algorithm 137

10 Appendix B 140

B.1 The calibration matrix 140

11 Appendix C 142

C.1 19 November 2009 143

C.2 05 December 2009 144

C.3 14 July 2009 145

C.4 23 July 2009 146

12 Appendix D 147

D.1 Normalized TEC compared with CHAMP 147

D.2 Normalized TEC gradients compared with CHAMP 148

13 References 150

List of tables

Table 6-1. RMS error (values are in TECU) of the VTEC map obtained with spherical harmonics and wavelets at low resolutions. Only the VTEC coefficients where there is ray coverage were considered. The percentage of basis functions with non-zero coefficients is also shown and, within brackets, the number in absolute value. 81
Table 6-2. RMS error (values are in TECU) of the VTEC map obtained with spherical harmonics and wavelets at high resolutions. Only the VTEC coefficients where there is ray coverage were considered. The percentage of basis functions with non-zero coefficients is also shown and, within brackets, the number in absolute value. 81
Table 6-3. RMS error of reconstructed ionosphere for spherical harmonics and discrete Meyer with and without the Gaussian noise term. The percentage of basis functions with non-zero coefficients is shown and, within brackets, the number in absolute value. 84
Table 6-4. RMS error (values are in TECU) of the VTEC map obtained with a 3DVar scheme using spherical harmonics and discrete Meyer with a noise term added to the observations. Only the VTEC coefficients where there is ray coverage were considered. The percentage of basis functions with non-zero coefficients is also shown and within brackets the number in absolute value 89
Table 7-1. RMS error of the VTEC map obtained with MIDAS and wavelets using CHAMP as reference for the 22nd October 2009. The maximum cross-correlation values (C_{MAX}, normalized to 100) between the reconstructed values and CHAMP, and the MSSIS index are also included. The percentage of basis functions with non-zero coefficients is also shown. For each basis function the time-smooth (TS) and time-sparse (TP) results are illustrated. 120
Table 7-2. The table shows different indices calculated from normalized VTEC maps obtained with standard MIDAS (SM) and wavelets using CHAMP as reference. The indices are the RMS error (values are in TECU), the maximum cross-correlation values (CMAX, normalized to 100) between the reconstructed values and CHAMP and the MSSIS index. The percentage of basis functions with non-zero coefficients is also shown. The time-smooth (TS) and time-sparse (TP) results are illustrated for each basis function and for different case studies. Each index was calculated by averaging the values obtained for each case study (five in total). An average of those (normalized) four indices is also included. 130

List of figures

Figure 2.1. The ion production rate can be seen as the combination of the radiation intensity (from the sun) and the neutral densities of the species in the atmosphere. 6
Figure 2.2. Run of the IRI2012 empirical model on the day of 5 October 2014 at 21:00 UT. The map shows the TEC obtained from the empirical model. Values are in TECU. 8
Figure 2.3. CHAMP. 17
Figure 2.4: Picture of the Digital Ion Drift Meter (DIDM). 18
Figure 2.5. Ion spectrum and plasma lines [*Beynon and Williams*, 1978] (© IOP Publishing. Reproduced with permission. All rights reserved). 22
Figure 3.1. A simplified picture of the tomography scan scenario. The satellites transmit a signal received by the ground station. Each signal carries an information about the TEC integrated along the ray path and it is used to calculate the electron content along the ray path (represented by the grey pixels in the grid for one particular ray). 24
Figure 3.2. The block diagram illustrates the main components of MIDAS. The blocks modified in this thesis are shown in purple. 33
Figure 4.1. The ionosphere is discretized into a 3 dimensional grid with dimensions $L = \mathrm{L_V} \times \mathrm{L_{H_x}} \times \mathrm{L_{H_y}}$. The grid is stored in a column vector n whose coefficients are ordered by columns, as seen in the figure. 37
Figure 4.2. The ionosphere was created according to this normalized profile, and the same profile has been used to create the simulated ionosphere. 41
Figure 4.3. Simulated ionosphere produced for scaled versions of one IRI profile. Two structures with a peak around 300 km are present. The green rays represent the ray path between the satellites and the receivers. 41
Figure 4.4. The original TEC profile (blue) is compared with reconstructed TEC (black) by applying a regularization with $\alpha = 0.1$ fixed and $\varrho = 0.1 \dots 10$. It can be noted that, as ϱ increases, the regularization tends to reduce the rate of change of the electron content. 42
Figure 4.5. The influence of the parameter α in the inversion result is shown here. The image has been obtained with $\varrho = 0.1$ and with different values of $\alpha = 0.1 \dots 1$. As α gets bigger, the inversion minimizes the solution trying to better match the electron content with the n_0 profile (e.g. coming from a model). In this particular example, n_0 has been considered a zero vector. 43

Figure 4.6. Simulated ionosphere produced for scaled versions of one IRI profile. Different structures with a peak around 300 km are present. The green rays represent the ray path between the satellites and the receivers. 44
Figure 4.7. The RMS error between the true (simulated) and reconstructed ionosphere is shown for different values of ϱ and α. The figure shows the RMS error for: a) ϱ and α varying betwwen 0 and 40; b) α varying between 0 and 40 and ϱ equals to 3. Generally these curves present a minimum which is selected for a particular value of the two tuning coefficients. ... 44
Figure 4.8. The resolution matrix (right) gives information about the reliability of the solution obtained through inversion techniques. The columns are called point spread functions (PSF) and are representative of the actual resolution of a particular cell in the grid (left). If the PSF is not a delta like function, the value of the cell is spread to the neighbours and its resolution is decreased. 46
Figure 4.9. The signal n_i can be represented as the sum of different basis functions ϕ_{ij} weighted with the coefficients x_j. 47
Figure 4.10. The multi-resolution decomposition allows decomposing the signal into different levels. Each level takes into account different details and resolutions. 52
Figure 4.11. A signal g (top signal) can be split into different levels of approximation. Each level contains a different "version" or approximation of the signal, in terms of details as well as resolution. According to wavelet theory, it can be divided into a part which takes into account an averaged behaviour (the two left signals), and into a series of other versions which capture the details needed to describe the whole signal (the signals on the right). Each level of approximation can be reconstructed by means of the previous levels according to the formula $\mathrm{V_i} = \mathrm{V_0} \oplus \left(\oplus_{\mathrm{k=0}}^{\mathrm{i-1}} \mathrm{W_k}\right)$. 53
Figure 5.1. A two dimensional pictorial representation of the minimum solution for the ℓ_2 regularization. The ℓ_2 ball is a circle that corresponds to the points with the same norm $\|x\|_2^2$. The grey and thick line represents all possible solutions that satisfies $\|z - \mathrm{AK}x\|_{\mathrm{C}}^2 < \delta$. The unique solution is found at the intersection of the circle with the line..... 59
Figure 5.2. Two dimensional pictorial representation of the minimum solution for the $\ell 1$ regularization. The ℓ_1 ball is a diamond-shape that corresponds to the points with the same norm x_1. In constrast, the grey and thick line represents all possible solutions that satisfies $\|z - \mathrm{AK}x\|_{\mathrm{C}}^2 < \delta$. The unique solution is found at the intersection of the diamond-shape and the line. The intersection is likely to be at the edges of the diamond-shape, making the solution sparse. 61
Figure 5.3. Artificial signal used to compare the different regularization terms described in this chapter. 66

Figure 5.4. The reconstruction obtained with the ℓ_2 regularization (black) and the original signal (red) is shown in this figure. The regions with no data coverage are indicated in purple. 67
Figure 5.5. The reconstruction obtained with the ℓ_1 regularization using DB4 (black) and Haar (green) and the original signal (red) is shown in this figure. The regions with no data coverage are indicated in purple. 68
Figure 5.6. The reconstruction obtained with TV regularization (black) and the original signal (red) is shown in this figure. The estimation from the ℓ_2-regularized solution was used as input for the TV algorithm. The regions with no data coverage are indicated in purple. 68
Figure 5.7. The reconstruction obtained with TV regularization (black and green) and the original signal (red) is shown in this figure. The estimations from the ℓ_1-regularized solution using DB4 (black) and Haar (green) were used as input for the TV algorithm. The regions with no data coverage are indicated in purple. 69
Figure 6.1. a) discretised Meyer basis function for a particular scale and translation; b) Daubechies 4 basis function with the same scale and translation as a); c) a Fourier sinusoid component of the spherical harmonic basis functions. Basis functions are shown normalized to one and are interpolated for ease of viewing. 74
Figure 6.2. Simulated ionosphere with structures added to IRI2012. Values are in TECU (10^{16} electrons per m^{-2}). 78
Figure 6.3. Number of rays with ground stations (black dots). 78
Figure 6.4. Reconstructions obtained at low resolution with masked out VTEC values where there is no ray coverage for: a) spherical harmonics; b) discrete Meyer; c) Daubechies 4; and without the mask for d) spherical harmonics; e) discrete Meyer; f) Daubechies 4. Values are in TECU (10^{16} electrons per m^{-2}). 79
Figure 6.5. Reconstructions obtained at high resolution with masked out VTEC values where there is no ray coverage for: a) spherical harmonics; b) discrete Meyer; c) Daubechies 4; and without the mask for d) spherical harmonics; e) discrete Meyer; f) Daubechies 4. Values are in TECU (10^{16} electrons per m^{-2}). 80
Figure 6.6. A pictorial representation of the representativity error. The sketch shows the ambiguity given by the discretization of the ionosphere over a grid. Three different rays cross the voxel and bring three different information from the ionosphere (an ionospheric structure is represented in purple). Since only one value of electron density can be estimated per voxel, the discrepancy of the measurement from those three rays is called representativity error. 82
Figure 6.7. Low resolution reconstructions without Gaussian noise term for: a) spherical harmonics; c) discrete Meyer; and with Gaussian noise term for: b) spherical harmonics; d) discrete Meyer. Vertical TEC values where there is no ray coverage are masked out. 83

Figure 6.8. High resolution reconstructions without Gaussian noise term for: a) spherical harmonics; c) discrete Meyer; and with Gaussian noise term for: b) spherical harmonics; d) discrete Meyer. Vertical TEC values where there is no ray coverage are masked out. 84
Figure 6.9. Scatter plot of the estimated offsets (y-axis) versus the true offsets (x-axis) with a) spherical harmonics and b) discrete Meyer at low resolution. 85
Figure 6.10. Scatter plot of the estimated offsets (y-axis) versus the true offsets (x-axis) with a) spherical harmonics and b) discrete Meyer at high resolution. 86
Figure 6.11. Multi-Resolution (MR) Map for the high resolution case with discrete Meyer basis functions. Each box represents the scale of the basis function and its position. 87
Figure 6.12. Model-aided reconstruction obtained with a) spherical harmonics and b) discrete Meyer at high resolution. A noise term (zero-mean Gaussian with 1TEC unit of standard deviation) was added to the observations 88
Figure 6.13. Number of rays and GPS ground stations (blue), TRANSIT ground stations (red), TRANSIT satellite pass (purple) and radar scan path (black). 90
Figure 6.14. Reconstructions obtained at low resolution with masked out VTEC values where there is no ray coverage for: a) spherical harmonics; b) discrete Meyer; and without the mask for c) spherical harmonics; d) discrete Meyer. Values are in TECU (10^{16} electrons per m^{-2}). 91
Figure 6.15. Reconstructions obtained at high resolution with masked out VTEC values where there is no ray coverage for: a) spherical harmonics; b) discrete Meyer; and without the mask for c) spherical harmonics; d) discrete Meyer. Values are in TECU (10^{16} electrons per m^{-2}). 92
Figure 6.16. Southward longitudinal Incoherent Scatter (IS) radar scan starting at 03:21:20UT on 30th September 2000. Values of electron density are in 10^{11} particles/m^3. 93
Figure 6.17. Cross sections from low resolution tomographic reconstructions for: a) spherical harmonics; b) discrete Meyer; and from high resolution tomographic reconstructions for: c) spherical harmonics; d) discrete Meyer. Values of electron density are in 10^{11} particles/m^3. . 94
Figure 6.18. Normalized number of rays from GPS ground stations. It is also shown in purple the satellite CHAMP pass. 95
Figure 6.19. Reconstruction during the day of 02 January 2010 at 00:50UT. The map shows TEC calculated with sparse regularization and DM wavelets. TEC is shown only where there is data coverage. 95
Figure 6.20. Normalized electron density obtained from the reconstructed map (blue/circle) at the same altitude and position of CHAMP (red/cross) with sparse regularization and DM wavelets 96
Figure 6.21. Reconstruction during the day of 02 January 2010 at 00:50UT. The map shows TEC calculated with TV regularization using the sparse regularization reconstruction as input. TEC is shown only where there is data coverage. 97

Figure 6.22. Normalized electron density obtained from the reconstructed map (blue/circle) at the same altitude and position of CHAMP (red/cross) with TV regularization. 97
Figure 7.1: Pictorial representation of a time window consisting of three frames. Each frame contains observations (in yellow) which are distributed only partially on a hypothetical grid covering the continents Europe, Africa and Asia. The combination of the frames produces a better coverage, but the description of the time variation of the measurement between one frame to the other must be provided. 105
Figure 7.2. Normalized number of available rays for two different times. The coverage corresponds to a grid over Europe with 2.5 degree resolution in latitude and longitude. 108
Figure 7.3. Reconstructions obtained with (a-c) standard MIDAS (left), time-smooth regularization with (d-f) Daubechies 4 and (g-i) discrete Meyer, and time-sparse regularization with (j-l) Daubechies 4 and m-o) discrete Meyer. The images show TEC maps of 23 July 2009 at 23:40 UT for three different values of α and ϱ. 109
Figure 7.4. Reconstructions obtained with (a-c) standard MIDAS, (d-f) Daubechies 4 and (g-i) discrete Meyer with time-smooth regularization. The images show TEC maps every 30 minutes starting from 01:10 UT of 20 October 2009. 111
Figure 7.5. Reconstructions obtained with (a-c) standard MIDAS, (d-f) Daubechies 4 and (g-i) discrete Meyer with time-sparse regularization. The images show TEC maps every 30 minutes starting from 01:10 UT of 20 October 2009. 112
Figure 7.6. Comparison between normalized CHAMP and normalized TEC maps by using a) Daubechies 4; b) discrete Meyer; c) MIDAS with the time-smooth algorithm. CHAMP is shown in terms of normalized electron density as coloured circles over the reconstruction maps. The colour is proportional to the normalized electron density. The bottom plot shows, instead, the normalized electron density (black) and normalized TEC for Daubechies 4 (red) and discrete Meyer (orange) and MIDAS (green) along the satellite CHAMP pass. The plot shows two axes: the latitude (degrees) and the CHAMP time when the measurement was taken. TEC maps are calculated at 01:40 UT of 22 October 2009. 114
Figure 7.7. Comparison between normalized CHAMP and normalized TEC maps by using a) Daubechies 4; b) discrete Meyer; c) MIDAS with the time-sparse algorithm. CHAMP is shown in terms of normalized electron density as coloured circles over the reconstruction maps. The colour is proportional to the normalized electron density. The bottom plot shows, instead, the normalized electron density (black) and normalized TEC for Daubechies 4 (red) and discrete Meyer (orange) and MIDAS (green) along the satellite CHAMP pass. The plot shows two axes: the latitude (degrees) and the CHAMP time when the measurement was taken. TEC maps are calculated at 01:40 UT of 22 October 2009. 116
Figure 7.8. Normalized (a) TEC and (b) TEC gradients interpolated over time and space with a spline polynomial function for Daubechies 4 (red), discrete Meyer (orange) and MIDAS

(green). The CHAMP normalized electron density (black) is also shown. The results were calculated for the 22nd October 2009 with the time-smooth regularization. 117
Figure 7.9. Normalized (a) TEC and (b) TEC gradients interpolated over time and space with a spline polynomial function for Daubechies 4 (red), discrete Meyer (orange) and MIDAS (green). The CHAMP normalized electron density (black) is also shown. The results were calculated for the 22nd October 2009 with the time-sparse regularization..... 118
Figure 7.10. Reconstructions obtained with a-c) standard MIDAS (left), time-smooth regularization (centre) with d-f) Daubechies 4 and g-i) discrete Meyer and time-sparse regularization (right) with j-l) Daubechies 4 and m-o) discrete Meyer. The images show TEC maps every 30 minutes starting from 00:00 UT of 20 November 2009. 122
Figure 7.11. Reconstructions obtained with a-c) standard MIDAS (left), time-smooth regularization (centre) with d-f) Daubechies 4 and g-i) discrete Meyer and time-sparse regularization (right) with j-l) Daubechies 4 and m-o) discrete Meyer. The images show TEC maps every 30 minutes starting from 22:00 UT of 05 December 2009. 123
Figure 7.12. Reconstructions obtained with a-c) standard MIDAS (left), time-smooth regularization (centre) with d-f) Daubechies 4 and g-i) discrete Meyer and time-sparse regularization (right) with j-l) Daubechies 4 and m-o) discrete Meyer. The images show TEC maps every 30 minutes starting from 23:30 UT of 14 July 2009..... 124
Figure 7.13. Reconstructions obtained with a-c) standard MIDAS (left), time-smooth regularization (centre) with d-f) Daubechies 4 and g-i) discrete Meyer and time-sparse regularization (right) with j-l) Daubechies 4 and m-o) discrete Meyer. The images show TEC maps every 30 minutes starting from 22:40 UT of 23 July 2009..... 125
Figure 7.14. Reconstruction obtained with time-smooth regularization with a) Daubechies 4 and c) discrete Meyer and with time-sparse regularization with b) Daubechies 4 and d) discrete Meyer. The regularization parameter was selected differently between the two methods (TS and TP) in order to have a comparable number of basis functions used in the reconstruction..... 127
Figure 7.15. Normalized (a-d) TEC and (e-h) TEC gradients interpolated over time and space with a spline polynomial function for Daubechies 4 (red), discrete Meyer (orange) and MIDAS (green). The CHAMP normalized electron density (black) is also shown. The results were calculated for different days with the time-smooth regularization..... 128
Figure 7.16. Normalized (a-d) TEC and (e-h) TEC gradients interpolated over time and space with a spline polynomial function for Daubechies 4 (red), discrete Meyer (orange) and MIDAS (green). The CHAMP normalized electron density (black) is also shown. The results were calculated for different days with the time-sparse regularization. 128

List of abbreviations

3DVAR	-	3-Dimensional Variational (data assimilation technique)
4DVAR	-	4-Dimensional Variational (data assimilation technique)
AK	-	Averaging Kernel
APL	-	Applied Physics Laboratory
ART	-	Algebraic Reconstruction Technique
BG	-	Backus-Gilbert kernel
CDDIS	-	Crustal dynamics data information system
CHAMP	-	CHAllenging Minisatellite Payload
CIDR	-	Coherent Ionospheric Doppler Receiver
CIT	-	Computerized Ionospheric Tomography
COSMIC	-	Constellation Observing System for Meteorology, Ionosphere, and Climate
CT	-	Computerized Tomography
DA	-	Data Assimilation
DB4	-	Daubechies 4 (wavelet)
DFT	-	Discrete Fourier Transform
DM	-	Discrete Meyer (wavelet)
DoD	-	U.S. Department of Defense
DWT	-	Discrete Wavelet Transform
EOF	-	Empirical Orthonormal Function
EUV	-	Extreme Ultra-Violet
FGP	-	Fast Gradient Projection

FISTA	-	Fast Iterative Shrinkage-Thresholding Algorithm
GAIM@USC/JPL	-	Global Assimilative Ionospheric Model (by USC/JPL)
GAIM@USU	-	Global Assimilation of Ionospheric Measurements (by USU)
GALILEO	-	Named after the Italian astronomer Galileo Galilei, it is European Global Satellite Navigation System.
GLONASS	-	GLObal Navigation Satellite System or Globalnaya Navigazionnaya Sputnikovaya Sistema
GPS	-	Global Positioning System
HA	-	Haar (wavelet)
HRCIT	-	High Resolution Computerised Ionospheric Tomography
IDA4D	-	Ionospheric Data Assimilation 4-Dimensional
IFM	-	Ionospheric Forecast Model
IGRF	-	International Geomagnetic Reference Field
IGS	-	International GNSS Service
IMF	-	Interplanetary Magnetic Field
IRI	-	Ionospheric Reference Model
IRI2012	-	International Reference Ionosphere, version 2012
IS	-	Incoherent Scatter (radar)
ISTA	-	Iterative Shrinkage-Thresholding Algorithm
JHU	-	Johns Hopkins University
JPL	-	Jet Propulsory Laboratory
MART	-	Multiplicative Algebraic Reconstruction Technique
MIDAS	-	Multi-Instrument Data Analysis System
MRA	-	Multi-Resolution Analysis
MURI	-	Multidisciplinary University Research Initiatives
NmF2	-	Electron density peak in the F2 layer

NNSS	-	Navy Navigation Satellite System
NRL	-	Naval Research Laboratories
PPP	-	Precise Point Positioning
RLS	-	Regularised Least Square (algorithm)
RMS	-	Root Mean Square (error)
SIRT	-	Simultaneous Iterative Reconstruction Technique
SM	-	Standard MIDAS
SSBN	-	Nuclear-Powered Ballistic Missile Submarine
sTEC	-	Slant Total Electron Content
TEC	-	Total Electron Content
TECU	-	Total Electron Content Unit
TID	-	Travel Ionospheric Disturbance
TSVD	-	Truncated Singular Value Decomposition
TV	-	Total Variation
UHF	-	Ultra High Frequency
UNAVCO	-	University Navstar Consortium
USC	-	University of Southern California
USU	-	Utah State University
VHF	-	Very High Frequency
VTEC	-	Vertical Total Electron Content

Notations and symbols

μ - Real refractive index

c - Speed of light

v - Propagation speed (radio-wave)

η - Complex refractive index

χ - Imaginary refractive index

μ_{gr} - (Real) group refractive index

μ_{ph} - (Real) phase refractive index

λ_0 - Characteristic wavelength in the vacuum

λ - Signal wavelength

ω - Angular frequency

f_N - Plasma frequency

u - Packet velocity

$\boldsymbol{k}$ - Wave vector

P_{gr} - Group path

P_{ph} - Phase path

P_S - Real signal path

d_I - Ionospheric delay

P - Pseudo-range

Φ - Carrier Phase

ρ - True range

dt - Receiver clock bias

dT - Satellite clock bias

d_T - Tropospheric error

d_M - Multipath error

ϵ_p^r - Receiver hardware dispersive error (pseudo-range)

ϵ_p^s - Satellite hardware dispersive error (pseudo-range)

ϵ_Φ^r - Receiver hardware dispersive error (carrier phase)

ϵ_Φ^s - Satellite hardware dispersive error (carrier phase)

N - Integer ambiguity

P_{L1} - Pseudo-range at L1 frequency

P_{L2} - Pseudo-range at L2 frequency

Φ_{L1} - Carrier phase at L1 frequency

Φ_{L2} - Carrier phase at L2 frequency

P_I - Double differenced pseudo-range

Φ_I - Double differenced carrier phase

b_P^r - Receiver dispersive component residuals (pseudo-range)

b_P^s - Satellite dispersive component residuals (pseudo-range)

b_Φ^r - Receiver dispersive component residuals (carrier phase)

b_Φ^s - Satellite dispersive component residuals (carrier phase)

μ_I - Phase refractive index in the ionosphere

f - Signal frequency

X - Electron frequency

Y - Magnetic field

Z - Collision frequency

$Y_{T,L}$ - Transversal (Longitudinal) magnetic field

n - Electron density

ε_0 - Permittivity of free space

m_e - Mass of the electron

k - Constant equals to 80.6 $\mathrm{m^3H_z^2}$

TEC - Total Electron Content

σ - Radar cross section

r_e - Effective radius of the electron

e - Charge of the electron

γ - Radar backscattering angle

h - Altitude

Δf - Half-power bandwidth

T_e - Electron temperature

κ - Boltzmann constant

D - Debye length

Δf - Doppler frequency shift

V - Velocity of ion-acoustic or electro-acoustic wave

$F(\Lambda)$ - Frequency of ion-acoustic or electro-acoustic wave

T_i - Ion temperature

$\boldsymbol{z}$ - Receiver measurement vector

$\mathbf{A}$ - Geometric matrix

$\boldsymbol{n}$ - Electron density vector

ϱ, α - Regularization parameters for ℓ_2 and ℓ_1 norm

Q - Observed data domain

L - Electron density vector length

L_V - Vertical dimension of the ionospheric grid

L_{H_x} - Horizontal (latitude) dimension of the ionospheric grid

L_{H_y} - Horizontal (longitude) dimension of the ionospheric grid

M - Observation vector length

$\mathbb{R}$ - Set of real numbers

$\hat{\boldsymbol{z}}$ - Projected observation vector through forward problem

$\widehat{\boldsymbol{n}}$ - Estimated electron density vector trough inversion

$\mathbf{W}_{\mathrm{m}}$ - Weighted matrix in Tikhonov regularization

$\boldsymbol{n_0}$ - A-priori information

$\delta\widehat{\boldsymbol{n}}$ - Residual estimated electron density

$\mathbf{R_m}$ - Model resolution matrix

$\mathbf{R_d}$ - Data resolution matrix

$\mathbf{K}$ - Basis function matrix

$\mathbf{K_V}$ - Vertical basis function matrix

$\mathbf{K_H}$ - Horizontal basis function matrix

V - Number of basis functions

$\boldsymbol{\phi_i}$ - i-th basis function of $\mathbf{K}$

$\boldsymbol{\varphi_i}$ - i-th basis function of $\mathbf{K_H}$

$\boldsymbol{x}$ - Basis functions coefficient vector

$\hat{\boldsymbol{x}}$ - Estimated basis functions coefficient vector through inversion

$y_l^m(\theta, \phi)$ - Spherical harmonic

P_l^m - Legendre polynomial

K_l^m - Normalized constant for spherical harmonics

$\psi(t)$ - Wavelet mother functions

$\Psi_{a,b}$ - Wavelet function with dilation a and translation b

V_i - i-th scaling function subspace

W_i - i-th wavelet subspace

$\mathbf{B}$ - Projection matrix of observation offsets

$\boldsymbol{b}$ - Observation offset vector

$\widehat{\boldsymbol{b}}$ - Estimated observation offset vector

$\mathcal{P}(.)$ - Generic regularization term

$\mathbf{C}$ - Calibration matrix

$\mathbf{P}$ - Regularization matrix

η_γ - Shrinkage function

Chapter 1
Introduction

The ionosphere is an ionized medium that is the result of the interaction between the Earth's atmosphere and the Sun's activity, and extends from about 70 km up to 500 km in altitude. A broad spectrum of radiation is emitted continuously from the Sun and is absorbed in the atmosphere where the plasma originates. Plasma is a fundamental state of the matter and is populated by ions and electrons. The Sun emits also charged particles in the form of the solar wind. Those particles reach the Earth's magnetic field where they travel along the magnetic field lines. Those particles are then injected directly into the atmosphere. This represents a secondary source of ionization denoted as particle precipitation. Similarly to clouds, the electrons form structures in the ionosphere, which evolve continuously in time and space, according to season and the Sun's activity.

The study of the ionosphere has become very important because of its interactions with radio-wave signals. Services affected by ionospheric phenomena are broad and include satellite operations, aviation (communication and navigation), electric power distribution and Trans-Atlantic radio communications [*Hargreaves*, 1995]. The vulnerability and disruption of those services depend on ionospheric activity and the possibility to estimate its behaviour. However, the prediction of the evolution of the ionosphere is very difficult and often almost impossible to calculate.

Computerized Ionospheric tomography (CIT) is a technique that allows the state of the ionosphere to be estimated and studied in terms of electron density and to produce 4-Dimensional electron density maps. Techniques that implement CIT algorithms rely on measurements of integrated electron density (or Total Electron Content, TEC), which are generally obtained from ground or satellite-based receivers that measure the signal from GNSS satellites. The measurements are calculated from the delay of the received GNSS signal and converted to TEC. Contrary to other problems like, for example, medical tomography, CIT is very challenging. The difficulties are caused by the uneven distribution of data and limited-angle measurements. Their coverage, in fact, tends to be sparse and receivers can also occur in clusters of different size. Those limitations make the ability to resolve the horizontal structure of the ionosphere through CIT challenging. For those reasons, regularization techniques need to be employed in order to guarantee a stable solution.

A common regularization in CIT applies a smoothing factor in time and space. This produces reasonable results but can sometimes lose the information of the structures in the ionosphere, which could appear smoothed and less defined. This can be a problem since sharp TEC gradients can cause disruption to the GNSS signal. An erroneous solution would be to run a high resolution reconstruction. In fact, this would make the CIT problem highly undetermined and the reconstruction unreliable. Because the structures in the ionosphere occur in different positions and with different scale/sizes, better performances could be achieved if the CIT algorithm would be implemented in view of those structures instead of smoothness, which still remains an important parameter in regions where data are not available. Furthermore, the possibility to adapt the solution according to the data coverage would be of particular interest. In fact, small-scale structures can be potentially resolved in regions with good data coverage.

This thesis develops, applies and tests new methods based on wavelet decomposition, which are compared with more standard ones (generally based on a smooth regularization term or spherical harmonics) by using simulated and real data. The thesis was developed on the basis of three main concepts:

- *Wavelets*. The ionospheric structures appear at different scale and positions. This supported the idea of using wavelet basis function as they are also defined in terms of scale and position.
- *Multi-Resolution Analysis (MRA)*. The decomposition of the ionosphere through wavelet basis functions is better exploited with MRA. It consists on a multiscale approximation of the structures that can be resolved with CIT. Because of the problem of lack of data, it is not possible to have a good definition of the ionospheric structures everywhere. Therefore, it is important that where data are limited, the solution provides a smooth reconstruction. On the other hand, where data are available the solution should attempt to reconstruct the real scale of the structure.
- *Sparse Regularization*. An interesting property of wavelets is given by their ability to represent a signal (in this case the structures of the ionosphere) in a compact or sparse way. It means that the information about the state of the ionosphere can be reconstructed through CIT in terms of just a few coefficients. This implies that a better distinction can be made between the actual information and noisy artefacts that could corrupt ionospheric reconstructions.

This wavelet-based approach is new in CIT and, therefore, has never been applied before. The objectives of this thesis comprise the investigation of new methods that could overcome the limitations imposed by data coverage and limitations imposed by smoothing methods. In particular, the objectives of the thesis can be summarised as:

- Replacing spherical harmonics with wavelets to accommodate MRA horizontally.
- Implementing a 3-Dimensional method based on sparse regularization, wavelets and MRA.
- Extending the sparse regularization method to a 4-Dimensional one that includes the time-dependent component of the ionosphere.
- Testing the performance of discrete Meyer and Daubechies 4 wavelet families.
- Assessing the ability of the methods to resolve different scale structures of the ionosphere.

Below, a brief summary of the content of each chapter of the thesis is described. Chapter 2 provides the background knowledge about the ionosphere and its behaviour. The interaction of the electromagnetic wave is also described, by focusing on the Global Navigation Satellite System (GNSS) signals. Then, the derivation of TEC from the GNSS signal and the origin of the measurements biases are explained. The same chapter includes a description of the main instruments that were used to validate the CIT reconstructions with real data.

Chapter 3 contains the description of the evolution of CIT starting from the first experiments by Austen [*J. R. Austen et al.*, 1986]. It also illustrates the state of the art in CIT and a list of the different algorithms that were implemented by various Universities or research centres during past years.

The definition of the mathematical problem for CIT is given in Chapter 4. It describes the mathematical notation and a general solution for tomography based on Tikhonov regularization. It also defines the vertical and horizontal basis functions that were used to describe the vertical and horizontal variation of the ionosphere. Then, it focuses on the definition of wavelets and MRA.

The new methods implemented in this thesis are explained in Chapter 5. It revises different regularization terms including the sparse regularization that was implemented with the Fast Iterative Shrinkage-Thresholding Algorithm (FISTA) [*Amir Beck and Marc Teboulle*, 2009]. Then, the Total Variation (TV) algorithm [*A. Beck and M. Teboulle*, 2009] is also explained. The latter is used to reduce the artefacts in high resolution CIT reconstruction by preserving the edges of the structures.

The first results on sparse regularization using wavelets are presented in Chapter 6. The results compare two wavelets (discrete Meyer and Daubechies 4) with spherical harmonics, which are implemented with a standard version of Tikhonov regularization. The analysis is performed in terms of robustness of the methods and ability to recover the measurement biases and structures by using simulated data. The methods are intended for 3-dimensional

reconstruction of the ionosphere and some results are also shown using real data. Furthermore, the chapter points out the difficulties obtained when sparse regularization is used for high resolution tomography and illustrates the advantages of the TV algorithm.

The time-dependency of the ionosphere is described in Chapter 7. The sparse regularization method is extended to a 4-dimensional reconstruction of the ionosphere, where the fourth dimension is time. The chapter compares MIDAS with two implementations of the sparse regularization. The first one consists in using three-dimensional wavelets (in space and time), whilst the second uses two-dimensional wavelets in space and an additional smoothing term for the time dimension. The analysis is performed using real data.

Finally, conclusions and future works are described in Chapter 8.

The Multi-Instrument Data Analysis System (MIDAS) was used as the framework within which the methods and experiments described in this thesis were developed. MIDAS is a software package developed at the University of Bath for tomography. Although it has been used mainly for ionospheric purposes, it includes capabilities that widen to other fields like medical tomography.

Chapter 2
Ionosphere: principles and measurements

Introduction

This chapter outlines the principles of the ionosphere and its interaction with radio-waves. Some well-established instruments that can be used to measure the ionosphere are also described

In particular Section 2.1 gives a simple description of the main processes that generate the ionosphere. The structure of the ionosphere and how it interacts with radio waves are described in Section 2.2. Section 2.3 introduces the techniques that can be used for sounding the ionosphere, which include details of the Global Positioning System (GPS) signal that provides the main measurements used in this thesis. Those measurements will then be used in Chapter 3 to describe the imaging of the ionosphere by means of tomographic algorithms. Conclusions are in Section 2.4.

2.1 The Ionosphere

The first evidence for the existence of a conductive medium in the upper region of the atmosphere came with Marconi's experiment in 1901 where a radio signal was transmitted from Cornwall (UK) to Newfoundland (Canada). The existence of this conductive medium was suggested by Kennelly and, later, by Heaviside. They argued that because of the Earth's curvature, the radio signal must have been reflected from another ionized medium. In 1926, R. Watson-Watt coined the name ionosphere and since 1936 it came into common use [*Hargreaves*, 1995].

The sun emits radiation in a wide wavelength range that penetrates the atmosphere and is absorbed by the neutral species of the atmosphere with different intensities according to altitude, latitude and longitude. The main radiations involved in the ionosphere are the X-Ray and Extreme Ultra-Violet (EUV) bands. They are the principal source of ionization that shapes the ionosphere in different layers (where different wavelengths are absorbed). Another source of ionization is given by energetic particles. The sun continuously releases those particles with the solar wind. The interaction of the solar wind with the Earth's geomagnetic field can allow some of the particles to penetrate into the atmosphere, causing a process known as impact ionization. Although their impact in the formation of the ionosphere is minimal at middle latitudes, they can become the main source of ionization at high latitude. In

fact, at the poles the magnetic field lines converge perpendicularly to the Earth driving those particles directly in the ionosphere [*Barclay*, 2003].

The presence in the ionosphere of those neutral species is important in the formation of the ions. The Chapman model [*Chapman*, 1931] is a good example to explain the basic ionosphere formation. This model explains the daily variation in the ionosphere given a quiet solar activity at low and middle latitudes. It considers the ionization rate dependent on the neutral densities and radiation intensity coming from the Sun (see Figure 2.1). Therefore, the generation of ions is associated to only the photoionization process by means of the EUV radiation.

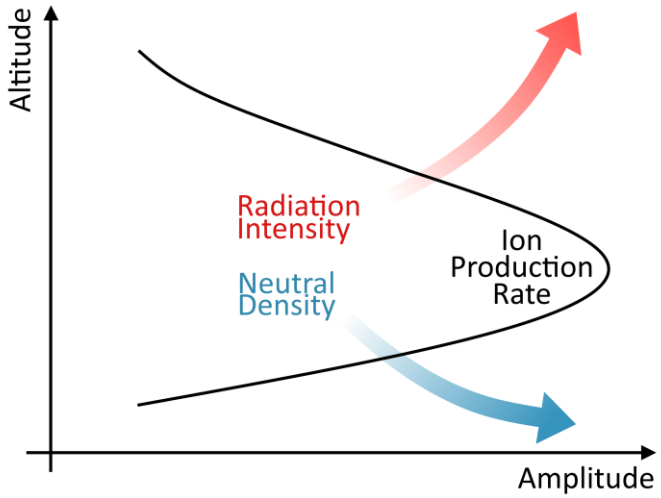

Figure 2.1. The ion production rate can be seen as the combination of the radiation intensity (from the sun) and the neutral densities of the species in the atmosphere.

The actual processes that drive the ionospheric behaviour are actually more complicated than the ones described in the Chapman model. Nevertheless, it remains an important tool to explain the basic theory of the ionosphere. The ions and electrons formed from the solar radiation react and recombine with the other gaseous (neutral) species to produce further ions [*Hargreaves*, 1995]. These photochemical processes produce a dynamical equilibrium that depends on the production and loss rate of the electrons and ions. Three main processes are involved [*Barclay*, 2003]:

- *Photoionization*. The neutral gases are ionized by the EUV photons producing ions plus electrons.
- *Collision*. The collision between ions and neutral molecules produce an energy which ionizes the molecules through a charge exchange reaction.
- *Recombination*. Ions and electrons recombine.

The importance of these processes and the effect of their combination varies with altitude. This generates the typical layered structure of the ionosphere. Three main layers can be recognized on an electron density profile:

- D-Layer (60-90 km). The neutral atmosphere in this layer is denser than in the other layers, and the electron collision frequency is therefore higher and consequently the recombination.
- E-Layer (105-160 km). This is, with the F1-Layer, well described by the Chapman model, where the photochemical processes of production and loss get involved.
- F-Layer (160-300 km). It is divided in two sub-layers, F1 (160-180 km) and F2 (180-300 km). The F1 still follows the Chapman model quite well in a quiet period (i.e. it can be well explained with the photochemical reactions), but the F2 does not generally. It is the layer with the highest electron density concentration and where the electron peak is located (F2-peak). The strong photoionization and the low neutral density make the loss process slower prolonging the life of the electrons.

During the day, the electron density reaches its maximum and decreases to its minimum density at night and just before sunrise. The amount of electron density, and the relative day-night fluctuation, is also dependent on the solar activity, which can be indicated in terms of sun-spot numbers. Sunspots are measured from the Sun surface and look like dark spots due to their lower temperature. During maximum solar activity the electron density tends to be generally higher than during minimum solar activity. Furthermore, during the night the E-Layer, which is present when the ionization process caused by the sun is stronger, disappears and the F2 electron density peak (NmF2) rises up. The season also influences ionospheric activity, making the NmF2 generally higher during the equinoxes and lower in the summer [*Hargreaves*, 1995].

The geographical variation of the ionosphere is illustrated in Figure 2.2. The image shows the vertical Total Electron Density (VTEC), which is defined as the integrated electron content along the vertical in each point of the map. The values showed in the maps are in TEC units (TECU); 1 TECU corresponds to the integrated electron content in 1 metre square cross-section along the ray path and is measured in 10^{16} electrons per metres squared.

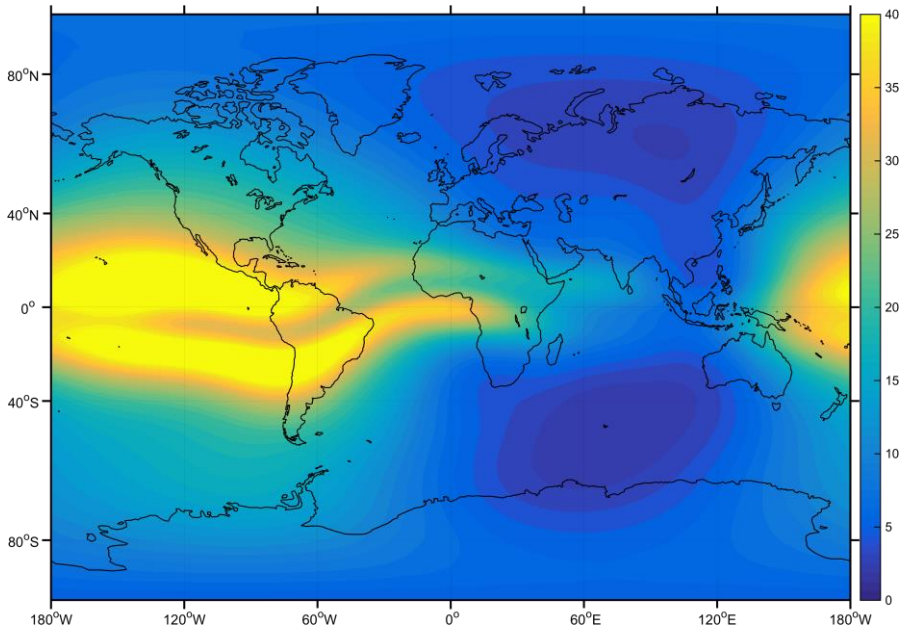

Figure 2.2. Run of the IRI2012 empirical model on the day of 5 October 2014 at 21:00 UT. The map shows the TEC obtained from the empirical model. Values are in TECU.

The image was obtained by a run of the International reference Ionosphere (version 2012) empirical model for the day of 5 October 2014 at 21:00 UT. It can be noticed the presence of large-scale features like enhancements, depletions and troughs in terms of TEC variation. Sometimes smaller scale structures can originate. Those structures can cause events in the signal like diffraction/interference. These events can occur when the electromagnetic wave travels in proximity of the edges of an irregularity. The scale of the irregularity can vary from 100 to 1000 metres and their effect can persist for few seconds or several minutes [*Davies*, 1990].

2.1.1 Principles of signal propagation in the ionosphere

A radio signal transmitted through the ionosphere experiences episodes of refraction/diffraction. Those phenomena are the cause of distortion and retardation in the signal. This section will outline the principles of a signal propagating in a medium such as the ionosphere. This section is based on the references found in [*Davies*, 1990]

A signal propagating in a medium will travel with a propagation speed v which depends on the medium itself and can be expressed in terms of its refractive index μ as [*Kaplan and Hegarty*, 2006]

$$\mu = \frac{c}{v} \tag{2.1}$$

where c is the speed of light. Thus, a variation of the refractive index causes a delay in the signal with respect to vacuum where the velocity is the speed of light and the refractive index

is unitary. In addition to the retardation, the refractive index can also cause an attenuation, which is considered by introducing an imaginary refractive index part χ

$$\eta = \mu - i\chi \tag{2.2}$$

The refractive index of the atmosphere changes with the altitude and the effects on the signal can be identified as coming mainly from the troposphere and ionosphere affecting the signal in different ways. The troposphere introduces a non-dispersive delay (i.e. the refractive index doesn't depend on the frequency) and takes into account variables like the pressure, temperature and water vapour. By contrast, the ionosphere is a charged medium which affects the signal with effects like scintillation and retardation. It is also a dispersive medium causing different frequencies to travel at a different speed [*Barclay*, 2003]. It is the medium that introduces the biggest delay and perturbation at Very High Frequency (VHF) and Ultra High Frequency (UHF). Furthermore it is well known that, in order to pass through the ionosphere, a signal needs to have a carrier frequency higher than a threshold. This threshold depends on the launch angle and the electron concentration at the F2 peak. The free electrons absorb the energy of the signal and, if the density of the electrons is relatively low (i.e. there are not many collisions between neutrals and charged particles, namely collision frequency), they release that energy at the same frequency and the signal still propagates. Where, instead, the density is very high, the increased number of collisions makes the electrons lose more energy, compromising the propagation (e.g. causing a delay in the received signal from the satellite). The energy released in both cases is not directional, but is spread in different directions causing attenuation and multipath.

The refractive index is divided in phase μ_{ph} and group μ_{gr} refractive indices. This is due to the different effects that the ionosphere (as a medium) impinges on the phase and envelope of the signal. When the signal crosses the ionosphere the phase velocity of the signal increases [*Davies*, 1990] but, because the following relationship holds

$$\frac{v_{ph}}{\lambda} = \frac{c}{\lambda_0} \tag{2.3}$$

where c is the speed of light, λ_0 the characteristic wavelength in the vacuum and λ the wavelength of the signal, it means that the wavelength of the signal becomes longer. The fact that the phase velocity increases can be seen as a fast advancement of the phase and this effect is generally referred as *phase advance*.

The group refractive index μ_{gr}, instead, is responsible for the propagation delay of the signal intended as a wave packet (i.e. made up of different frequencies). The relationship between the phase and group refractive index is given by

$$\mu_{gr} = \frac{d(\omega\mu_{ph})}{d\omega} \tag{2.4}$$

where ω is the angular frequency associated with λ. By knowing the velocities and the refractive indices it is possible to calculate the group and phase paths (in distance unit) [*Davies*, 1990]. The group path P_{gr} represents the distance or spatial delay (with the ionospheric delay included) of the wave packet traveling along the ray path S and is given by

$$P_{gr} = \int_S \mu_{gr} \cos\alpha \, ds \tag{2.5}$$

In the same way it is possible to define the phase path P_{ph} as the distance in free space that can produce the same phase variation as the one obtained along the ray path S, i.e.

$$P_{ph} = \int_S \mu_{ph} \cos\alpha \, ds \tag{2.6}$$

The angle α represents the angle between the ray and the direction of phase propagation. By considering the angle α very small (i.e. isotropic medium [*Davies*, 1990]), the cosine term in Equations (2.5) and (2.6) can be neglected.

The real signal path P_S is simply calculated as

$$P_S = \int_S ds \tag{2.7}$$

In the case of a signal propagating through the ionosphere, the extra-delay d_I introduced by the medium can be obtained by subtracting the real path with the group or phase path

$$d_I = P_S - P = \int_S (1 - \mu) ds \tag{2.8}$$

where P represents both the group or phase path and, in the same way, μ the equivalent group and phase refractive index for the ionosphere. The distance d_I has a different interpretation from the real path P_S and it is associated with the expected distance that can explain the traveling time (as measured from phase or envelope).

2.2 GPS, from the origins to the application

The beginning of experimental studies of radio signal propagation through the ionosphere and orbit determination by means of Doppler measurements came with the launch of Sputnik I on 4 October 1957 by the Soviet Union. The launch of Sputnik defined a new era of the space race in particular between Soviet Union and United States. Despite the concerns of the period

about the implications in the Cold War, this first artificial Earth satellite ever orbiting in space led to important scientific and technological discoveries and advancements that provided the basis for the development of the current positioning systems like the U.S. GPS, the European system GALILEO, and the Russian GLObal Navigation Satellite System (GLONASS), which are generally classified under the common name of Global Navigation Satellite Systems (GNSSs).

The GPS was developed by the U.S. Department of Defense (DoD) and became operational in June of 1995 with 24 satellites then extended to 32. The GPS system provides positioning and timing services for civilian and military applications.

The GPS provides the fundamental measurement for the tomographic methods used for the imaging of the ionosphere in this thesis. This Section describes Transit, the predecessor of GPS. Transit data are used in conjunction with GPS data in Chapter 5. Then, the GPS signal structure and how it can be used to probe the ionosphere are also described. Other instruments (used in this thesis) that sound the ionosphere are also illustrated.

2.2.1 Transit, the ancestor of GPS

The Navy Navigation Satellite System (NNSS), also known informally as Transit, was an all-weather global navigation system designed by the Applied Physics Laboratory (APL) including user equipment and ground control system [*Danchik*, 1998]. Its operational use lasted for almost 32 years and provided the most reliable and precise system for positioning at that time. The system was continuously improved during the years and all the achievements and results of the mission were essential to the space science and technology of the current years.

Frank T. McClure (APL) in the spring of 1958 started the transit program on the basis of the Doppler tracking discoveries by William H. Guier and George C. Weiffenbach at Johns Hopkins University (JHU) and APL, i.e. that the Sputnik orbit could be determined by the measurement of the Doppler (by chance the only measurement they had available) of the received signal [*Guier and Weiffenbach*, 1997]. This led to the definition of the navigation problem: if the position of the satellite were accurately known, then Doppler from the signal could tell an observer on the ground his unknown position[1].

The origin of the development of Transit came from the need to fulfil the requirements of an accurate and precise position for the navigation system of Nuclear-Powered Ballistic Missile Submarines (SSBNs) of the U.S. Navy. The position was required to be available several

[1] From https://directory.eoportal.org/web/eoportal/satellite-missions/t/transit

times during the day and with an accuracy of 0.1 nmi (about 185.2 metres). Furthermore the user was required to use passive receivers and to obtain data in real time [*Danchik*, 1998].

The Transit satellite system was made public in 1967 by the Vice President Hubert Humphrey and released for commercial use by ships and friendly nations. However, the system became widely used after the development of a low-cost receiver made in the early 1970.

2.2.2 GPS signal propagation

A GPS receiver provides a measure of the time the signal transmitted from a satellite k needs to travel to a receiver p. The GPS system has two fully operative signals which can be used, at two frequencies indicated as L1 (1.5 GHz) and L2 (1.2 GHz). Each satellite transmits a well-known digital code (pseudorandom noise code, PRN) whose replica, at the receiver, is used to estimate the travel time. This time measurement is called pseudorange P or carrier phase Φ according to whether the code or the phase of the carrier signal is used. By crossing the atmosphere the signal is distorted and delayed by a series of effects which, at one particular frequency, can be modelled according to the following equations

$$P = \rho + c(dt - dT) + d_I + d_T + d_M + \epsilon_p^r + \epsilon_p^s \tag{2.9}$$

$$\Phi = \rho + c(dt - dT) - d_I + d_T + d_M + \lambda N + \epsilon_\Phi^r + \epsilon_\Phi^s \tag{2.10}$$

where ρ is the true range, which only depends on the geometry of the problem, dt and dT are the receiver and satellite clock biases, d_I and d_T are the errors introduced by the ionosphere and troposphere, d_M is the error due to the multipath, and ϵ_p^r and ϵ_p^s are the hardware dispersive components (which depend on the frequency) in the pseudorange measurement including the noise. In the same way, ϵ_Φ^r and ϵ_Φ^s corresponds to the hardware dispersive components for the carrier phase. The term N, instead, is the integer ambiguity in the phase cycle measurement. N introduces a delay (in distance unit) that is proportional to the wavelength λ of the signal.

The two different signs that appear in the d_I terms in the above equations are due to the fact that the pseudorange suffers a positive group-delay, while in the carrier phase this effect, which has approximately the same magnitude, is negative and it is referred as phase-advance.

Because of the very efficient demodulation technique, it is possible to handle very weak GPS signals, even when they are below the noise level. When the signal is detected the code is generally perfectly reconstructed. Furthermore, the effects due to the frequency bias and spread are also very well compensated by a tracking loop filter. For this reason the next section will focus on the delay only while all the attenuation effects will not be considered.

It will be seen in the next section that by double differencing the GPS signal at the L1 and L2 frequencies it is possible to isolate the ionospheric term. This term is important as it can be related directly to the electron density of the ionosphere. In ionospheric tomography this term is the main observable that allows reconstructing three dimensional images of the ionosphere.

2.2.3 TEC measurements from GPS

From the ionospheric point of view the electron density is the main observable that can be exploited to describe the ionospheric effects on the signal. The GPS signal can be used to retrieve the estimation of the slant Total Electron Content (sTEC) between the satellite and receiver, which is defined as the integrated electron content along the ray path and is measured in TECU.

Due to the frequency-dependence of the ionospheric effects, it is possible to estimate the term d_I of (2.9) and (2.10) by differencing the GPS signal at the two frequencies L1 and L2. This allows having an estimation of sTEC (along the satellite-receiver path) that can be used to compensate the ionospheric delay in case of positioning or timing applications. In other applications like ionospheric tomography, this estimation is used as input data from which the three-dimensional electron density can be computed (see Chapter 3 for a more comprehensive description of ionospheric tomography).

The non-dispersive components can be cancelled by taking the difference at the two frequencies L1 and L2 of Equation (2.9) and Equation (2.10), obtaining [*Mannucci et al.*, 1999]

$$P_I = P_{L2} - P_{L1} = \left(d_{I,L2} - d_{I,L1}\right) + b_P^r + b_P^s \tag{2.11}$$

$$\Phi_I = \Phi_{L1} - \Phi_{L2} = \left(d_{I,L2} - d_{I,L1}\right) + b_\Phi^r + b_\Phi^s \tag{2.12}$$

where b_P^r, b_P^s, b_Φ^r and b_Φ^s are the dispersive component residuals related to receiver (r) and satellite (s), which cannot be removed by differencing. The integer ambiguities $\lambda_1 N_1$ and $\lambda_2 N_2$ have been included in b_Φ^s. The dispersive multipath-residual biases in P_I contribute as a significant noise component [*Jakowski*, 1996] which makes the sTEC estimations less accurate. A more accurate estimation of the ionospheric term can be done from the carrier phase. Unfortunately this has the disadvantage that a calibration is required as will be seen later.

The refractive index in the ionosphere can be modelled through the Appleton equation. It models the phase refractive index in the ionosphere μ_I depending on the collision frequency (between the electrons and neutrals) and Earth's magnetic field [*Barclay*, 2003] and the electron content

$$\mu_I^2 = \left(\mu_{ph} - i\chi_{ph}\right)^2 = \\ = 1 - \frac{X}{1 - iZ - \frac{Y_T^2}{2(1 - X - iZ)} \pm \sqrt{\frac{Y_T^4}{4(1 - X - iZ)^2} + Y_Z^2}} \tag{2.13}$$

where the real part of the refractive index μ_{ph} takes into account the retardation and the bending of the signal, while the imaginary part, χ_{ph}, accounts for the absorption. X, Y, and Z describe respectively the electron density, geomagnetic field and collision frequency effects and the subscripts T and Z are depicting the transverse and longitudinal components of the geomagnetic field with reference to the wave normal direction. The equation above is particularly difficult to use, so for practical use further assumptions/simplifications are used. If the influence of the geomagnetic field is ignored and the collision frequency is neglected, Equation (2.13) becomes

$$\mu_{ph}{}^2 = 1 - X = 1 - \frac{e^2}{4\pi^2 \varepsilon_0 m_e f^2} n = 1 - k\frac{n}{f^2} \tag{2.14}$$

where f is the signal frequency, n is the electron density, ε_0 is the permittivity of free space, m_e is the mass of the electron, e is the charge of the electron and k is defined as

$$k = \frac{e^2}{4\pi^2 \varepsilon_0 m_e} = 80.6 \tag{2.15}$$

where the units are $m^3 H_Z^2$. It can be noted from Equation (2.14) that the phase refractive index must be between 0 and 1. Therefore, the condition $k\,n/f^2 > 1$ must hold, which corresponds to $f > f_N$ where the plasma frequency f_N is defined as

$$f_N = \frac{ne^2}{4\pi \varepsilon_0 m_e} \tag{2.16}$$

The case $f == f_N$ corresponds to the vertical reflection of a electromagnetic wave.

Finally, the phase refractive index can be obtained from the phase refractive index with the following approximation based on the binomial expansion series can be used

$$\mu_{ph} = \sqrt{1 - k\frac{n}{f^2}} \cong 1 - k\frac{n}{2f^2} \tag{2.17}$$

Equation (2.17) relates the electron density n to the phase refractive index μ_{ph} which will be used in Equation (2.11) and Equation (2.12). According to the simplification this approximation is valid for a solar minimum period without disturbed conditions and is

especially suitable for middle latitudes. Higher order effects can be included and become important for high-precision applications (e.g. Precise Point Positioning, PPP) but this is not part of this dissertation. A review of higher order ionospheric effects can be found in [*Petrie et al.*, 2011].

Therefore, according to Equation (2.4) the group refractive index can be obtained from the phase refractive index as

$$\mu_{gr} \cong \frac{d}{d\omega}\left(\omega\left(1 - k\frac{n}{2f^2}\right)\right) = 1 + k\frac{n}{2f^2} \tag{2.18}$$

The ionospheric delay can now be calculated as (Equation (2.8))

$$d_I = \int_{TX}^{RX} (1 - \mu_{gr})ds = \; k\frac{TEC}{2f^2} = \frac{40.3}{f^2}TEC = \; \frac{1}{f^2}I \tag{2.19}$$

where $I = 40.3\,TEC$. This puts a direct proportionality between the observable d_I and the total electron content (TEC), obtained from Equation (2.18) by integrating the electron content n_e between the satellite TX and the receiver RX along the ray path s as

$$TEC = \int_{TX}^{RX} n\,ds \tag{2.20}$$

The ionospheric delay in the carrier phase can be calculated in the same way. It can be noticed that, with the above approximations, the only difference lies in the sign, which is negative for the carrier phase. For this reason the phase term d_I is usually referred as phase advance.

From the Equations (2.11) and (2.12) it is possible to write [*Mannucci et al.*, 1999]

$$P_I = P_{L2} - P_{L1} = I\left({}^{1}/_{f_{L2}^2} - {}^{1}/_{f_{L1}^2}\right) + b_P^r + b_P^s \tag{2.21}$$

$$\Phi_{\mathrm{I}} = \Phi_{L1} - \Phi_{\mathrm{L2}} = \\ = I\left({}^{1}/_{f_{L2}^2} - {}^{1}/_{f_{L1}^2}\right) + (\lambda_{L1}N_{L1} - \lambda_{L2}\,N_{L2}) + b_\Phi^r + b_\Phi^s \tag{2.22}$$

The two equations (2.21) and (2.22) measure the same ionospheric delay, but, because the integer phase ambiguities $\lambda_{L1}N_{L1}$ and $\lambda_{L2}\,N_{L2}$ cannot be compensated in the carrier phase, they still remain unknown in the equation. For this reason the differential carrier phase doesn't provide an absolute value for the TEC like the differential pseudorange could potentially do. They are however used together to calibrate the measurement. However, the pseudorange tends to be noisier while the carrier phase is generally more stable especially at low elevation

angles where the multipath effect starts to become important. Instead, observations of differential phase can be generally considered noise free from the point of view of the instrument which inherently smooths over noise, but the measurement arc between a single receiver and satellite is uncalibrated or biased.

The dispersive pseudorange errors can be reduced by smoothing the difference between the pseudorange and the shifted version of the carrier phase (which is cleaner but has an ambiguity term) in a least square sense [*Jakowski et al.*, 2001].

Daily calibration to the differential pseudorange code is necessary to take into account variations of the differential biases in the clocks [*Mannucci et al.*, 1999]. Those solutions to calibrate the GPS-based ionospheric measurements are discussed in different publications (e.g. [*Mannucci et al.*, 1999] and [*Jakowski*, 1996]) and are based on a combination of both P_I and Φ_I. Note that within a short time interval and until the signal loses its lock, the integer ambiguity and the residual terms can be considered constant. This can be used to help calibrate sTEC as described before. However, it is also possible to calibrate the observations by mean of tomographic tools in the inversion process. This has been shown to have advantages [*Dear and Mitchell*, 2006], particularly in cases where the hardware biases vary over time. This calibration method will be discussed in Chapter 5.

2.3 Instruments used for validation

In this thesis two other instruments are used to validate the results, a satellite mounting the Planar Langmuir Probe and an incoherent scatter radar.

The CHAllenging Minisatellite Payload (CHAMP) is a German small satellite mission whose primary scientific purposes were related to gravity field, magnetic field and the atmosphere/ionosphere. Because of its polar low-Earth orbit (between about 300 and 450 km), it is a suitable instrument for studying the F-region of the ionosphere with in situ instrumentation [*Reigber et al.*, 2002].

The Incoherent Scatter (IS) radar has been operated since the beginning of 1960s [*W E Gordon*, 1958]. Its measurements, obtained from the scattering of the transmitted signal through plasma irregularities in the ionosphere, provided important information about both the bottom and top side ionospheric composition. The radar allowed extending the range of measurements that between 1930 and 1960 came mainly from ionosondes and were limited to the bottom-side of the ionosphere only.

2.3.1 CHAMP

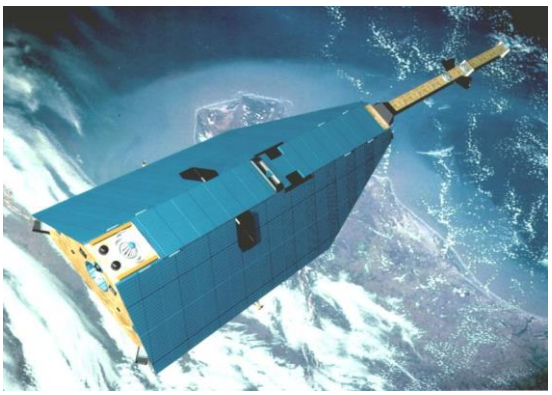

Figure 2.3. CHAMP.

The CHAMP mission was originally proposed by GeoForschungsZentrum (GFZ) and the former German Space Agency DARA (now DLR) with the partnership of the National and Space Administration (NASA), the Centre National des Études Spatiales (CNES), and the Air Force Research Laboratories (AFRL).

The mission started on 15 July 2000 when CHAMP was launched with a Russian COSMOS launch vehicle into an almost circular, near-polar orbit. Its initial altitude was around 450 km and the inclination about 87.3 degrees.

The payloads of the satellite can be grouped according to their mission target:

- *Gravity*. The new generation GPS receiver TRSR-2 (by NASA) for the tracking of low orbit satellites and GPS constellation and high-precision orbit determination, and the electrostatic STAR accelerometer (by CNES), a high-precision three-axes accelerometer for measuring the surface forces accelerations.
- *Magnetics*. The fluxgate magnetometer (by DTU) that together with the laser retro reflector (by GFZ) was used to probe the vector components of the Earth's magnetic field. The overhauser Magnetometer (by LETI) was used to provide an absolute in-flight calibration capability for the fluxgate magnetometer.
- *Atmosphere/Ionosphere*. The same instrumentations for the magnetic and gravity field contributed to the characterisation of the state and dynamics of the neutral atmosphere and ionosphere and the GPS receiver could be used for GPS radio-occultation measurements. In addition, the digital ion drift meter (by AFRL) was used to make in-situ measurements of the ion distribution and its moment within the ionosphere. Quantities like electron density, and temperature can be then derived from its measurements.

The anticipated life of the mission was estimated as 5 years, when CHAMP would have decayed its altitude to 300km. In fact the highly successful mission lasted until 19 September 2010 when it was officially concluded.

The Digital Ion Drift Meter (DIDM) payload has been used in Chapter 6 and Chapter 7 to produce the validation of the ionospheric reconstructions. For this reason it will be briefly described below. The DIDM was developed by Amptek, Bedford, MA USA, for AFRL with the aim to detect the angle of arrival of thermal plasma ions [*Cooke et al.*, 2003] from which the ion distribution and its moments can be retrieved.

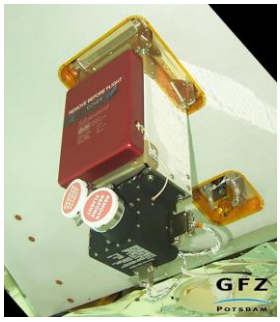

Figure 2.4: Picture of the Digital Ion Drift Meter (DIDM)[2].

The DIDM can operate in two different modes:

- Drift Meter (DM). Its principle of operation consists in a pin-hole camera that images the ion distribution on a Micro-Channel Plate (MCP). The MCP converts the ions into a pulse of about 10^6 electrons. Then, the pulse position is located by means of an anode consisting in a 16x128 pixels array. The distribution of the ions in the velocity space can be determined by accumulating the ions that impact into the plate. Different ions parameters, such as drift angle, temperature and density, can be recovered by matching a tabulated set of expectations.
- Retarding Potential (RPA). It energises the ions that enter through the pin-hole camera to the MCP. Therefore, it determines the energy of the incoming ions as well as their velocity, assuming the ion mass to be that of oxygen.

In addition, the DIDM mounts also the Planar Langmuir Probe (PLP), which provides auxiliary data used to interpret the ion drift measurements and to monitor the spacecraft potential, ion density and electron temperature.

[2] Credits: http://op.gfzpotsdam.de/champ/systems/index_SYSTEMS.html.

Two sensors were originally designed in order to provide simultaneous DM and RPA measurements but also higher measurement rates and redundancy. They were positioned 10 degrees apart in respect to the nominal flight direction. This configuration allowed improving the angular coverage for the horizontal ion drift. Unfortunately, after the first power-up it was noticed there was a failure in one of the sensors and one single sensor was used eventually. This affected the performances and the statistical quality of the data. Because of this, the initial precision for the ion drift measurement of 25-50 m/sec was reduced to 130 m/sec.

2.3.2 Incoherent scatter radar

The physical basis of the incoherent scatter (IS) radar was suggested in 1906 when J.J. Thomson showed that the electron can scatter an incident electromagnetic radiation with an equivalent radar cross section σ equal to [*Thomson*, 1906]

$$\sigma = 4\pi(r_e \sin\gamma)^2 \tag{2.23}$$

where γ is the angle between the direction of the incident wave and the observer direction [*Beynon and Williams*, 1978] and r_e is the effective radius of the electron defined as [*Hargreaves*, 1995]

$$r_e = \frac{e^2}{4\pi\varepsilon_0 m_e c^2} \tag{2.24}$$

where e and m_e are the charge and mass of the electron respectively, ε_0 is the permittivity of free space and c is the speed of light in vacuum.

Unfortunately, the technology available in that period was not sufficient to detect the weak signal originated by the scattering defined by Equation (2.23) and they had to wait until the technological advancements driven by the Second World War to overcome those limitations and to achieve the required instrument sensitivity [*Hargreaves*, 1995]. In 1958 it was suggested that even though free electrons scatter radio waves weakly in an ionized medium, a powerful radar with a large enough antenna could detect the incoherent backscatter from the free electrons in and above the earth's ionosphere [*W E Gordon*, 1958]. Furthermore, both density and temperature of the electrons could be measured as a function of height. In the same paper a simple theory of scattering was also proposed. Given a transmitted signal of wave vector $\boldsymbol{k}$ defined as

$$|\boldsymbol{k}| = \frac{2\pi}{\lambda} \tag{2.25}$$

and supposing the plasma composed of thermal electrons, distributed randomly with a constant average density equivalent to a mean spacing of δ and providing that $|\boldsymbol{k}| \ll 2\pi/\delta$,

then the electrons are totally independent of each other. In the ionosphere, at an altitude of 100-2000 km, the value of δ is always less than 10^{-3} m [*Beynon and Williams*, 1978], which easily satisfies the previous condition of $\lambda \gg \delta$. Due to the random motion of the electrons the scattering would experience random phases. This would make the scattering completely incoherent and suggested that the total scattered power was proportional to $\sigma n_e(h)$, i.e. given the power the electron density $n_e(h)$ at different altitudes h could be calculated. In order to avoid coherent scattering λ should be chosen smaller than the smallest irregularity of the ionosphere. Structures with size smaller than 10 m are known to occur for example in aurora zones, which makes the upper limit for λ equal to 1 metre. Gordon, in the same paper, predicted that due to the random thermal motion of the electrons, the spectrum of the received signal would have experienced a wide range of Doppler shifts, whose half-power bandwidth Δf was equivalent to

$$\Delta f = 4(\kappa T_e / m_e \lambda^2)^{1/2} \tag{2.26}$$

where κ is the Boltzmann constant, T_e the electron temperature, m_e the electron mass and λ the signal wavelength.

The same year, Bowles showed the first results of actual observations of echos using a new high-power transmitted built at Long Branch, Illinois (US). The results were presented in an early unpublished report of the National Bureau of Standards on 18 September 1959 and then published in [*Bowles*, 1961].

The results of the experiment showed a spectrum that was ten times smaller than the one predicted by Gordon. It was clear at this stage that a new theory needed to be developed.

The new theory of scattering is more complicated than a simple incoherent scattering by free electrons and describes a quasi-coherent scattering by thermally induced electron-ion acoustic waves but, for some reasons, the name Incoherent Scatter (IS) radar was maintained. Those and the following comments are from [*Beynon and Williams*, 1978].

The new theory considers the electrostatic interaction between ions and electrons. The ions tend to be attracted towards electrons while their thermal velocities and their mutual repulsion tend to disperse them. The extension of influence of the ions is represented by the Debye length [*Davies*, 1990]

$$D = \left(\frac{\varepsilon_0 k T_e}{e^2 n}\right)^2 \tag{2.27}$$

where ε_0 is the permittivity of free space, κ the Boltzmann's constant, T_e the electron temperature, e the charge of an electron and n the electron density.

The Debye length gives an indication of the region of influence of an ion. If the wavelength of the radar is smaller than D, the scattering can be considered incoherent and depending on the thermal random motion of the free-electrons in the ionosphere. In this case the spectrum would agree to the one predicted by Gordon. For a radar application, the wavelength is usually bigger than D. In this case the scattering is strongly controlled by the surrounding ions. The random thermal motion of the electrons produces pressure gradients that generate longitudinal waves known as *electro-acoustic* (or *plasma*) waves and, because of the electrostatic coupling, also *ion-acoustic* waves. Furthermore, the increase of electron (or ion) density produces an increment in the local net charge of the ionosphere. Thus, additional electrostatic forces are generated that need to be added to the forces due to pressure gradients. The scattered signal that is observed by the radar is, therefore, caused by the fluctuation of those waves.

The ion-acoustic and electro-acoustic waves can travel upward or downward. That causes a Doppler frequency shift equal to

$$\Delta f = \pm 2\frac{V}{\lambda} = \pm\frac{V}{\Lambda} = \pm F(\Lambda) \tag{2.28}$$

Where $\Lambda = \lambda/2$, $F(\Lambda)$ is the frequency of the wave and V the velocity of the wave. Considering the existence of two waves, it follows that the spectrum of the scattered signal will consist of four components (upward and downward) depending on the ion-acoustic $F^+(\Lambda)$ and electro-acoustic $F^-(\Lambda)$ waves

$$F^+(\Lambda) = \frac{1}{\Lambda}\left(\frac{kT_i}{m_i}\left(1+\frac{T_e}{T_i}\right)\right)^{0.5} \tag{2.29}$$

$$F^-(\Lambda) = f_p\left(1+\frac{12\pi^2 D^2}{\Lambda^2}\right)^{0.5} \tag{2.30}$$

where T_i and T_e are the ion and electron temperature respectively, m_i is the ion mass and f_p is the plasma frequency.

The overall spectrum (Figure 2.5) is composed of a broad range of frequencies and its shape is controlled by an attenuation mechanism known as the Landau damping mechanism. It describes the transfer of energy that occurs when the charged particles move in the same direction as the wave. It causes attenuation or enhancement of the wave and acceleration or

deceleration of the particle according to whatever the wave is respectively faster or slower than the particle.

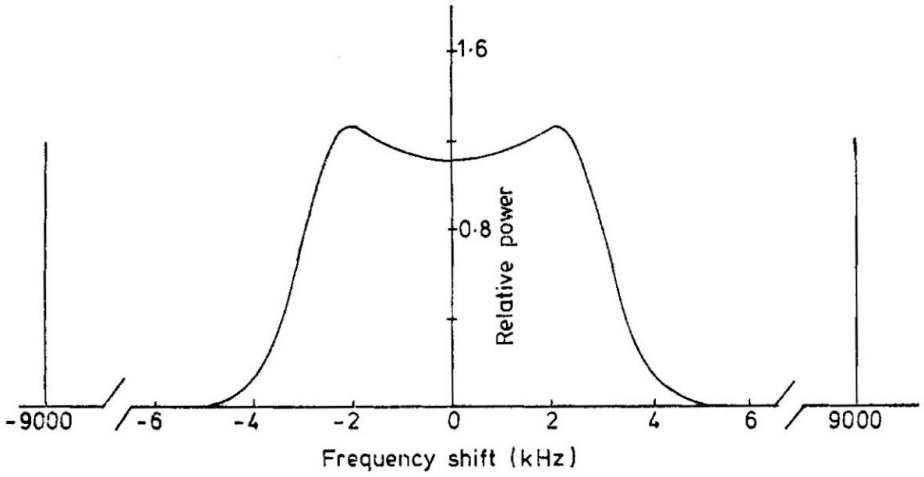

Figure 2.5. Ion spectrum and plasma lines [*Beynon and Williams*, 1978] (© IOP Publishing. Reproduced with permission. All rights reserved).

The spectrum of Figure 2.5 is composed of four spectral lines, which depend on the two waves travelling in two different directions. The ion-acoustic wave travels with a velocity that is generally slightly higher than the ions and the wave is, therefore, attenuated. This causes the two spectral lines to merge together giving the characteristic "double-hump" shape. The ion spectrum is then broadened due to the thermal motion of the ions. The steepness of the spectrum and the height of the peaks of the spectral lines increase as the ratio T_e/T_i becomes bigger. The electron-acoustic waves, instead, travel with a speed that is generally much greater than the electrons. The attenuation of the wave is therefore very little and the two spectral lines remain sharp. If there is an influx of suprathermal electrons, the spectral lines are enhanced and still remain sharp.

The IS spectrum represented in Figure 2.5 is centred at the transmitted frequency of the radar. For incoherent-scatter experiments the radar frequency is chosen as $\lambda \gg D$. In normal circumstances the bulk of the scattered power is in the ion spectrum but if there is an influx of suprathermal electrons or the plasma is excited by a strong incident radio wave (e.g. ionospheric heating), then the plasma lines can be enhanced and become more prominent.

The parameters that can be directly measured by a radar, supposing an adequate signal-to-noise ratio, are:

- Electron density.
- Electron and ion temperature.
- Ion mass.

- Plasma drift velocity.
- Ion-neutral collision frequency.
- Electric currents.

Several other parameters can be determined from the ones listed above like wind speed or temperature of the neutral atmosphere but some can require a-priori assumptions.

2.4 Summary

This chapter outlined the main characteristics and principles of the ionosphere in relation to radio propagation and gives an idea of the complexity of such a medium.

The characterization of the ionosphere is usually done (at least for the purposes of this thesis) according to the electron density.

GPS dual frequency signals can be used to observe the integrated electron density (or TEC) along the satellite-receiver ray path. This measurement technique is fundamental to the research in this thesis. It is, in fact, used in ionospheric tomography in order to reconstruct four-dimensional electron density maps of the state of the ionosphere (see Chapter 4 for more details).

The electron density can also be observed using incoherent scatter radar and in situ plasma observations. These observations are important in this thesis for verification of results.

Chapter 3
State of the art

Introduction

This chapter illustrates the literature review for Computerized Ionospheric Tomography (CIT). Starting from the first experiments on simulated data from Austin et al. (Texas), the main progresses that brought CIT to the current advancement are recalled in Section 3.1. Early validations and results are in Section 3.2. The available methods developed by different institutions are presented in Section 3.3 while conclusions are in Section 3.4.

3.1 Origins and evolution of ionospheric tomography

This section provides the literature review of CIT with comments on the main achievements and aspects. In order to do this, a brief introduction to the problem that is solved by CIT must be given. More details on the solution of the problem can be found in Chapter 4.

In a classical CIT approach the ionosphere is divided into a grid where each voxel (a three dimensional pixel) is not necessarily equidistant from each other one or has the same size. Figure 3.1 shows two satellites transmitting a signal that is received at the ground stations (the three square black boxes). The measurement is indicative of the electron content integrated along the line of sight between satellites and receivers (sTEC). Each ray is then used to retrieve the electron content $\boldsymbol{n}$ of the cell it crosses (the grey pixels in the grid for the thicker ray path).

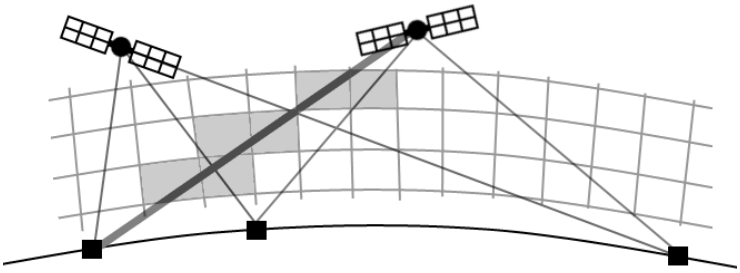

Figure 3.1. A simplified picture of the tomography scan scenario. The satellites transmit a signal received by the ground station. Each signal carries an information about the TEC integrated along the ray path and it is used to calculate the electron content along the ray path (represented by the grey pixels in the grid for one particular ray).

The problem, namely forward problem, is usually written as

$$\boldsymbol{z} = \mathbf{A}\boldsymbol{n} \tag{3.1}$$

where $\boldsymbol{z}$ is the receiver measurement and **A** a matrix that takes into account the geometry of the problem and links the unknown (i.e. the electron content $\boldsymbol{n}$ of the ionosphere) with $\boldsymbol{z}$. The inversion of the matrix **A** results in $\boldsymbol{n}$, i.e. the unknown that needs to be estimated. What appears like a simple operation has actually a very complicated solution. The errors of the instruments due to the ionosphere plus the problem of the limited number of ground stations, make the solution typically not unique and unstable. The solution requires a great effort and involves knowledge from different research fields like medical imaging, physics and maths. Usually, because of the instability of the inversion where data are missing, it is crucial to include some external aids or assumption in order to reconstruct as truthful as possible the ionospheric scenario observed through the instruments (refer to Chapter 5 for a better explanation of the regularization term used to stabilize the inversion). The assumptions of stationarity of the medium and geometric limitations such as incomplete data coverage have been addressed in the last years and improvements have been obtained [*G S Bust and Mitchell*, 2008]. However, those limitations are important to consider in ionospheric tomography in order to correctly interpret the reconstructions.

The reconstruction techniques in CT can be divided in three main groups (from [*Leitinger*, 1996]):

- *Radon-derivating methods*. The first studies of the inverse problem and its solutions were proposed by *Radon* [1917]. The method didn't find any practical application until 1956 where it was first applied to radio astronomy [*Bracewell*, 1956]. Starting from the invention of the X-ray computerized tomography scanner [*Newbold*, 1972], the Radon transform has been used mainly for medical purposes in diagnostic radiology. The Algebraic Reconstruction Technique (ART) is the first application of Kaczmarz's method to Radon's integral equation [*Natterer*, 2001] and has derived the Simultaneous Iterative Reconstruction Technique (SIRT), which is a derivative version of the original ART. Both are an iterative method where the current version of the solution is used to calculate the next iteration according to the following equation

$$\boldsymbol{n}^{k+1} = \boldsymbol{n}^{k} + \varrho_k \frac{z_i - \sum_{j=1}^{J} A_{ij} n_j}{\sum_{j=1}^{J} A_{ij} A_{ij}} \boldsymbol{A}_i \tag{3.2}$$

  where $\boldsymbol{A}_i$ is the i-th row of the matrix **A**, J is the total number of rays used for the reconstruction and i and j are the indices for the row and columns respectively. The parameter ϱ is used to make the solution convergent. The ART and SIRT derive the correction for the next iteration from the root mean square (RMS) error difference between the observed data $\boldsymbol{z}$ and the reconstructed version $\boldsymbol{n}^k$(here supposed to be a

row vector) at the step k. Both the solutions tend to introduce strong gradients and the initial guess shape is usually lost along the iterative process and, because of the ray-by-ray approach, each ray equation will introduce an error in the pixel just altered by the previous equation originating a "salt and pepper" noise. This effect is stronger in ART than in SIRT, which is an averaged and then smoothed version. The Simultaneous Algebraic Reconstruction Technique combines, instead, the main positive characteristics of ART and SIRT and adds a simultaneous correction for all the rays in a single step.

Another technique is the Multiplicative Algebraic Reconstruction Technique (MART) which uses the RMS and maintains the shape of the initial guess. It is part of the maximum-entropy algorithms i.e. part of the reconstructing theory from incomplete data according to probabilistic information. For these and other reasons it is, among the Radon techniques, the most used.

- *Model Parameter Fitting*. Part of this method is discussed in this thesis. It solves the problem by means of the least-squares fitting and using tuning parameters in order to stabilize the solution and make it more data or model-like. The model can come from a physics model (data-assimilation) or can be just a minimization of the sharpest gradients in order to reduce the noise due to incomplete data. These methods often use basis functions in order to decompose the vertical profile and the horizontal variation. These basis functions help to reduce the dimensionality of the problem and to catch the actual variation of the structures isolating, as a consequence, the noise or artefacts. A common technique to create some basis functions is the Truncated Singular Value Decomposition (TSVD). It is basically used to stabilize the reconstruction avoiding the bigger coefficients which could lead to artefacts and instability [Kunitake et al., 1995]. Another new approach are wavelets. They haven't still found a practical application and only the classic Haar [Amerian et al., 2010] and B-Spline [Durmaz and Karslioğlu, 2011; M. Schmidt et al., 2008; Zeilhofer et al., 2009] basis have been so far tested.
- Fourier Reconstruction. It is based on the Fourier Slice Theorem but hasn't found any practical application [*Leitinger*, 1996].

Computerized Tomography (CT) is the process of reconstructing an image from a series of its projections [*Jeffrey R. Austen et al.*, 1988] and was formerly used principally in medical research [*Hounsfield*, 1975] with X-ray tomography and, later, to study ocean structures too [*Munk and Wunsch*, 1979]. In 1986 the first feasibility study about using medical tomographic algorithms for CIT purposes was proposed by *J. R. Austen et al.* [1986]. They presented a simulation study in order to demonstrate the potential of the technique. By using the SIRT

algorithm they reconstructed the VTEC comparing artificial structures also in presence of noise. The author used a triangular vertical density profile as initial condition for the algorithm. The profile peak was chosen to be close to that used in the simulation. The algorithm was based on work by [*Censor*, 1983], which was preferred for problems with limited angle geometry as in CIT.

In 1988 a more exhaustive study was presented, demonstrating a good correlation between the model and the reconstruction [*Jeffrey R. Austen et al.*, 1988]. They compared ART and SIRT algorithms in order to reproduce a 2-Dimensional picture of the ionosphere through a model of the ionosphere evaluating different initial guesses, noise in the data and number of iterations. All the techniques proposed were derived from medical applications. MART was used to show the potentiality of such an algorithm in ionospheric tomography where data are incomplete [*T.D. Raymund et al.*, 1990]. MART, in fact, was considered an algorithm suitable for problems with incomplete data and was proposed by *R Gordon et al.* [1970]. The model to simulate the observations was obtained from IS radar measurements and an initial guess set from back projection methods combined with Chapman profiles [*Hargreaves*, 1995]. The work suggested the necessity of introducing a priori information in tomography especially for the vertical profile.

Due to the substantial differences between the medical and ionospheric scenarios a series of analyses were done. In fact the principal difference is that, whereas in medical application each experiment benefits from dedicated equipment, which implies the best set of measurements, in the ionospheric applications there are a series of limitations due to the large scale of the problem, measurement restrictions and a non-static medium. For this reason the bigger difference between the ionospheric tomography and the X-ray tomography lies in the receivers [*H Na and Lee*, 1992]. Because of the limitations due to the nature of the ionospheric experiment, application of a CT technique could lead to limited resolution in the reconstructed image. A discussion of these and further limitations are given in [*Yeh and Raymund*, 1991], where three experimental limitations were considered by using radio beacon data:

- *View angle limitations.* The number of available rays, and therefore the number of available observations used for the CIT inversion, is in general limited. The curvature of the Earth, in fact, in addition to the measurements at low grazing angles, where the signal is particular weak because of a stronger refraction effect, reduce the number of available rays which can be used in the reconstruction.

- *Receiver density limitations.* The TEC variation is sampled by the receiving antennas in the ground stations. The reconstructed information of the ionosphere could be reduced in the case of an under-sampling condition.
- *Finite receiver limitations.* In practical considerations the numbers of ground stations are finite. This affects the resolution in the reconstructed image and, as argued in the paper, the vertical resolution is usually worse than the horizontal one. This problem leads to solutions like using vertical basis functions to constrain the solution to be better defined in the vertical direction.

In addition to this, the paper addressed the need to implement a time-dependent algorithm that could take into account the temporal variation of the ionosphere during the data collection period. The solution to this problem was proposed in a later work [*C.N. Mitchell and Spencer*, 2003].

Another transformation, rather than the Radon ones, was proposed in [*Afraimovich et al.*, 1992]. With simulated data, a one dimensional linear algebraic matrix inversion based on least squares minimization was implemented. It was called *weighted and damped least squares* and was adapted from acoustic tomography [*Menke*, 1989]. The optimal reconstruction (the result from the inversion) made use of the available data from the observed measurements and the a priori information where observations were missing. It used empirical orthonormal functions for the vertical profiles and a power law spectrum, based on a reduced set of Fourier basis functions, to extrapolate the horizontal variations of the ionosphere (the same approach with Fourier basis functions was illustrated in [*H Na and Biswas*, 1994]). This method was extended to two dimensional functions by [*T. D. Raymund*, 1994] with vertical empirical models, which, however, was argued to include enough information in the reconstruction process but sufficient for only the large smooth features. They solved, in fact, for the least square error by exploiting the orthogonal decomposition algorithm described in [*H Na and Lee*, 1991] making the solution constrained in the space spanned by the model. This resolved the not uniqueness solution and opened the way toward optimized solutions (see Section 3.3) still based on the minimization of the least square error. The algorithm was intended for large-scale features but it was argued that the residual TEC error described the features different from the background and that other basis functions would better reconstruct smaller scale features.

It was clear at this stage the importance of including a priori information into the reconstruction algorithms where data are missing or where the reliability is diminished, especially for the vertical profiles [*T.D. Raymund et al.*, 1990]. In 1993 two algorithms were developed that incorporated a priori information for resolution enhancements [*H R Na*, 1993]:

the orthogonal decomposition algorithm and the direct Fourier method with bi-directional extrapolation. The former used Legendre polynomials in both directions of the two dimensional grid. It indicated the advantages of using separable basis functions for the vertical (such as IRI) and the horizontal dimensions. The second method was a Fourier domain reconstruction, which didn't depend on the vertical profile distribution, and tried to maximize the amount of known information from both the dimensions in the Fourier domain. In [*Sutton and Na*, 1994] it was argued that a priori information must be used to minimize the incomplete data and to handle the very poor vertical resolution. They used vertical basis functions derived from singular value decompositions from the original data to restrict the solution to lie in any particular most probable solution but limited by the a priori information. This choice of basis functions gave a good degree of freedom in the solution and a good enhancement was shown compared to the case where no a priori information was included. *Fougere* [1995] used a set of linear combinations of orthonormal basis functions coming from Chapman profiles. These profiles were used to create horizontal rays to enhance the inversion results especially in the vertical. He used a maximum entropy method as a good choice to maximize the overall available information. Good results were obtained on inverting a synthetic ionosphere derived from Gaussians. In 1997 a three-dimensional grid was used to show the potential of reconstructing a three-dimensional picture of the ionosphere [*Kuklinski*, 1997]. The experiment used data that simulated a chain of 50 NNSS (see Chapter 2) receivers spread over United States collecting data from a single NNSS satellite pass.

3.2 Early validation results and further improvements

This section aims to present the first validations of tomographic techniques by using real data. The sequence of the experiments thereafter illustrated follows the comprehensive review paper by *G S Bust and Mitchell* [2008].

The first application of tomographic reconstruction techniques with real data was conducted by the M.V. Lomonosov Moscow State University group [*Andreeva et al.*, 1990]. They implemented an iterative algorithm and collected signals from navigation satellites from three ground stations placed in Murmansk, Kem and Moscow in March and April 1990. The results showed two dimensional pictures with two depletions and some complex structures with high gradients. The same year, TEC data collected from a set of U.K. stations were used to reconstruct an image of the ionosphere on the night of the 15th December. They used the series expansion method, namely SIRT, and an initial guess obtained from Chapman layers estimated on a monthly mean data from 1968 to 1970 (solar maximum). The reconstruction looked reasonable and confirmed the potential of the technique. A similar experiment was proposed from the University of Wales, Aberystwyth [*S. E. Pryse and Kersley*, 1992]. A chain

of two receivers located in Kiruna (Sweden) and Oulu (Finland), collected TEC measurements during a satellite pass on 4 September 1986. This experiment was the first to use the European Incoherent SCATter (EISCAT) radar as an independent instrument to validate the tomographic reconstructions. The CIT images, in agreement with the radar scan, showed a northward gradient of the electron density. On December 1990 a special campaign was carried out in United Kingdom. In this occasion a chain made up of four receivers located in Lerwick, Aberdeen, Hawick and Aberystwyth were used to record TEC from eight consecutive satellite passes [*S.E. Pryse et al.*, 1993]. The different images obtained from tomographic reconstructions were in good agreement with the f_oF_2 measured from an ionosonde and illustrated the expected behaviour of the mid-latitude trough.

Another experiment used EISCAT as the independent instrument for validation [*Kersley et al.*, 1993]. The radar ran in a specific meridional mode designed for the comparison with CIT reconstructions. The Ionospheric Reference Model (IRI) was used to create the background for the ionosphere by ingesting also ionosonde data. Results showed a notable improvement in the reconstructed maps. This experiment illustrated the importance to include ionosonde data in the reconstruction in order to improve the poor capability of the methods in resolving vertical profiles.

An important step forward was obtained when phase difference techniques were employed in place of calibrated TEC values [*Kunitsyn et al.*, 1994a; *Kunitsyn et al.*, 1994b]. The idea, already introduced by [*Leitinger et al.*, 1975] was here proposed with some experiments. In those works phase differences between adjacent raypaths were demonstrated to be more sensitive in the reconstruction of localized electron density gradients than methods based on calibrated TEC values. In fact, their results suggested that the offset for the calibration of TEC was particularly difficult to estimate in presence of strong gradients of electron density.

In a joint Russian-American campaign two different approaches (one from the American and one from the Russian group) were analysed [*Foster et al.*, 1994]. The experiment used four different receivers located in the eastern United States and Canada. Data were recorded from NNSS satellites and the Russian Cicada satellites during a midlatitude geomagnetic storm. The American group employed the method describes by [*T. D. Raymund*, 1994] and implemented the ionospheric background model as described in [*Daniell*, 1991]. The Russian group, instead, utilized a horizontally stratified ionosphere as model background. The results were compared with the Millstone Hill incoherent scatter radar and showed the possibility of imaging the changes of the ionospheric structures during the storm by the two methods.

Further improvements were then introduced in CIT techniques and demonstrated with experimental validations. [*Markkanen et al.*, 1995] presented a new method where different

peak heights were introduced in the reconstruction process based on the Bayesian approach. The method was tested on simulated and experimental data from the Cicada satellites. The EISCAT radar was used to compare the reconstructions. The radar was reasonably in agreement with the tomographic reconstructions but the ambiguity on resolving ionospheric vertical profiles was not solved. *Kunitake et al.* [1995] presented an extended version of Singular Value Decomposition (SVD) used to reconstruct the state of the ionosphere over Japan. The method involved the truncation of the smallest SVD coefficients in order to reduce instabilities and noise from the data. The comparison with ionosondes and oblique sounders showed a good improvement of the results by the truncated SVD method. In [*C.N. Mitchell et al.*, 1995] a chain of four receivers was used to image field-aligned irregularities of the ionosphere over Scandinavia. The MART algorithm was used and the results were compared with the EISCAT radar. The images showed a latitudinally narrow trough with a boundary blob on its poleward wall.

3.3 State of art in CIT

During past years many different CIT algorithms have been proposed to reconstruct the ionosphere. GPS ground stations are used to provide observations for the inversion algorithms but other kinds of instruments (e.g. ionosondes or radio-occultation measurements) are in general assimilated to improve the reconstruction especially for the vertical profile. Furthermore, physics or empirical models that describe the main ionospheric behaviour are also used to provide an initial background of the ionosphere and to aid the inversion. They provide a good set of values which would otherwise be unavailable due to poor coverage of receivers (e.g. on the oceans). A physics model can also be used to forecast the future state of the ionosphere. The following techniques try to minimize a functional cost where the information coming from instruments and models (physics or empirical) are combined and weighted according to the reliability of the information, and eventually interpolated.

The following sections will describe some of the most significant algorithms used to reconstruct and predict the ionospheric behaviour. They use a priori information to drive the solution of the inversion and make it as close as possible to the truth.

3.3.1 IDA4D

The Ionospheric Data Assimilation four-dimensional (IDA4D) was developed by the University of Texas at Austin [*G Bust et al.*, 2000; *G Bust et al.*, 2004; *G Bust et al.*, 2007b]. It is based on a three Dimensional Variational (3DVAR) data assimilation algorithm described by the Naval Research Laboratories (NRL) for meteorological analysis [*Daley*, 1991; *Daley and Barker*, 1998]. It uses a minimization technique [*G Bust et al.*, 2000] to produce, at each time step, an electron density estimation used as an initial guess for the next step. A four-

dimensional map of the electron density distribution of the ionosphere is eventually produced. It can digest sTEC from different instruments like GPS receivers, incoherent scatter profiles, DORIS and Beacon data, and additionally include data from spaced based sources like the Scintillation and Tomography Receiver in Space (CITRIS), CHAMP and the Constellation Observing System for Meteorology, Ionosphere, and Climate (COSMIC) together with radio-occultation data [*G Bust et al.*, 2007a]. The 3DVAR algorithm implements a Gauss-Markov Kalman filter to provide an estimate of the electron density in the next-time (future) of the analysis. The 3DVAR tries to minimize, with a least squares metric, the modelled and measured observations. The measurements and the models are weighted out in the inversion algorithm according to the reliability of each single observation. This determines what observation prevails in the reconstruction.

The electron density is solved with its logarithmic to ensure that the electron density is positively defined. At high latitudes, IDA4D ingests high latitude drift velocities obtained from the Assimilative Mapping of Ionospheric Electrodynamics (AMIE) algorithm [*G Bust et al.*, 2007a].

3.3.2 MIDAS

The Multi-Instrument Data Analysis System (MIDAS) is a software package to solve inverse problems. It was developed at the University of Bath and evolved in three different versions:

- MIDAS version 1. It uses horizontal basis functions in the form of the spherical harmonics (see for example [*Meggs et al.*, 2005]). Those basis functions were used in a three-dimensional time-dependent algorithm [*C.N. Mitchell and Spencer*, 2003], whose inversion was based on Singular Valued Decomposition (SVD).
- MIDAS version 2. It is implemented with a Kalman filter based technique [*Spencer and Mitchell*, 2007], where the minimization aims to diagonalize a covariance matrix that is used to describe the statistic of the state (e.g. the electron density).
- MIDAS version 3. This is a 4-D data assimilation technique based on the 3DVAR technique. The method which has similarities with the linear Gauss-Newton solver explained in [*Adler et al.*, 2007], where the a priori information is calculated in order to model a sort of smoothness of the image. The Tikhonov regularization is used to solve the inversion in MIDAS and is explained in Chapter 4. The method can also incorporate basis functions.

MIDAS was used as the framework within which the methods and experiments described in this thesis were developed. Figure 3.2 shows a block diagram of the main components of MIDAS and the blocks (purple) that were modified (see Chapter 1) to produce the results in this thesis.

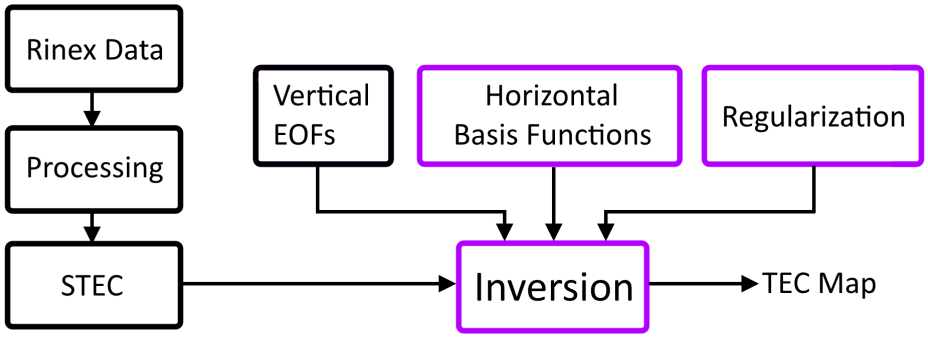

Figure 3.2. The block diagram illustrates the main components of MIDAS. The blocks modified in this thesis are shown in purple.

RINEX files are processed and converted sTEC observations, which are used to produce TEC maps of the ionosphere through the inversion. MIDAS can also include orthonormal functions to describe spatial variations in the ionosphere along the vertical and horizontal dimensions (see Chapter 4). The vertical basis functions (namely Empirical Orthonormal Functions, EOFs) are used to constrain the vertical profile according to a priori information coming from an empirical model (e.g. IRI). A regularization term is also used to stabilize the inversion (see Chapter 4).

3.3.3 GAIM (by USU)

The Global Assimilation of Ionospheric Measurements (GAIM) is a tomographic tool developed by the Utah State University (USU) team [*Schunk et al.*, 2004]. GAIM can exploit different data sources coming from ground-based locations (e.g. GPS ground station receivers, ionosondes) and space-based platforms. It provides both the specification and forecast of the Ionosphere state. There are two different versions, GAIM-GM and GAIM-FP [*R Schunk et al.*, 2005].

The GAIM-GM algorithm implements a 3DVAR algorithm which is fed by the electron density distribution obtained from the Ionospheric Forecast Model (IFM) [*Schunk et al.*, 1997]. It takes into account the electrons, the five main ion species that contribute to create the ionosphere (namely NO+, O2+, N2+, O+ and H+), electron and ion temperatures and plasma drift due to the Earth's geomagnetic field. The filter implemented in GAIM-GM is a Gauss-Markov Kalman algorithm [*Scherliess et al.*, 2006; *R W Schunk et al.*, 2005] and it is the same as described for IDA4D. It is a recursive least square procedure which gives an improvement, from a statistical point of view, by minimizing the differences between all inputs coming from observations and physics models. GAIM-GM is a mature algorithm and it has been selected for operational use at the Air Force Weather Agency in Ohama, Nebraska.

By contrast, GAIM-FP is a full physics data assimilation model. It implements a more sophisticated Kalman filter than the Gauss-Markov filter of GAIM-GM. The physics model covers a higher altitude range and includes six ion species (the same of GAIM-GM plus He+). This model allows improvements in the results by providing a three-dimensional electric field (by means of the International Geomagnetic Reference Field, IGRF) that evolves with time, as well as other drivers of the ionosphere like neutral wind and particle precipitation [*R Schunk et al.*, 2005].

3.3.4 GAIM (by USC/JPL)

The Global Assimilative Ionospheric Model (GAIM) has been developed by the University of Southern California (USC) and Jet Propulsory Laboratory (JPL), under the Multidisciplinary University Research Initiatives (MURI) program sponsored by the U.S. Department of Defence.

GAIM can ingest different line-of-sight TEC measurements from, for example, ground-based GPS receiver networks and space-borne GPS receivers, and ionosondes. It uses a first principle ionospheric model, described through hydrodynamic equations, and given in the work of *Pi et al.* [2003], to numerically solve for ion and electron volume densities.

The algorithm uses two different approaches [*Mandrake et al.*, 2005] based on 3DVAR and 4-Dimensional Variational (4DVAR) techniques. The former is the same Kalman gain filter described for IDA4D [*Daley*, 1991] where a continuous data assimilation (or 4-D data assimilation) is applied by continuously assimilating data and propagating the model. Here an approximation of the Kalman filter is applied. As a huge amount of memory is required for the full Kalman filter, the covariance matrices are truncated by keeping only a non-zero covariance for a given voxel j, in a correlation volume defined around the same voxel j [*G Hajj et al.*, 2004]. This makes the estimation of the future time step covariances based on a resolution reduced grid.

The 4DVAR does not solve for the electron density in space but attempts to solve for the drivers of the ionosphere such as production rates, temperatures and wind [*G Hajj et al.*, 2004; *Wang et al.*, 2004; *Xiaoqing et al.*, 2004]. It aims to estimate the model-drivers that satisfy the minimization requirement.

3.4 Summary

This chapter presented the most important achievements in CIT during the last 30 years, which included the use of Fourier basis functions.

Because the inversion can produce some artefacts and the data can be affected by some errors, a good representation in terms of basis functions has some advantages. If the basis functions

are well adapted to the problem, the source space will be more efficiently represented in the new domain and all the variability captured. In this way, the real variability of the ionosphere can be better separated from the noisy artefacts introduced by the limitations due to the geometry of the problem. In applications based on structure analysis it can have remarkable advantages. For example, in electron microscopy, where a resolution around a nanometre is required, the noise reduction is definitely essential [*Frank*, 2006]. Although in CIT the resolution is limited by time and geometric limitations, such an approach has still the potential to reveal better the structures in the ionosphere. Different basis functions than Fourier based ones can be used in CIT, and it will be described in Chapter 4. This is the case of wavelet basis functions, which allow to better describe regional variations of TEC in the ionosphere by exploiting the concept of sparsity. The concept implies the possibility to represent the ionosphere with only a considerably small set of wavelet basis functions. Therefore, the solution is found by looking for the minimum number of basis functions in order to efficiently represent the ionosphere from the information provided by the observations. The method has the obvious advantage of better representing the different scale structures of the ionosphere in a more efficient way. The theory of sparsity and its applications to CIT have been analysed in Chapter 5 and Chapter 7 by using 2-Dimensional and 3-Dimensional wavelets respectively.

Chapter 4
Ionospheric tomography

Introduction

This chapter describes the general CIT problem, and solutions that will be used as the basis to develop the new methods described in Chapter 5. In particular, Section 4.1 defines the CIT problem over a lattice and gives the basic mathematical notation. An implementation of the solution through Tikhonov regularization is described in Section 4.2. The inclusion of vertical and horizontal basis functions to describe the electron density spatial variation of the ionosphere is discussed in Section 4.5. The same section also describes the Multi-Resolution Analysis (MRA) through wavelets. Finally, Section 4.6 includes the summary with conclusions and further discussions.

4.1 Definition of the problem of ionospheric tomography

CIT is an inverse geophysical problem. In this section the mathematical description of the problem is provided. The forward problem defines the common scenario of a measurement intended as the observation of an entity (or source). The inverse problem, instead, defines the opposite situation, i.e. the definition of the entity that produced the observation. The latter problem is what is solved in tomography. In mathematical notation this means that given a set of observed data z, tomography aims to determine the model parameters or sources n. In CIT, the basic concept is to use data sets from the GPS constellation (different observations can be used, such as radars, radio occultation, etc.) to measure the electron content along the signal path from different ground stations placed around the Earth.

The forward problem can be defined as

$$z = A(n) \tag{4.1}$$

where z is the observed data and n is the source data. The operator A can be seen as a projection of the source into the observation set.

Therefore, the source data is obtained by inverting the operator A

$$n = A^{-1}(z) \tag{4.2}$$

The inversion is a critical process. In fact, the solution must satisfy the conditions of uniqueness and stability. The simple concept of using real data confirms from a physics point

of view the existence of a unique ionosphere and, therefore, the existence of a unique solution. Unfortunately, almost all geophysical problems are ill-posed. This means that the conditions listed above are not satisfied; however it does not mean that a physically meaningful solution cannot be found [*Zhdanov*, 2002]. In an ill-posed problem an infinite number of solutions exists but only a subset of the whole set contains the most reasonable solutions.

Focusing on ionospheric tomography, the source n corresponds to the electron density of the ionosphere, whilst the observed data z represents the TEC measured along the receiver and the satellite ray path. The receivers can be satellite-based or ground-based. The latter is considered in this thesis. In CIT, one of the main problems is the ground stations coverage. In fact, the distribution of ground stations around the globe is uneven and sparse, and this causes data gaps (e.g. in the Oceans where ground stations cannot be placed) in the observed data, which makes the problem *underdetermined* and, thus, not invertible.

The ionosphere is usually discretized in a finite number of samples (or voxels, i.e. three-dimensional pixels). For a three-dimensional ionosphere with dimension $\mathrm{L} = \mathrm{L_V} \times \mathrm{L_{H_x}} \times \mathrm{L_{H_y}}$, the discretization is made in a grid according to the following figure

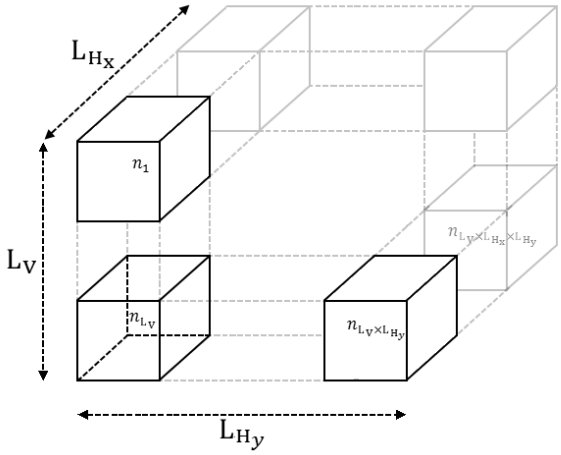

Figure 4.1. The ionosphere is discretized into a 3 dimensional grid with dimensions $\boldsymbol{L = \mathrm{L_V} \times \mathrm{L_{H_x}} \times \mathrm{L_{H_y}}}$. The grid is stored in a column vector $\boldsymbol{n}$ whose coefficients are ordered by columns, as seen in the figure.

Each sample is then stored in a column vector $\boldsymbol{n}$ containing all the discretized values of the ionosphere grid, i.e.

$$\boldsymbol{n} = \left[n_1,\ n_2, \ldots,\ n_{\mathrm{L_V}+1}, \ldots,\ n_{\mathrm{L_V}\times\mathrm{L_{H_y}}}, \ldots,\ n_{\mathrm{L_V}\times\mathrm{L_{H_x}}\times\mathrm{L_{H_y}}}\right]^T \tag{4.3}$$

where n_i represents a defined value of electron density in a given voxel of the grid.

The observed data $\boldsymbol{z}$ are generated from sTEC measurements, whose definition is recalled in Equation (4.4). As described in Chapter 2, the sTEC is the electron density integrated along the ray path between the satellite and receiver, which can be discretized, for a single ray r, as

$$\mathrm{TEC(r)} = z_{\mathrm{r}} = \sum_{\mathrm{s}=1}^{L} n_{\mathrm{s}} \cdot a_{\mathrm{rs}} \tag{4.4}$$

where n_{s} is the electron content in the position s of the grid, and a_{rs} is the length of the s-th ray within the voxel s. Therefore, if the operator A is linear, Equation (4.1) can be rewritten as a product of matrix and vector

$$\boldsymbol{z} = \mathbf{A}\boldsymbol{n} \tag{4.5}$$

where $\boldsymbol{z}$ and $\boldsymbol{n}$ are two column vectors of size M and L respectively, i.e.

$$\boldsymbol{z} = [z_1,\ z_2,\ z_3, \ldots,\ z_M]^T \tag{4.6}$$

$$\boldsymbol{n} = [n_1,\ n_2,\ n_3, \ldots,\ n_L]^T \tag{4.7}$$

Consequently the matrix $\mathbf{A}$ is a $M \times L$ matrix of the operator A. Each element of the data vector $\boldsymbol{z}$ is considered as one measurement coming from one GNSS ray path. Equation (4.5) makes the assumption of a straight ray path. This is a good approximation that works well to describe the large structures in the ionosphere [*Gorbunov et al.*, 1997].

Because of the limited distribution of ground receivers, usually $M < L$ and the system is generally underdetermined.

4.2 Tikhonov regularization

An underdetermined problem has an infinite number of solutions. In order to make the solution unique, a regularization (or penalty) term is generally included. This term provides extra information that guides the solution toward the most reasonable one.

In order to solve Equation (4.5) an estimated version of the source model $\widehat{\boldsymbol{n}}$ (or pseudo-inversion) needs to be found. It provides an estimation of the real observed data $\boldsymbol{z}$ as

$$\hat{\boldsymbol{z}} = \mathbf{A}\widehat{\boldsymbol{n}} \tag{4.8}$$

where $\hat{\boldsymbol{z}}$ is the predicted observation which has to be close enough to the real observed data vector $\boldsymbol{z}$. By defining a metric, it is possible to have a measure of the "distance" between the real and predicted data, and trying to reduce it as much as possible. In this way, it is possible to evaluate the accuracy of the inverse problem [*Zhdanov*, 2002]. Any residual in the difference between $\hat{\boldsymbol{z}}$ and $\boldsymbol{z}$ can be due to a number of factors such as errors in the

measurements, inexact description of the projection operator **A** or, as described here, approximation from the regularization term.

The inversion can be solved by minimizing a cost function (see Appendix A) through a least square algorithm that leads to the following equation

$$\widehat{\boldsymbol{n}} = \left(\mathbf{A}^{\mathrm{T}}\mathbf{A}\right)^{-1}\mathbf{A}^{\mathrm{T}}\boldsymbol{z} \tag{4.9}$$

where $\widehat{\boldsymbol{n}}$ is the pseudo-inverse solution. Unfortunately, the pseudo-inverse solution is unstable because the undetermined problem [*Zhdanov*, 2002] makes the inversion singular. A solution for ill-posed problems was proposed by the Russian mathematician Andrei Tikhonov in 1977. He introduced a regularization method that makes an ill-posed problem well-posed. The regularization acts on the operator **A** which, instead of considering all the possible values in N space, it selects only a subspace which gives more reliable sources and, therefore, solutions. As a consequence, the solution is highly dependent on the type of regularization used in the inversion. Tikhonov introduced the following parametric functional cost [*Tikhonov and Arsenin*, 1977]

$$f^{\varrho}(\boldsymbol{n},\boldsymbol{z}) = \|\mathbf{A}\boldsymbol{n} - \boldsymbol{z}\|^2 + \varrho\|\mathbf{W}_{\mathrm{m}}\boldsymbol{n}\|^2 \tag{4.10}$$

The solution $\widehat{\boldsymbol{n}}$ is then obtained by minimizing the functional cost $f^{\varrho}(\boldsymbol{n},\boldsymbol{z})$ with respect to $\boldsymbol{n}$

$$\widehat{\boldsymbol{n}} = \min_{\boldsymbol{n}} f^{\varrho}(\boldsymbol{n},\boldsymbol{z}) \tag{4.11}$$

which provides the following solution (namely the Regularised Least Square solution, RLS)

$$\widehat{\boldsymbol{n}} = \left(\mathbf{A}^{\mathrm{T}}\mathbf{A} + \varrho\,\mathbf{W}_{\mathrm{m}}^{\mathrm{T}}\mathbf{W}_{\mathrm{m}}\right)^{-1}\mathbf{A}^{\mathrm{T}}\boldsymbol{z} \tag{4.12}$$

where $\mathbf{W}_{\mathrm{m}}$ is a weighted matrix (usually diagonal) and ϱ is a parameter that sets a trade-off between the best fitting and the most reasonable stabilization [*Zhdanov*, 2002]. A small ϱ value will reduce the regularization effect. Therefore, if the problem is ill-conditioned, the solution would become unstable. The instability may give rise to artefacts/noise in the solution. As the ϱ value gets bigger, the regularization will drive the solution according to the definition of $\mathbf{W}_{\mathrm{m}}$. The weighted matrix can be defined as the Jacobian (first-order partial derivatives) or Hessian (second-order partial derivatives) matrix. As a consequence, if ϱ is too big the regularization would smooth the solution to the point that structures would be barely visible.

To help the inversion, an initial guess $\boldsymbol{n_0}$ can be used (see Appendix A) in the functional cost as follows

$$f^{\lambda}(\boldsymbol{n},\boldsymbol{b}) = \|\mathbf{A}\boldsymbol{n} - \boldsymbol{z}\|^2 + \varrho\|\mathbf{W}_{\mathrm{m}}\boldsymbol{n} - \mathbf{W}_{\mathrm{m}}\boldsymbol{n_0}\|^2 \tag{4.13}$$

The vector $\boldsymbol{n_0}$ brings other information, which normally comes from a physics model, and is part of the data assimilation technique which is not discussed in this dissertation. The initial guess provides a background of values that are available where the data from the rays are missing or not sufficient. Increasing the ϱ parameter will force the solution to behave like the physics model along with the choice of $\mathbf{W}_{\mathrm{m}}$ (i.e. if the derivative matrix is used, the solution will behave like the gradient of the model).

If $\mathbf{W}_{\mathrm{m}}$ is the Hessian matrix, by setting the vector $\boldsymbol{n_0}$ to zero (for example) the minimization algorithm will force the solution to have a small rate of change for big values of ϱ and, as a consequence, the solution would appear smooth and with little detail. This is true in the case of standard Tikhonov regularization, where the model is set to zero. However, for completeness, the pseudo solution is thereafter described by including the term $\boldsymbol{n}_0$. In this case the final solution will consist in the model $\boldsymbol{n}_0$ and the complementary information $\delta\widehat{\boldsymbol{n}}$ extracted from the observations, which was not present in the model. Thus

$$\widehat{\boldsymbol{n}} = \boldsymbol{n}_0 + \delta\widehat{\boldsymbol{n}} \tag{4.14}$$

where $\boldsymbol{n}_0$ is the initial guess (or background) and,

$$\delta\widehat{\boldsymbol{n}} = \left(\mathbf{A}^{\mathrm{T}}\mathbf{A} + \varrho\mathbf{W}_{\mathrm{m}}^{\mathrm{T}}\mathbf{W}_{\mathrm{m}}\right)^{-1}\mathbf{A}^{\mathrm{T}}(\boldsymbol{z} - \mathbf{A}\boldsymbol{n}_0) \tag{4.15}$$

is the information added to the background obtained from the measurements. The inversion in this case is calculated on the difference between the observation data and the projection of the background, because this residual information is the component missing in the background.

Since Equation (4.13) could still be unstable because $\mathbf{W}_{\mathrm{m}}$ could not be full rank, another parameter α can be introduced that could further stabilize the inversion

$$f^{\varrho,\alpha}(\boldsymbol{n},\boldsymbol{z}) = \|\mathbf{A}\boldsymbol{n} - \boldsymbol{z}\|^2 + \varrho\|\mathbf{W}_{\mathrm{m}}\boldsymbol{n} - \mathbf{W}_{\mathrm{m}}\boldsymbol{n_0}\|^2 + \alpha\|\boldsymbol{n} - \boldsymbol{n_0}\|^2 \tag{4.16}$$

The tuning parameter α sets the result to be like the model in terms of absolute values. Usually α is set to a small value, but in general it is chosen such that $\alpha \ll \varrho$.

4.2.1 Performance of Tikhonov regularization by varying its tuning parameters

The formulation of Equation (4.16) combines two different regularizations terms governed by the parameters ϱ and α. These terms depend on the model $\boldsymbol{n_0}$, but only the first one is defined through the matrix $\mathbf{W}_{\mathrm{m}}$. In order to illustrate the effects of these regularizations, two simple two-dimensional experiments were set up, which analysed the regularization with partial and full data coverage. The Tikhonov regularization was implemented by setting $\boldsymbol{n_0}$ to zero and

$\mathbf{W}_m$ to the Hessian matrix. A simple two-dimensional ionosphere was simulated on the basis of a normalized profile taken from the International Reference Ionospheric (IRI) model shown in Figure 4.2.

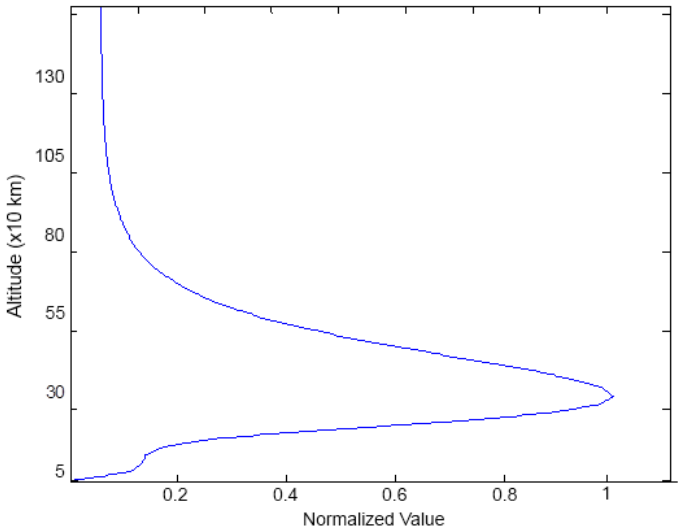

Figure 4.2. The ionosphere was created according to this normalized profile, and the same profile has been used to create the simulated ionosphere.

The final two-dimensional ionosphere was created by scaling the same profile with different values. The resulting simulated ionosphere is shown in Figure 4.3 together with a hypothetical arrangement of rays (green rays in the figure).

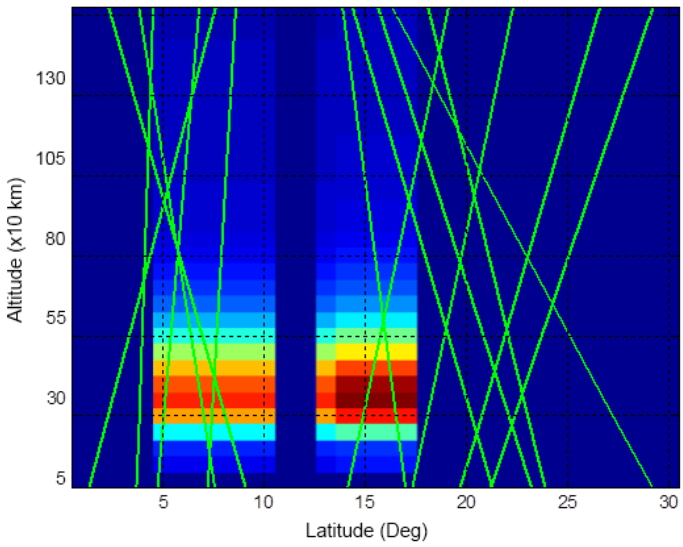

Figure 4.3. Simulated ionosphere produced for scaled versions of one IRI profile. Two structures with a peak around 300 km are present. The green rays represent the ray path between the satellites and the receivers.

This scenario considered an ionosphere with no GPS coverage on certain regions and with a total of 15 ray paths.

In Figure 4.4 the ϱ parameter is moving from 0.1 to 10 in order to show its effect in the reconstruction. The α parameter has been set to 0.1. This is a minimum value chosen empirically in order to stabilize the inversion and avoid artefacts in the reconstruction. A small value of ϱ gives a good result in zones where the rays cross the grid. A bigger value of ϱ still produces a good reconstruction where data is available but also smoothen the region where there are no rays. This identifies ϱ as a non-critical parameter, and its value can be chosen arbitrarily according to necessity. It has to be noted that, where the rays are missing, the regularization applies a solution which consists of minimizing the rate of change of gradient. Therefore, in the data gap zone, the solution tends to be a straight line. As expected, the results show the impossibility of discerning the two structures where rays are not present and without any reliable initial guess. A big value of ϱ results in a solution that minimizes the regularization term $\|\mathbf{W}_m \boldsymbol{n}\|^2$ where $\boldsymbol{n_0}$ was set equal to zero.

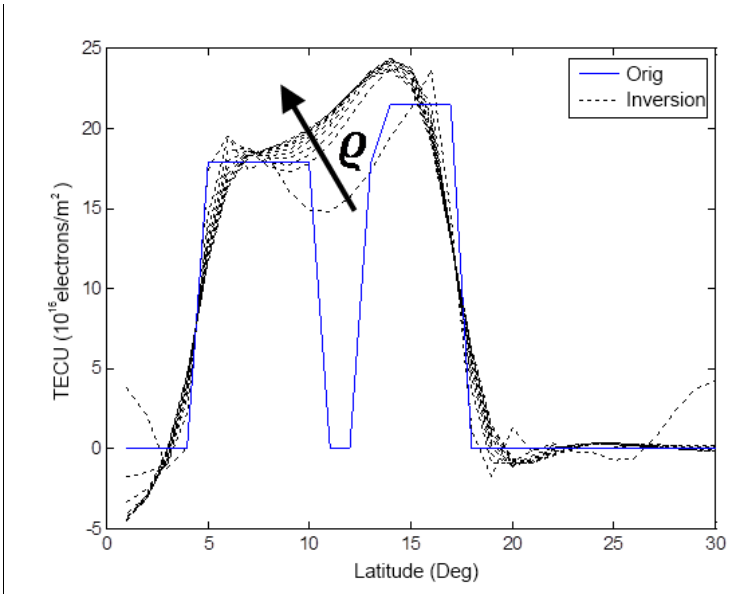

Figure 4.4. The original TEC profile (blue) is compared with reconstructed TEC (black) by applying a regularization with $\alpha = 0.1$ fixed and $\varrho = 0.1 \ldots 10$. It can be noted that, as ϱ increases, the regularization tends to reduce the rate of change of the electron content.

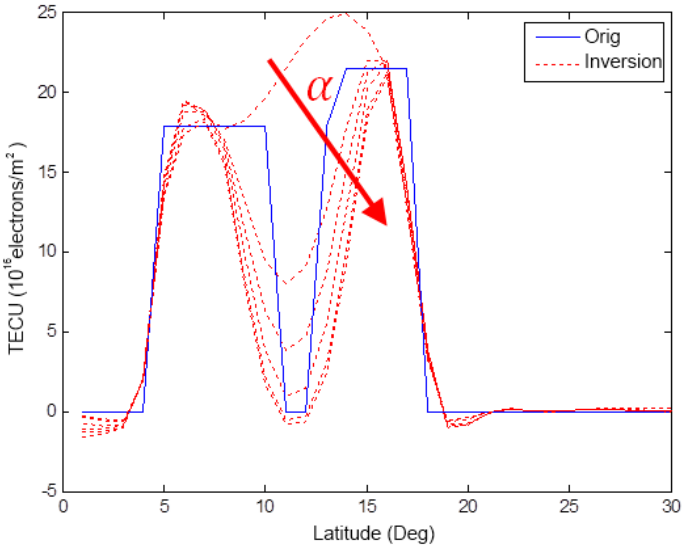

Figure 4.5. The influence of the parameter α in the inversion result is shown here. The image has been obtained with $\varrho = 0.1$ and with different values of $\alpha = 0.1 \ldots 1$. As α gets bigger, the inversion minimizes the solution trying to better match the electron content with the n_0 profile (e.g. coming from a model). In this particular example, n_0 has been considered a zero vector.

Figure 4.5 represents the same inversion, but with different parameters compared with Figure 4.4. Here ϱ is fixed to the value 0.1, and α is varying from 0.1 to 1. Also, in this case the two minimum values are chosen in order to avoid singularities, and therefore artefacts, in the inversion. As α increases, the TEC values tend toward zero (where there are no rays, and, thus, information), constraining the solution to the model (set to zero in this scenario).

The two parameters ϱ and α are important for the inversion result. However, it seems that results are more sensitive to α, when they are used together.

The following case shows a more uniform distribution of rays with a simulated ionosphere that was created from scaled versions of one IRI profile (Figure 4.6).

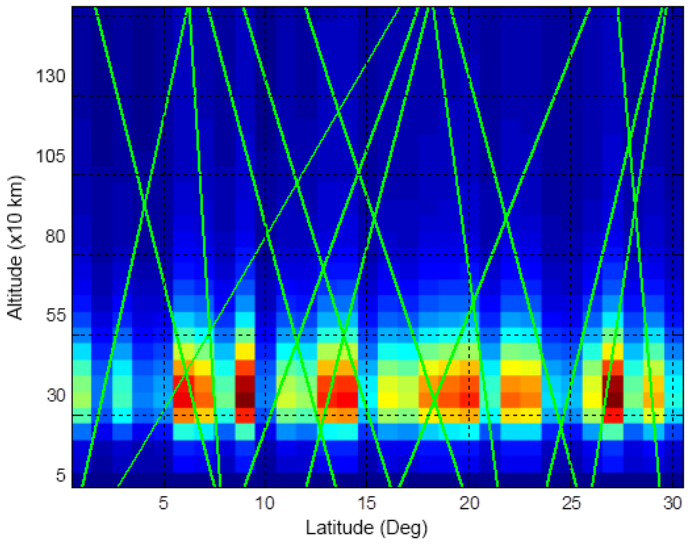

Figure 4.6. Simulated ionosphere produced for scaled versions of one IRI profile. Different structures with a peak around 300 km are present. The green rays represent the ray path between the satellites and the receivers.

Figure 4.7 shows the root mean square (RMS) error between the true ionosphere and the reconstructed one for different values of ϱ and α. The tuning parameters ϱ and α select, from all the possible solutions, the one that better represents the problem. The minimum RMS value is, in fact, reached with values of ϱ and α bigger than zero. Better results are obtained with the combination of both the tuning parameters. However, since bigger values produce results that move away from the optimum one, it is desirable to keep them small. The correct choice of these two values is important as they select a compromise between detail and artefacts in the reconstruction. Small values produce results with more detail but also artefacts, while bigger values produce results with fewer details but less artefacts.

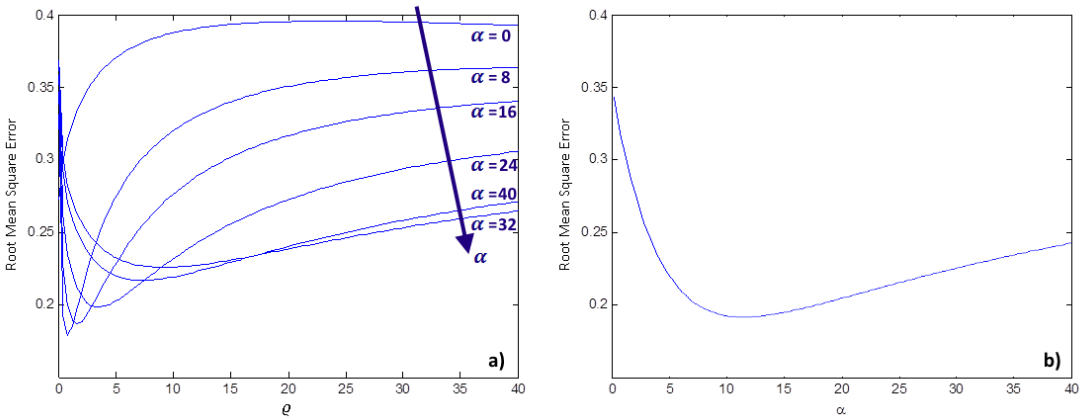

Figure 4.7. The RMS error between the true (simulated) and reconstructed ionosphere is shown for different values of ϱ and α. The figure shows the RMS error for: a) ϱ and α varying betwwen 0 and 40; b) α varying between 0 and 40 and ϱ equals to 3. Generally these curves present a minimum which is selected for a particular value of the two tuning coefficients.

4.3 The weighting matrix

The weighting matrix $\mathbf{W}_{\mathrm{m}}$ introduced in Equation (4.10) is defined according to the problem, and usually corresponds to the Jacobian (first-order partial derivatives) or Hessian (second-order partial derivatives) matrix. The most used approach attempts to solve the problem by searching for the smoothest solution, and this can be considered as an optimal approach [*Nesterov and Kunitsyn*, 2011]. The solution is found by means of the RLS technique that consists in the minimization of the functional cost in Equation (4.16), which depends on the values of ϱ and α.

A general solution considers the weighted matrix to be the Hessian matrix. As explained before, this is an optimum choice that allows the reduction of potential artefacts caused by the limited geometry of the problem and the uneven distribution of ray path coverage. The solution becomes smoother as ϱ increases as rapid variations of electron density between adjacent voxels will not be allowed.

There is no particular reason to use the Jacobian instead of the Hessian matrix, and usually they do not introduce an appreciable difference in the results. However, it must be noted that the second derivative is a stronger definition from a mathematical and physics point of view than the first derivative. Generally, from the physics aspect, the real world does not have discontinuities, and functions are infinitely differentiable. Therefore, the second derivate assumption can be considered as a good choice. Other constructions, even if not explicit, exist and are explained in Chapter 5. Another application of the regularization is in the definition of the time dimension when observations are collected within a long time window. A time-regularization term can be constructed similarly to the one defined in this chapter for the horizontal variation of the electron density of the ionosphere. The application of the regularization in time is explained in Chapter 7 with an analysis based on real GPS data.

4.4 Resolution Matrix

Section 4.3 showed the possibility to solve the inverse problem and identify a unique solution by means of a regularization term. However, the reliability of the reconstruction could be very limited, especially in zones where data are completely missing or few. The resolution matrix identifies the zones where the inversion has a good reliability and where it has not. It also provides a good picture about the data dispersion.

According to Equation (4.12) the pseudo-inverse solution is

$$\widehat{\boldsymbol{n}} = \left(\mathbf{A}^{\mathrm{T}}\mathbf{A} + \varrho\, \mathbf{W}_{\mathrm{m}}^{\mathrm{T}}\mathbf{W}_{\mathrm{m}}\right)^{-1}\mathbf{A}^{\mathrm{T}}\boldsymbol{z} \tag{4.17}$$

where

$$\boldsymbol{z} = \mathbf{A}\boldsymbol{n} \tag{4.18}$$

By inserting Equation (4.18) into Equation (4.17) the definition of the model resolution matrix $\mathbf{R_m}$ is obtained

$$\begin{aligned}\hat{\boldsymbol{n}} = \left(\mathbf{A}^{\mathrm{T}}\mathbf{A} + \varrho\, \mathbf{W}_{\mathrm{m}}^{\mathrm{T}}\mathbf{W}_{\mathrm{m}}\right)^{-1}\mathbf{A}^{\mathrm{T}}\boldsymbol{z} = \\ \left(\mathbf{A}^{\mathrm{T}}\mathbf{A} + \varrho\, \mathbf{W}_{\mathrm{m}}^{\mathrm{T}}\mathbf{W}_{\mathrm{m}}\right)^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{A}\,\boldsymbol{n} = \mathbf{R_m}\,\boldsymbol{n}\end{aligned} \tag{4.19}$$

where

$$\mathbf{R_m} = \left(\mathbf{A}^{\mathrm{T}}\mathbf{A} + \varrho\, \mathbf{W}_{\mathrm{m}}^{\mathrm{T}}\mathbf{W}_{\mathrm{m}}\right)^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{A} \tag{4.20}$$

The aim of any inversion technique is to have the pseudo-inverse solution $\hat{\boldsymbol{n}}$ as close to the true solution $\boldsymbol{n}$ as possible; which implies that the resolution matrix $\mathbf{R_m}$ has to be an identity matrix. The measure of how far the resolution matrix is from being an identity gives information about the reliability of the inversion for a well-defined point. The columns of the resolution matrix are called Point Spread Functions (PSF) and represent the response of the inversion to a delta-like impulse in the model. It describes well the actual resolution of each cell. Figure 4.8 (right) represents a pictorial figure of an ideal resolution matrix with all ones in the diagonal apart from the sixth column where the values are spread into different pixels. The wide shape of the PSF means that the resolution associated to a given cell in the ionosphere grid is wider, and hence, its value oversteps the contiguous pixels (Figure 4.8 left). In this case, the value of the cell is not representative of the single cell itself, but of a wider area making the resolution poorer.

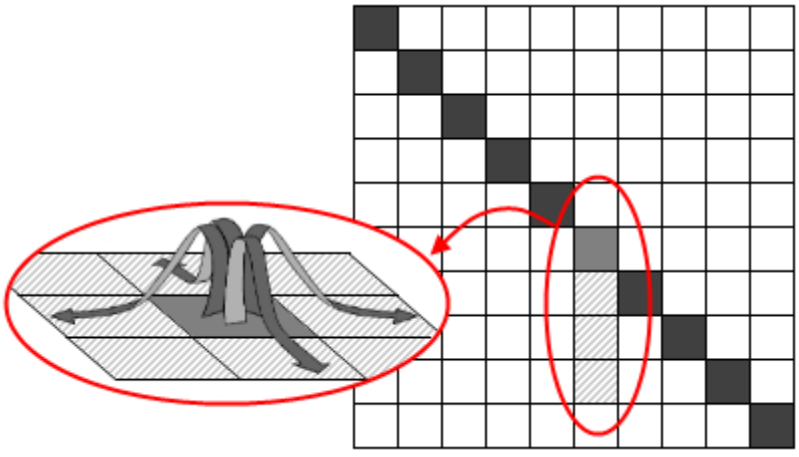

Figure 4.8. The resolution matrix (right) gives information about the reliability of the solution obtained through inversion techniques. The columns are called point spread functions (PSF) and are representative of the actual resolution of a particular cell in the grid (left). If the PSF is not a delta like function, the value of the cell is spread to the neighbours and its resolution is decreased.

In contrast, the rows of the resolution matrix are known as the Backus-Gilbert (BG) kernel or Averaging Kernel (AK). The BG kernel is defined as the operator that projects the true source

$\boldsymbol{n}$ into the approximated solution $\hat{\boldsymbol{n}}$. If the rows are delta-like functions, then the approximated solution will not be estimated (averaged) by any other value but itself in the true solution. Therefore the approximated solution will coincide with the true solution.

In the same way as the model resolution matrix $\mathbf{R_m}$, it is possible to define the data resolution matrix $\mathbf{R_d}$. According to Equation (4.17) the data generated by the pseudo solution $\hat{\boldsymbol{n}}$ is given by

$$\hat{\boldsymbol{z}} = \mathbf{A}\hat{\boldsymbol{n}} = \mathbf{A}\left(\mathbf{A}^{\mathrm{T}}\mathbf{A} + \lambda\, \mathbf{W}_{\mathrm{m}}^{\mathrm{T}}\mathbf{W}_{\mathrm{m}}\right)^{-1}\mathbf{A}^{\mathrm{T}}\boldsymbol{z} = \mathbf{R_d}\boldsymbol{z} \tag{4.21}$$

where

$$\mathbf{R_d} = \mathbf{A}\left(\mathbf{A}^{\mathrm{T}}\mathbf{A} + \lambda\, \mathbf{W}_{\mathrm{m}}^{\mathrm{T}}\mathbf{W}_{\mathrm{m}}\right)^{-1}\mathbf{A}^{\mathrm{T}} \tag{4.22}$$

With the same consideration of the model resolution matrix, if the data resolution matrix is an identity, the prediction errors are zero ($\hat{\boldsymbol{z}} = \boldsymbol{z}$) and $\boldsymbol{z}$ is completely resolved in $\hat{\boldsymbol{n}}$.

4.5 Basis function decomposition

The inverse problem has been so far described on a three-dimensional grid where each voxel has contributed independently from the others. The regularization term was then applied in order to better correlate nearby voxels, for example by smoothing big variations of electron densities in the reconstruction. Another approach is the use of basis functions that exploits the vertical and the horizontal correlation of structures in the ionosphere.

The space of the source data $\boldsymbol{n}$ can be decomposed into the sum of weighted basis functions (Figure 4.9).

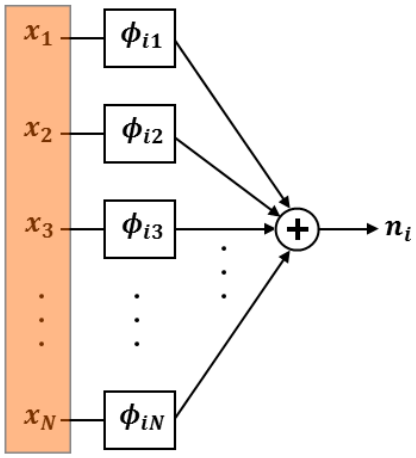

Figure 4.9. The signal n_i can be represented as the sum of different basis functions ϕ_{ij} weighted with the coefficients x_j.

The Fourier transform is an example of how a signal can be decomposed into the sum of harmonics with different wavenumbers. The decomposition can be applied to both the vertical profiles and the horizontal variations of electron densities. Furthermore, the vertical and horizontal basis functions can be coupled together in the set represented by the matrix **K**. According to Figure 4.9 it is possible to rewrite the problem of Equation (4.5). Assuming there are a total of V basis functions,

$$\boldsymbol{z} = \mathbf{A}\boldsymbol{n} = \mathbf{A}\mathbf{K}\boldsymbol{x} \tag{4.23}$$

where

$$\mathbf{K} = [\boldsymbol{\phi_1},\ \boldsymbol{\phi_2},\ \boldsymbol{\phi_3}, \ldots,\ \boldsymbol{\phi_V}] \tag{4.24}$$

with $\mathbf{K} \in \mathbb{R}^{\mathrm{L}\times\mathrm{V}}$, $\boldsymbol{x} \in \mathbb{R}^{\mathrm{V}\times 1}$, $\mathbf{A} \in \mathbb{R}^{\mathrm{M}\times\mathrm{L}}$, and $\boldsymbol{\phi_i} \in \mathbb{R}^{\mathrm{L}\times 1}$ is the i-th basis function. If the basis functions satisfy the following identity

$$\mathbf{K}^{\mathrm{T}}\mathbf{K} = \mathbf{I} \tag{4.25}$$

the basis functions form an orthonormal set. Thus, the solution with the basis function decomposition becomes (Appendix A)

$$\hat{\boldsymbol{x}} = \left(\mathbf{K}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{A}\mathbf{K} + \varrho\mathbf{K}^{\mathrm{T}}\mathbf{W}_{\mathrm{m}}^{\mathrm{T}}\mathbf{W}_{\mathrm{m}}\mathbf{K} + \alpha\mathbf{K}^{\mathrm{T}}\mathbf{K}\right)^{-1}\mathbf{K}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\boldsymbol{z} \tag{4.26}$$

where x_i represents the i-th coefficient associated with the basis function $\boldsymbol{\phi_i}$.

Good basis functions can extrapolate information from the observed data through the inversion and isolate noise (artefacts) from the useful data. In general, for structural analysis (where structures are investigated), the damping stage of the coefficients $\hat{\boldsymbol{x}}$ becomes important. A common technique used in inversion problems is the shrinkage function (for example [*D.L. Donoho*, 1992; *D. L. Donoho and Johnstone*, 1994a; *Fodor and Kamath*, 2003]). It allows removing those coefficients that could potentially cause artefacts in the reconstruction. More details on the shrinkage function and its implementation in the regularization term can be found in Chapter 5. Furthermore, by selecting only a small representative set of basis functions it is possible to reduce the size of the problem described by Equation (4.23) and, therefore, the related computational problem. This is, for example, the case in the vertical Empirical Orthonormal Functions (EOFs) described in the following subsection.

4.5.1 Vertical Basis Functions

The vertical (along the altitude) resolution from a GPS measurement is usually very poor [*T. D. Raymund et al.*, 1994], and it can affect the inversion technique by creating TEC profiles

which are not true representations of the ionosphere. A common procedure is to introduce some a priori information into the reconstruction algorithm by using orthogonal decomposition [*Sutton and Na*, 1994]. It constrains the vertical distribution of the reconstructed profile to a space which depends on the a priori information. Singular Value Decomposition (SVD) [*Hansen*, 1987] is a method that can be used to extrapolate a set of vertical basis functions from, for example, the IRI-95 model [*Fremouw et al.*, 1992; *Howe et al.*, 1998; *C.N. Mitchell and Spencer*, 2003]. A minimum of three basis functions can be used to define the height of the peak and width of the electron density profile along the altitude. One direct consequence of this approach is that the number of unknowns in the Equation (4.5) is dramatically reduced. This also decreases the computational effort to calculate the inversion. The solution can thus be written as

$$n_m = \sum_{l=1}^{L} k_{ml} \cdot x_l \tag{4.27}$$

where x_l is the l-th coefficient that weight each vertical basis $\boldsymbol{k}_l$, and n_m the m-th element of $\boldsymbol{n}$. In matrix notation it is possible to rewrite the Equation (4.5) by inserting Equation (4.27) as

$$\boldsymbol{z} = \mathbf{A}\mathbf{K}_\mathbf{V}\boldsymbol{x} \tag{4.28}$$

where $\mathbf{K}_\mathbf{V}$ is the matrix containing vertical basis functions. The solution of Equation (4.28) is the same as that of Equation (4.26).

These vertical basis functions are usually obtained from empirical models (e.g. IRI), and, as such, are called Empirical Orthonormal Functions (EOFs). In addition to reducing the order of the problem, they also compensate for the fact that the matrix **A** is sparse. The source data $\boldsymbol{n}$ are represented by the coefficients $\boldsymbol{x}$ of the EOFs, giving a good representation of the whole vertical profile even if data are missing. For example, it is sufficient to have a single ray crossing the whole two-dimensional grid in order to reconstruct the entire profiles in the same grid. It must be noted that the EOFs vary significantly with the day, hour, season and sun activity. This variation should be taken into account when global maps are calculated [*Erturk et al.*, 2009].

4.5.2 Horizontal Basis Functions

For the ionosphere, vertical basis functions represent a scenario that completely differs from the horizontal variation of electron content. Thus, it is useful to define a new set of basis which best suit the horizontal behaviour, in addition to the vertical basis functions.

One classical approach to obtain the horizontal basis function is by means of SVD. Historically, the method has been used for image restoration and is useful for reducing noise

[*Huang and Narendra*, 1975]. It is also often used to reduce singularities in an inverse problem (for example [*Amerian et al.*, 2010; *G A Hajj et al.*, 1994]).

Others basis functions are also adopted that take into account factors such as periodicities (using spherical harmonics) or localized variations (using wavelets). These are discussed in this section.

4.5.2.1 Spherical harmonics

Spherical harmonics define a complete set of basis functions over the sphere and consist of a Fourier basis for longitude increments at fixed latitudes and a Lagrange polynomial (along the latitude) at fixed longitudes. The Fourier basis functions have a long domain and are made up of sinusoids. For this reason, they are able to identify periodicity (in terms of wavenumber) but fail where the signal has local variation. Due to the direct relationship with the wavenumber, the transformed domain is referred to as the wavenumber domain.

Spherical harmonics basis functions are defined as [*Press et al.*, 2007]

$$Y_{l,m}(\theta,\phi) = K_{l,m}\, P_l^m(\cos\theta) e^{im\phi} \tag{4.29}$$

where θ and ϕ are the latitude and longitude respectively, P_l^m is the Legendre polynomial and the normalization constant K_l^m is

$$K_l^m = \sqrt{\frac{(2l+1)(l-m)!}{4\pi(l+m)!}} \tag{4.30}$$

Equation (4.29) define a spherical harmonic in terms of two integer parameters: the order l and the degree m^2, which are related by $-l \leq m \leq l$. In practice, the order l acts similarly to a frequency in the spherical domain, which can be limited to a maximum value in order to create a band-limited approximation of a signal.

4.5.2.2 Wavelets

Wavelets are functions used to decompose a signal into different refined versions in which the feature of the signal can be extrapolated. The difference between the spherical harmonics transform and wavelets is that, while spherical harmonics can identify the presence of a particular wavenumber in a signal, they cannot localize the instance (time or distance) when that wavenumber appeared. In contrast, wavelets allow the localization (in time or distance) of a particular event, but they are not directly correlated with the wavenumber.

A wavelet family is defined in terms of dilation a and translation b of a mother function $\psi(t)$

$$\Psi_{a,b}(t) = \frac{1}{\sqrt{a}}\Psi\left(\frac{t-b}{a}\right) \tag{4.31}$$

where the factor $1/\sqrt{a}$ is chosen to satisfy the orthonormal constraint

$$\langle \Psi, \Psi \rangle = 1 \tag{4.32}$$

and $\langle \cdot \rangle$ represents the inner product.

Thus, a signal $g(t)$ can be decomposed through wavelets according to the following transformation

$$G(a,b) = \int_{-\infty}^{+\infty} g(t)\frac{1}{\sqrt{a}}\Psi^*\left(\frac{t-b}{a}\right)dt \tag{4.33}$$

Generally, for practical application, the number of $\Psi(t)$ coefficients is reduced to a discrete subset by constructing an orthonormal set of discrete basis functions obtained by sampling the continuous ones [*Mallat*, 2008], i.e.

$$\psi_{ab}(s) = \Psi(2^a s - b\Delta s) \tag{4.34}$$

where s is a discrete variable, and Δs the distance between two contiguous samples s. The continuous scale parameter in Equation (4.34) has been replaced by a dilation scaled by a power of two, and the translation by multiples of the distance Δs [*Frank*, 2006].

4.5.3 Multi-Resolution Analysis

The wavelets are suitable basis functions which can be exploited to extract information from a signal independently from its structure [*Daubechies*, 1992]. A further important concept regarding wavelets is the multi-resolution analysis (MRA), which is based on defining a nested set of vector spaces [*Keinert*, 2003; *Stollnitz et al.*, 1995]. Here, one signal can be split into a series of different levels - each of them with different details and resolution - by using wavelet decomposition. Figure 4.10 gives a pictorial representation of the multi-resolution concept. A full resolution image (for example with a maximum level 2, L2) is decomposed into two versions: level zero (L0) and level one (L1), and the number of the level corresponds to the value of a in Equation (4.34). The L1 level resolution is reduced by a factor of two in comparison to L0, as described in Equation (4.34).

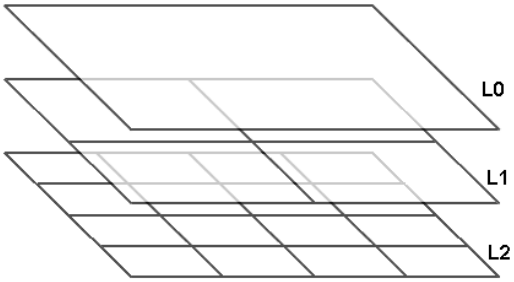

Figure 4.10. The multi-resolution decomposition allows decomposing the signal into different levels. Each level takes into account different details and resolutions.

Each level in Figure 4.10 can be seen as a coarser version of the level that comes after. In order to reconstruct the signal, the details which have been lost in this process have to be restored. Therefore, the wavelet transform has to be capable of separating the "averaged" or coarse version of a signal from the detailed part of the same signal. Thus, the wavelet basis consists of two sets of basis: the scaling function for the averaged version and the wavelet function for the detailed one.

As mentioned before, the nested subspace is the main concept used in MRA, which is defined as an infinite nested sequence of subspaces V_i with the following properties

$$\cdots \subset V_0 \subset V_1 \subset V_2 \subset V_3 \subset \cdots \tag{4.35}$$

The sum of all nested subspaces produces the full resolution version. Any subspace V_i represents the signal g with a different resolution and detail, and any function in V_{i+1} is a higher resolution approximation of the same function in V_i. In conjunction with subspace W_i from the wavelet function, the higher resolution is obtained as follows

$$V_i \oplus W_i = V_{i+1} \tag{4.36}$$

However, since wavelet and scaling subspaces are not distinct, the following identity is still valid

$$V_i = V_p \oplus \left(\bigoplus_{k=p}^{i-1} W_k \right) \tag{4.37}$$

This implies that W_k is also a subspace of V_i for $k < i$. The level i can be obtained by using the approximation V_p together with all the details at finer resolution [*Keinert*, 2003].

Figure 4.11 illustrates the MRA concept for a one dimensional signal. The top graph represents the signal g at its maximum resolution. Two figures are shown for each level of decomposition, the coarse (left) and detailed (right) part. Level 0 is the lowest resolution, which can be intended as the averaged value of the whole signal g. The finer part of the same

level is used to reconstruct the signal in its higher resolution version. By increasing the resolution the coarse part will also have better detail, while the finer one will have even more finer details.

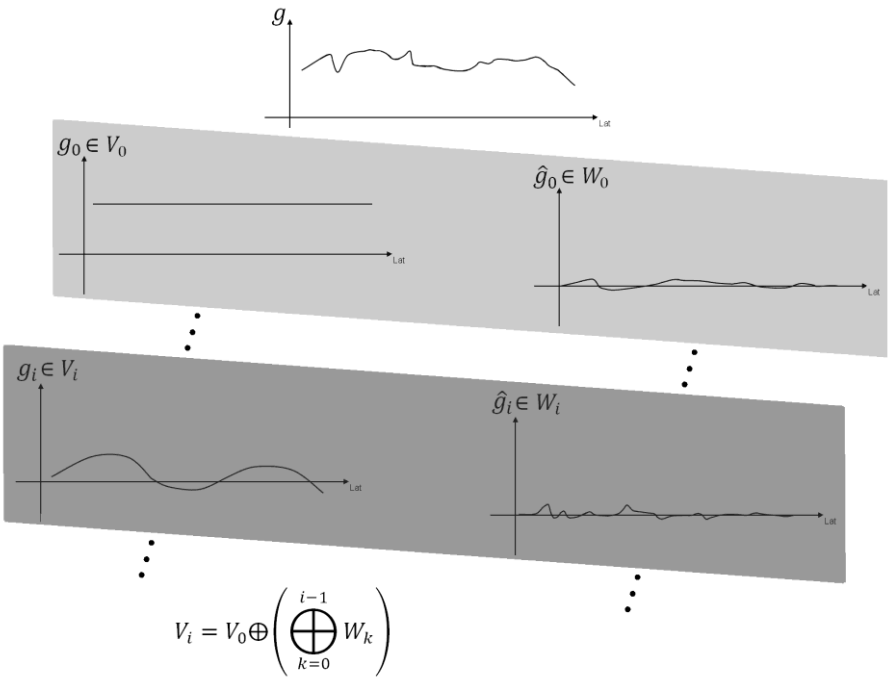

Figure 4.11. A signal g (top signal) can be split into different levels of approximation. Each level contains a different "version" or approximation of the signal, in terms of details as well as resolution. According to wavelet theory, it can be divided into a part which takes into account an averaged behaviour (the two left signals), and into a series of other versions which capture the details needed to describe the whole signal (the signals on the right). Each level of approximation can be reconstructed by means of the previous levels according to the formula $\mathbf{V_i = V_0 \oplus \left(\oplus_{k=0}^{i-1} W_k \right)}$.

The MRA is obtained with the Discrete Wavelet Transform (DWT), which is described according to scaling and wavelet functions. A common implementation of this is given by the following equations

$$g_{a-1}(b) = \sum_k h_1(k - 2b)\, g_a(k) \qquad g_{a-1}(b) \in V_{a-1} \tag{4.38}$$

$$d_{a-1}(b) = \sum_k h_2(k - 2b)\, g_a(k) \qquad d_{a-1}(b) \in W_{a-1} \tag{4.39}$$

The transformation consists of the convolution of translated versions (by a factor of b) between the recursion coefficients $h_1(k)$ and $h_2(k)$ and the signal g at the current resolution i. The equations above describe the decomposition of the signal g into two parts, a lower resolution g_{a-1} and the details d_{a-1}. The recursions coefficients are chosen according to the wavelet family that is used (e.g. Daubechies 4 or discrete Meyer). This implementation is commonly used in signal analysis because it can be implemented efficiently and with little computational complexity. For CIT a matrix representation (as in Equation (4.23)) of h_1 and

h_2 is more useful. The same decomposition, in fact, can be written according to Equation (4.37) by setting p equal to zero [*Stollnitz et al.*, 1995].

Therefore, MRA can be expressed through the following equation

$$\boldsymbol{z} = \mathbf{A}\boldsymbol{n} = \mathbf{A}\mathbf{K}_\mathbf{H}\boldsymbol{x} = \mathbf{A} \cdot [\boldsymbol{\varphi_1},\ \boldsymbol{\varphi_2},\ \boldsymbol{\varphi_3}, \ldots,\ \boldsymbol{\varphi_V}] \cdot \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_V \end{bmatrix} \tag{4.40}$$

where $\mathbf{K_H}$ is the matrix that contains the basis functions that describe the horizontal variation in the ionosphere, $\boldsymbol{\varphi_1}$ is the scaling function, and $\boldsymbol{\varphi_i}$ for $i = 2 \ldots V$ are the different translated and scaled versions of wavelets.

4.6 Summary

This chapter explained the theory of tomography, defined as an inverse problem, and how it can be applied to the ionosphere. In particular, the Tikhonov regularization was used to stabilize the inverse problem and produce a valid solution. The regularization uses different parameters and their effects in the solution were discussed with a simple simulation of a two-dimensional ionosphere. Furthermore, the concepts of vertical and horizontal basis functions and their use in decomposing the ionosphere were explained. Wavelet decomposition was then introduced, which allows the compact (or sparse) representation of structures in the ionosphere. This is the main concept used in this thesis. Sparsity is used to define a different regularization term, which depends on wavelets and their multi-resolution analysis. Chapter 5 will explain the use of a sparse regularization term together with 2-Dimensional wavelets, which will be used in Chapter 6 to describe the horizontal variation of the ionosphere. An extension of the method for 4-Dimensional ionospheric reconstructions will be described in Chapter 7, where 3-Dimensional wavelets will be used with the third dimension accounting for the time variation of the ionosphere.

Chapter 5
Sparse regularization techniques

Introduction

A key problem in ionospheric tomography is the sparse and inhomogeneous distribution of measurements. This is due to the lack of availability of receiver sites and the cost of deployment, which results in a set of observations that are non ideal in distribution. This chapter describes the theory of sparse regularization, and how it can be applied to the ionospheric problem. The algorithms used for the experiments are also discussed.

The sparsity of ground receivers was formerly addressed in [*M. Schmidt et al.*, 2008]. They observed that observations tend to have a clotted-like distribution, and that motivated the choice of using B-spline basis functions. B-splines are a family of wavelet basis functions, which are known for their properties of scalability, and for their ability to represent localized structures in a compact form. This is also known as MRA [*Mallat*, 2008], which defines the ability of the wavelets to describe structures at different scales and positions. In this thesis, wavelets were combined with a proper implementation of sparse regularization, fully embedded in the CIT algorithm. In fact, as discussed in [*D. L. Donoho*, 2006], the optimum combination of compactness and accuracy in the reconstruction cannot be obtained with simple threshold techniques, but requires a proper implementation of sparsity-constrained algorithms, which will be explained in this chapter and implemented on real and simulated case studies in Chapters 6 and 7. In this way, the sparse distribution of observations can be addressed more efficiently, as with the way structures are extrapolated from the observations. An efficient representation of the ionosphere in terms of wavelets allows the weakest structures to be extracted and better isolated from noise (i.e. artefacts in the reconstructed maps). Further advantages can be obtained with MRA: firstly, the uneven distribution of the observation can be tackled by the scalability of the wavelets. Secondly, the solution will appear as richer in detail as the number of observations allows it. This means that if the measurements are good enough in quality and number, the best-case reconstruction will look sharp and with a good level of detail. On the contrary, the solution will be represented without fine details and by using only the large scale basis functions. This is automatically obtained by selecting the most important coefficients that contribute to the reconstruction.

The sparse-regularization method with wavelets is believed to be its first implementation in CIT. It must be noted that sparse regularization is not limited to wavelets only, and any other basis function that can compactly represent the structures can be used.

Section 5.1 describes the inverse problem by including the biases and defines the final notation that will be used in Chapter 6 and Chapter 7. Section 5.2 discusses different regularization terms, in particular the sparse regularization. The implementation of the regularization technique is presented in Section 5.3, and in Section 5.4 a comparison of these different techniques is proposed with some simple examples. Conclusion and final discussion are in Section 5.5.

5.1 The calibration of relative measurements in CIT

The CIT problem described so far has intentionally neglected the offset that affect GPS measurements (see Chapter 2 for more details) in order to not overload the notation and the amount of information discussed. Therefore, Equation (3.23) needs to include the offsets $\boldsymbol{b}$, and becomes

$$\boldsymbol{z} = \mathbf{A}\mathbf{K}\boldsymbol{x} + \mathbf{B}\boldsymbol{b} \tag{5.1}$$

where **A** is the projection matrix that maps the electron content into measurements $\boldsymbol{z}$, and depends on the geometry of the problem, and **K** is the matrix containing the vertical and horizontal basis functions as described in Chapter 4. The projection matrix **B** maps the offset of each ray (observation) into a single offset for each receiver-satellite pair and is defined as

$$\mathbf{B}_{ij} = \begin{cases} 1 & \text{if } b_j \text{ is the offset of } z_i \\ 0 & \text{otherwise} \end{cases} \tag{5.2}$$

Therefore, the solution associated to the inverse problem is calculated as

$$\left(\hat{\boldsymbol{x}}, \hat{\boldsymbol{b}}\right) = \min_{(\boldsymbol{x},\boldsymbol{b})} f(\boldsymbol{x}, \boldsymbol{b}) \tag{5.3}$$

where $\hat{\boldsymbol{x}}$ and $\hat{\boldsymbol{b}}$ are the pseudo-solutions of the basis function coefficients and the offsets respectively, and

$$f(\boldsymbol{x}, \boldsymbol{b}) = \|\boldsymbol{z} - \mathbf{A}\mathbf{K}\boldsymbol{x} - \mathbf{B}\boldsymbol{b}\|^2 + \alpha\, \mathcal{P}(\boldsymbol{x}) \tag{5.4}$$

The function $\mathcal{P}(\boldsymbol{x})$ defines the regularization (or penalty) term that makes the inversion a well-posed problem. The regularization term is written with the general notation $\mathcal{P}(\boldsymbol{x})$ in order to associate it to different implementations (for example Tikhonov regularization). The parameter α sets a trade-off between the best fitting and the most reasonable stabilization [*Zhdanov*, 2002].

$\mathcal{P}(\boldsymbol{x})$ is supposed to operate on $\boldsymbol{x}$ only and not on the offsets $\boldsymbol{b}$, which will not be constrained by any assumptions coming from the regularization. The parameter α sets a trade-off between the best fitting and the most reasonable stabilization [*Zhdanov*, 2002]. This justifies the different notation of $\widehat{\boldsymbol{x}}$ in order to distinguish the approximation from the true $\boldsymbol{x}$.

The functional $f(\boldsymbol{x}, \boldsymbol{b})$ is highly computationally expensive and, therefore, is not practically useful.

By expanding Equation (5.3) and Equation (5.4) it is possible to write (see Appendix B)

$$f(\boldsymbol{x}) = \|\boldsymbol{z} - \mathbf{A}\mathbf{K}\boldsymbol{x}\|_{\mathbf{C}}^2 + \alpha\, \mathcal{P}(\boldsymbol{x}) \tag{5.5}$$

Where the calibration matrix $\mathbf{C}$ is formed from Laplacian matrices and is defined as

$$\mathbf{C} = \mathbf{I} - \mathbf{B}\left(\mathbf{B}^{\mathbf{T}}\mathbf{B}\right)^{-\mathbf{1}}\mathbf{B}^{\mathbf{T}} \tag{5.6}$$

A similar approach is described in [*Spencer and Mitchell*, 2011]. The differences are that in Equation (5.6) an explicit relationship between the estimated biases $\widehat{\boldsymbol{b}}$ and the reconstruction $\widehat{\boldsymbol{x}}$ is provided, which is shown later in Equation (5.8). The solution of Equation (5.5) is coincident with the solution that would be obtained from Equation (5.4).

The solution $\widehat{\boldsymbol{x}}$ is obtained by minimizing Equation (5.5) over $\boldsymbol{x}$, given that $\mathcal{P}(\boldsymbol{n})$ is differentiable. Thus,

$$\widehat{\boldsymbol{x}} = \left(\mathbf{A}^{\mathrm{T}}\mathbf{K}^{\mathrm{T}}\mathbf{C}\mathbf{A}\mathbf{K} + \alpha\, \frac{1}{2}\frac{\partial \mathcal{P}(\boldsymbol{x})}{\partial \boldsymbol{x}}\right)^{-1} \mathbf{K}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{C}\, \boldsymbol{z} \tag{5.7}$$

The offsets $\widehat{\boldsymbol{b}}$ can then be recovered as

$$\widehat{\boldsymbol{b}} = \left(\mathbf{B}^{\mathrm{T}}\mathbf{B}\right)^{-1}\mathbf{B}^{\mathrm{T}}(\boldsymbol{z} - \mathbf{A}\mathbf{K}\widehat{\boldsymbol{x}}) \tag{5.8}$$

The observations $\boldsymbol{z}$ generally have a negligible noise term, but in the presence of ionospheric structures and because the ionosphere is not a static medium, there could be small variations in sTEC even between nearby ray paths. Therefore, the discretization of the ionosphere into a grid is important, which causes the measure $\|\boldsymbol{z} - \mathbf{A}\mathbf{K}\boldsymbol{x}\|^2$ to never reach zero; allowing a representativity error due to the complexity of the medium to be acceptable. The minimization of Equation (5.5) aims to have the reconstruction matching the observations (where data are available) up to a residual noisy term. Therefore, the regularization term $\mathcal{P}(\boldsymbol{x})$ becomes the most important term, which will be discussed in Section 5.2.

5.2 The regularization term

In this section, the regularization term is described in more detail by focusing in particular on the theory of sparse regularization. Recently, the sparse regularization has become an interesting alternative to the standard Tikhonov approach (see e.g.[*D. L. Donoho et al.*, 2006]). This is due to its property of uniqueness of the solution and the advantages that a sparse solution implies in fields like data compression, signal extraction and noise removal. The theory is explained by using the general definition of the norm to express the measure of the quantity that is used in the regularization term. This term is used to stabilize and make the solution unique.

The regularization term $\mathcal{P}(\boldsymbol{x})$ described in Section 5.1 takes into account a measure that is used to regularize the inverse problem associated with Equation (5.5). The regularization term can be expressed in different ways. A proper choice of $\mathcal{P}(\boldsymbol{x})$ makes Equation (5.5) convex, which guarantees the existence of a global minimum. In general terms, the regularization can be seen as the measure of a distance of the current solution from the one that has minimum length. This length or distance can be computed in different ways, with different norms. The general definition of norm can be written as

$$\|\boldsymbol{a}\|_p = \left(\sum_i \|a_i\|^p\right)^{1/p} \tag{5.9}$$

where $p = 0,1,2 \ldots$ identifies the different ℓ_p-norm, and a_i is an element of the vector $\boldsymbol{a}$. Special cases are the ℓ_0-norm

$$\|\boldsymbol{a}\|_{\mathbf{0}} = \#(\boldsymbol{i}|\boldsymbol{a_i} \neq \mathbf{0}) \tag{5.10}$$

where # represents the number of elements, and the ℓ_∞-norm

$$\|\boldsymbol{a}\|_\infty = \mathbf{max}_{\boldsymbol{i}}(\|\boldsymbol{a_i}\|) \tag{5.11}$$

5.2.1 ℓ_2 regularization (or Tikhonov regularization)

The most classical approach to regularize the inverse problem associated with Equation (5.5) is by using an ℓ_2 norm (or Tikhonov regularization)

$$\mathcal{P}(\boldsymbol{x}) = \|\mathbf{PK}\boldsymbol{x}\|_2^2 \tag{5.12}$$

where matrix **P** is used to select only a subspace of the possible solutions and stabilize the solution of Equation (5.7) toward a physically acceptable solution. The ℓ_2 norm is defined as $\|\boldsymbol{a}\|_2 = \sqrt[2]{\sum_i |a_i|^2}$. The matrix **P** can be simply set to the identity matrix (see, e.g.[*C.N.*

Mitchell and Spencer, 2003]). If the basis functions are orthonormal, the energy in the transformed domain is maintained and then Equation (5.12) can be written as

$$\mathcal{P}(\boldsymbol{x}) = \|\boldsymbol{x}\|_2^2 \tag{5.13}$$

Figure 5.1 represents a two dimensional sketch (based on [*D. L. Donoho et al.*, 2006; *Shi et al.*, 2013]) of how the ℓ_2 regularization finds the minimum solution

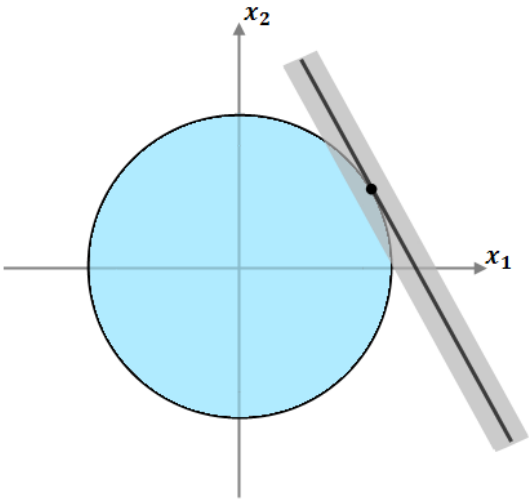

Figure 5.1. A two dimensional pictorial representation of the minimum solution for the ℓ_2 regularization. The ℓ_2 ball is a circle that corresponds to the points with the same norm $\|x\|_2^2$. The grey and thick line represents all possible solutions that satisfies $\|\boldsymbol{z} - \mathbf{AK}\boldsymbol{x}\|_{\mathbf{C}}^2 < \delta$. The unique solution is found at the intersection of the circle with the line.

In two dimensions, the ℓ_2 ball corresponds to a circle. It represents all the points with the same ℓ_2-norm $\|\boldsymbol{x}\|_2^2$ from Equation (5.13). The tick grey line represents all possible solutions which satisfy

$$\|\boldsymbol{z} - \mathbf{AK}\boldsymbol{x}\|_{\mathbf{C}}^2 < \delta \tag{5.14}$$

The black line, instead corresponds to all possible solutions equals to a fixed value of δ (e.g. zero). The final solution is found at the intersection of the black line with the ℓ_2 ball, and is described with a linear combination of $\boldsymbol{x_1}$ and $\boldsymbol{x_2}$. It corresponds to the minimum-length solution with the minimum residual. Figure 5.1 supposes that an exact solution is found, i.e. the left-hand term of the inequality of Equation (5.14) equals to zero. In this case the solution is defined as the vector $\boldsymbol{x}$ that satisfies the following problem

$$\|\boldsymbol{x}\|_2^2 \text{ subject to } \|\boldsymbol{z} - \mathbf{AK}\boldsymbol{x}\|_{\mathbf{C}}^2 = \delta \tag{5.15}$$

Therefore the real solution is illustrated by the black line in Figure 5.1. It must be noticed that the definition of the problem as in Equation (5.15) is different from the one defined in

Equation (5.14). However, a proper selection of the parameter α makes the two problems equivalent [*Bruckstein et al.*, 2009].

The identity matrix used here is the simplest form that can be used for the matrix **P**. Another suitable choice is the Laplacian matrix. In this case Equation (5.12) can be written as

$$\mathcal{P}(\boldsymbol{x}) = \|\nabla^2 \mathbf{K}\boldsymbol{x}\|_2^2 \tag{5.16}$$

This choice will favour the smoothness in the solution, preventing the solution having rapid oscillations.

5.2.2 ℓ_0 regularization (or sparse regularization)

Another interesting approach for regularization is obtained by promoting the solution to be sparse in terms of the number of basis function coefficients. This assumes that the basis functions can compactly represent the structures that characterize (in CIT) the ionosphere. A sparse solution, where only relatively few coefficients are the most important, has significant advantages. It allows the simplification of the problem, increasing the compression of data, and improving the noise removal from the data [*Tsaig and Donoho*, 2006]. The problem was initially studied by a number of authors (for example [*Natarajan*, 1995] and [*Daubechies et al.*, 2004]), where they described how overcomplete dictionaries (a big set of basis functions) may be more effective in areas regarding the compression or the extraction of a signal from a noisy source [*D. L. Donoho et al.*, 2006]. A dictionary is overcomplete if by removing some basis functions (or atoms) the signal is still well-represented with minimum or no distortion. That is, the signal can be represented just with a minimum number of basis functions.

This is equivalent to solving Equation (5.5) with the following regularization term,

$$\mathcal{P}(\boldsymbol{x}) = \|\boldsymbol{x}\|_0 \tag{5.17}$$

where $\|\boldsymbol{a}\|_0$ stands for the number of coefficients a_i that are not zero. By minimizing Equation (5.5) with Equation (5.17) not only the solution that produces the best agreement with the observations is found, but also the sparsest one at the same time. Unfortunately, the regularization of Equation (5.17) does not make the problem convex and, therefore, the solution may be affected by local minima. Furthermore, the complexity of this problem was also proven to be in general not computationally practical as the solution requires to exhaustively search for all the possible combinations of basis functions (the columns in **K**) that minimize the functional of Equation (5.5) [*Natarajan*, 1995].

5.2.3 ℓ_1 regularization for sparse regularization

The convex relaxation of Equation (5.17) is obtained with the ℓ_1 norm, which makes the problem described through Equation (5.5) more tractable. This choice comes naturally as it is

a somewhat half-way between the ℓ_0 and ℓ_2 norm. In this case, the regularization term can be written as

$$\mathcal{P}(\boldsymbol{x}) = \|\boldsymbol{x}\|_1 \tag{5.18}$$

where $\|\boldsymbol{a}\|_1 = \sum_i |a_i|$. It can be shown that the solution obtained through Equation (5.18), under certain hypotheses, is as sparse as the solution obtained with Equation (5.17) [*D. L. Donoho*, 2006; *Tropp*, 2004]. The theory requires that the dictionary **K** is incoherent, i.e. the coherence *M* is small, where

$$M = \max_{i \neq j} \left\| \langle \boldsymbol{\Phi}_\mathrm{i}, \boldsymbol{\Phi}_\mathrm{j} \rangle \right\| \tag{5.19}$$

$\boldsymbol{\Phi}_\mathrm{i}$ represents the i-th column of **K,** and $\langle .,. \rangle$ the inner product. Equation (5.19) considers that basis functions are normalized to one. Furthermore, $\boldsymbol{x}$ needs to be sufficiently sparse. That is, in CIT, the ionosphere can be represented sparsely with few basis functions. If those conditions are satisfied, then the solution is both the sparsest and the minimal ℓ_1 solution [*D. L. Donoho*, 2006], which means that the solution is unique.

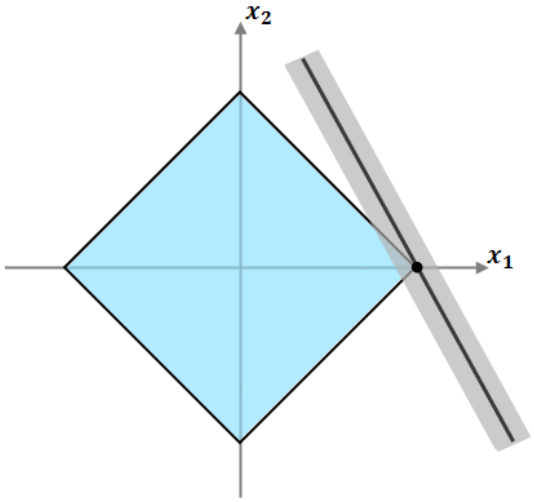

Figure 5.2. Two dimensional pictorial representation of the minimum solution for the ℓ_1 regularization. The ℓ_1 ball is a diamond-shape that corresponds to the points with the same norm $\|x\|_1$. In constrast, the grey and thick line represents all possible solutions that satisfies $\|\boldsymbol{z} - \mathbf{K}\boldsymbol{x}\|_\mathbf{C}^2 < \boldsymbol{\delta}$. The unique solution is found at the intersection of the diamond-shape and the line. The intersection is likely to be at the edges of the diamond-shape, making the solution sparse.

Figure 5.2 illustrates why the solution tends to be sparse with ℓ_1 regularization (based on [*D. L. Donoho et al.*, 2006; *Shi et al.*, 2013]). The ℓ_1 space is represented by the diamond-like shape. The points lying on the ℓ_1 diamond correspond to the values with the same norm $\|\boldsymbol{x}\|_1$.

The grey thick line, instead, corresponds to the points for which the first term $\|\boldsymbol{z} - \mathbf{AK}\boldsymbol{x}\|_{\mathbf{C}}^2$ of Equation (5.5) is constant and smaller than a certain quantity δ. The black line shows a solution with a fixed value of δ (e.g. zero). Finally, the solution is found at the intersection of the black line with the ℓ_1 diamond. The solution is described with a linear combination of $\boldsymbol{x_1}$ and $\boldsymbol{x_2}$. However, because of the diamond-shape of the ℓ_1 space, only one coefficient is needed to represent the solution. This shows, in a very simple way, the potential of the ℓ_1 regularization to produce a sparse solution. It is supposed in Figure 5.2 that an exact solution is found, and, as for the ℓ_2 regularization, the problem can be defined as in Equation (5.15).

Compared to Figure 5.1 where the solution is more likely to lie far from the axis, Figure 5.2 illustrates that, with ℓ_1 regularization the solution tends to lie on one of the axes (in this case $\boldsymbol{x_1}$); as the intersection occur in one of the edges of the ℓ_1 diamond. Thus, the solution is more likely to be sparse than the one obtained with the ℓ_2 regularization.

It must be noted that a perfect reconstruction is not always possible. This is mainly due to the errors and approximations that affect the measurements. In [*D. L. Donoho et al.*, 2006], it is shown that, even if the measurements are affected by errors and therefore the residual norm $\|\boldsymbol{z} - \mathbf{AK}\boldsymbol{x}\|_{\mathbf{C}}^2$ does not reach zero, the ℓ_1 regularization still produces a unique solution which is also the sparsest one. Furthermore, under appropriate conditions, it is also shown that the estimation will contain only basis functions that appear in the ideal sparse representation. The error in the estimated solution can then be quantified to be, at worst, proportional to the input noise level.

The theory described in this section assumes that a unique solution exists, and therefore that the source is sufficiently sparse in, for example, the wavelet domain. In the case of CIT, this may not be easy to demonstrate as a real image of the true ionosphere is unknown. A study from [*Adcock et al.*, 2013] tried to investigate this issue from the mathematical point of view. Most of the mathematical assumption, in fact, tends to fail or to be difficult to prove for real case scenarios. For this purpose, they introduced the concept of asymptotical sparsity, which claims that any signal source is asymptotically sparse as the information used to describe it increases. This means that, as the resolution (or sampling) increases, the signal becomes sparser. Conceptually, a higher resolution allows better definition of finer details. At a certain point the signal will not show details smaller than a certain scale. Therefore, at this point, increasing the resolution will not improve the reconstruction of the signal, although it will increase its sparsity as only a small number of small-scale basis functions will contribute at finer scales. This means that the reconstruction will look better as the resolution increases but, also, that a lot of measurements are needed to reconstruct the signal at high resolution; which may not be feasible in practical cases (e.g. due to costs). Fortunately, the concept of

compressing sensing states that, if properly described, the inverse problem can be solved at the same resolution using fewer measurements. Further, more important is that the sampling of measurements does not need to be uniform and follow the Nyquist theorem. The theories of compressing sensing and asymptotical sparsity are not used in this thesis, but some of the concepts can be seen in the results shown in the following chapters. Furthermore, these theories give a mathematical support that could justify the future use of sparse regularization for imaging the ionosphere at high resolution.

5.2.4 Total Variation

The ℓ_1 norm can be used similarly to the ℓ_2 norm to minimize the rate of change of the gradient. This is called Total Variation (TV)

$$\mathcal{P}(\boldsymbol{x}) = \|\boldsymbol{x}\|_{\mathrm{TV}} \tag{5.20}$$

The TV norm is used to minimize the number of variations in the gradient. In the case of a two dimensional matrix $\boldsymbol{x}$ of size $N \times N$, TV is defined as [*Chambolle*, 2004]

$$\mathrm{TV} = \sum_{1 \le i,j \le N} \left\| (\nabla \boldsymbol{x})_{i,j} \right\|_1 = \sum_{1 \le i,j \le N} \left\| \left((\nabla \boldsymbol{x})^1_{i,j}, (\nabla \boldsymbol{x})^2_{i,j} \right) \right\|_1 \tag{5.21}$$

where

$$\begin{aligned} (\nabla \boldsymbol{x})^1_{i,j} &= \begin{cases} \boldsymbol{x}_{i+1,j} - \boldsymbol{x}_{i,j} & if\ i < N \\ 0 & if\ i = N \end{cases} \\ (\nabla \boldsymbol{x})^2_{i,j} &= \begin{cases} \boldsymbol{x}_{i,j+1} - \boldsymbol{x}_{i,j} & if\ j < N \\ 0 & if\ j = N \end{cases} \end{aligned} \tag{5.22}$$

This is different from minimizing the Laplacian with the ℓ_2 norm, as in this case, the minimization will encourage the solution to have sharp gradients rather than promoting a smooth solution. In fact, TV is used to extract the signal from noisy data by maintaining the edges of the structures.

5.3 Implementation of the algorithms

The previous sections described the different norms that can be used to regularize the problem associated with Equation (5.5). In particular the ℓ_1 norm was discussed, including the TV norm. In this section, the algorithms that were used in this thesis to solve Equation (5.5) are described. Different algorithms exist to implement the same inversion techniques with different regularizations. As the comparison of the performance of different algorithms is not part of this dissertation, only the algorithms actually used in this thesis are described.

The inverse problem with $\mathcal{P}(\boldsymbol{x}) = \|\boldsymbol{x}\|_2^2$ is solved by using the MINRES algorithm of MATLAB. This is a standard algorithm and more information can be found in [*Barrett et al.*, 1994; *Choi and Saunders*, 2014; *Paige and Saunders*, 1982]. This section focuses on the description of the algorithm for sparse regularization.

In the case of ℓ_1 regularization, i.e. $\mathcal{P}(\boldsymbol{x}) = \|\boldsymbol{x}\|_1$, the Fast Iterative Shrinkage-Thresholding Algorithm (FISTA) was used. This is an elegant implementation that requires a single parameter γ and it was proposed in [*A. Beck and M. Teboulle*, 2009]. It is based on the Iterative Shrinkage-Thresholding Algorithm (ISTA) algorithm [*Daubechies et al.*, 2004] but with a proven faster convergence rate. It is an iterative gradient descendent algorithm where the main step is given by a nonlinear soft-threshold or shrinkage operator η_γ [*D.L. Donoho*, 1992; *D. L. Donoho and Johnstone*, 1994b]

$$\eta_\gamma(\hat{\boldsymbol{x}}_i) = \begin{cases} 0 & if\ |\hat{\boldsymbol{x}}_i| \leq \gamma \\ sgn(\hat{\boldsymbol{x}}_i)(|\hat{\boldsymbol{x}}_i| - \gamma) & if\ |\hat{\boldsymbol{x}}_i| > \gamma \end{cases} \tag{5.23}$$

The shrinkage is used to remove coefficients that are too small and considered to be under the level of uncertainty given by γ. In the case of Gaussian noise it has been demonstrated that the optimum threshold is given by [*D. L. Donoho and Johnstone*, 1994a; *Leadbetter et al.*, 1983]

$$\gamma_{Opt} = \epsilon\sqrt{2\log N} \tag{5.24}$$

where ϵ represents the noise level and N the length of the signal. In general, the parameter γ sets the trade-off between the best fitting and the sparsest solution.

The algorithm for FISTA is reported below [*Amir Beck and Marc Teboulle*, 2009; *Loris et al.*, 2010]

0. Set τ to $1/\max\left(\text{eig}\left(\mathbf{K}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{A}\mathbf{K}\right)\right)$
1. Set $t_1 = 1$, $\boldsymbol{y}_1 = \boldsymbol{x}_0$ (initial estimation)
2. For $j \geq 1$
3. $\quad \boldsymbol{x}_j = p_\tau(\boldsymbol{y}_j)$
4. $\quad$ Select the solution that minimize $\left(f(\boldsymbol{x}_j), f(\boldsymbol{x}_{j-1})\right)$
5. $\quad \boldsymbol{t}_{j+1} = \frac{1+\sqrt{1+4\boldsymbol{t}_j^2}}{2}$
6. $\quad \boldsymbol{y}_{j+1} = \boldsymbol{y}_j + \left(\frac{\boldsymbol{t}_j - 1}{\boldsymbol{t}_{j+1}}\right)(\boldsymbol{x}_j - \boldsymbol{x}_{j-1})$

where τ is used to set the threshold in the shrinkage operator, and

$$p_\tau(\boldsymbol{y}_j) = \eta_{\alpha\tau}\left(\boldsymbol{y}_j + \tau\mathbf{K}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}(\boldsymbol{z} - \mathbf{A}\mathbf{K}\boldsymbol{y}_j)\right) \tag{5.25}$$

The last method that is discussed is the Total Variation (TV). The same authors of FISTA described an algorithm to implement the TV [*A. Beck and M. Teboulle*, 2009]. As for FISTA, this is a gradient-based algorithm with a proven global rate of convergence and a simple implementation. It requires, as for FISTA, a single parameter α. The correct value of α is the one that reduces noisy oscillations (typical in inverse problems), but leaves or enhances the gradients in the reconstruction. The algorithm is based on FISTA, with the difference that at each j-th iteration, the subproblem p_τ is solved with the Fast Gradient Projection (FGP) algorithm

$$p_\tau(\boldsymbol{y}_n) = \text{FGP}(\boldsymbol{b}, \gamma) = \text{FGP}\left(\boldsymbol{y}_n - \frac{2}{\tau}\mathbf{K}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}(\mathbf{A}\mathbf{K}\boldsymbol{y}_n - \boldsymbol{z}), \frac{2\alpha}{\tau}\right) \tag{5.26}$$

The FGP algorithm is reported below

1. Algorithm for FPG($\boldsymbol{b}, \gamma$)
2. Set $(\boldsymbol{r}_1, \boldsymbol{s}_s) = (\boldsymbol{p}_0, \boldsymbol{q}_0) = \underline{\mathbf{0}}$ and $\boldsymbol{t}_1 = 1$
3. For $j \geq 1$
4. $\quad (\boldsymbol{p}_j, \boldsymbol{q}_j) = P_P\left[(\boldsymbol{r}_j, \boldsymbol{s}_j) + \frac{1}{8\gamma}\mathcal{L}^T\left(\boldsymbol{b} - \gamma\mathcal{L}(\boldsymbol{r}_j, \boldsymbol{s}_j)\right)\right]$
5. $\quad \boldsymbol{t}_{j+1} = \frac{1+\sqrt{1+4\boldsymbol{t}_j^2}}{2}$
6. $\quad (\boldsymbol{r}_{j+1}, \boldsymbol{s}_{j+1}) = (\boldsymbol{p}_j, \boldsymbol{q}_j) + \left(\frac{\boldsymbol{t}_k - 1}{\boldsymbol{t}_{k+1}}\right)(\boldsymbol{p}_j - \boldsymbol{p}_{j-1}, \boldsymbol{q}_j - \boldsymbol{q}_{j-1})$
7. Set $\boldsymbol{x} = \boldsymbol{b} - \gamma\mathcal{L}(\boldsymbol{r}_j, \boldsymbol{s}_j)$

The projection P_P defines the mapping of the pairs $(\boldsymbol{p}, \boldsymbol{q})$ into the set with pairs $(\boldsymbol{r}, \boldsymbol{s})$ that satisfies

$$\|r_{ij}\| \leq 1 \; \forall i, j \tag{5.27}$$

$$\|s_{ij}\| \leq 1 \; \forall i, j \tag{5.28}$$

The projection described in the equations above can be easily implemented as

$$r_{ij} = \frac{p_{ij}}{\max\{1, \|p_{ij}\|\}} \tag{5.29}$$

$$s_{ij} = \frac{q_{ij}}{\max\{1, \|q_{ij}\|\}} \tag{5.30}$$

5.4 Comparison of the algorithms

This section discusses the performance of the different algorithms described previously. A simple one dimensional case is used to test the efficiency of the algorithms in the scenarios of noisy and small number of observations. The uneven distribution and small number of observations are simulated by random undersampling of the signal. A Gaussian noise term is then added.

Figure 5.3 shows the artificial signal $\boldsymbol{n}$ used to simulate the observations $\boldsymbol{z}$.

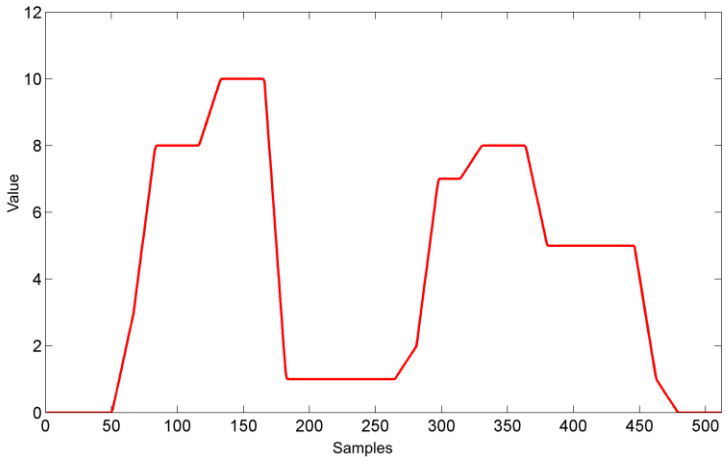

Figure 5.3. Artificial signal used to compare the different regularization terms described in this chapter.

The observations $\boldsymbol{z}$ are obtained by

$$\boldsymbol{z} = \mathbf{H}\boldsymbol{n} + \boldsymbol{\varepsilon} \tag{5.31}$$

where **H** is the projection matrix that is created by emulating a simple blurring effect defined by a Gaussian-shaped function, and $\boldsymbol{\varepsilon}$ is the noise vector described by a Gaussian distribution with zero mean and a standard deviation of two.

The matrix **K** of the basis functions is created by using the wavelet families Daubechies 4 (DB4) and Haar (HA). The inverse problem is defined as the minimization of

$$\hat{\boldsymbol{x}} = \min_{\boldsymbol{x}}\big(F(\boldsymbol{x})\big) \tag{5.32}$$

$$F(\boldsymbol{x}) = \|\boldsymbol{z} - \mathbf{A}\mathbf{K}\boldsymbol{x}\|_{\mathbf{C}}^2 + \alpha\,\mathcal{P}(\boldsymbol{x}) \tag{5.33}$$

where $\boldsymbol{x}$ is the vector containing the coefficients of the basis functions. It is defined as $\boldsymbol{x} = \mathbf{K}^{\mathrm{T}}\boldsymbol{n}$.

Some observations in $\boldsymbol{z}$ are then removed to represent a simple simulation of some limitations that are common in ionospheric tomography, in particular the uneven distribution of observations.

As mentioned before, two different families of wavelet basis functions are used: DB4 and HA. The different characteristics of these two wavelets give an insight of the importance of choosing the basis functions. Haar basis functions, for example, have sharp edges and little support. Therefore, they are suitable for representing strong gradients and very localized structures. In contrast, DB4 have a smoother shape with a slightly larger support which make them more suitable for representing structures with a gradual variation. In some aspect, DB4 is closer to the real world, where the variations are described continuously in time.

Figure 5.4 shows the reconstruction obtained with the ℓ_2 regularization. The solution looks smooth and the global trend of the profile is reconstructed. The purple sections in the image show the regions with no data coverage.

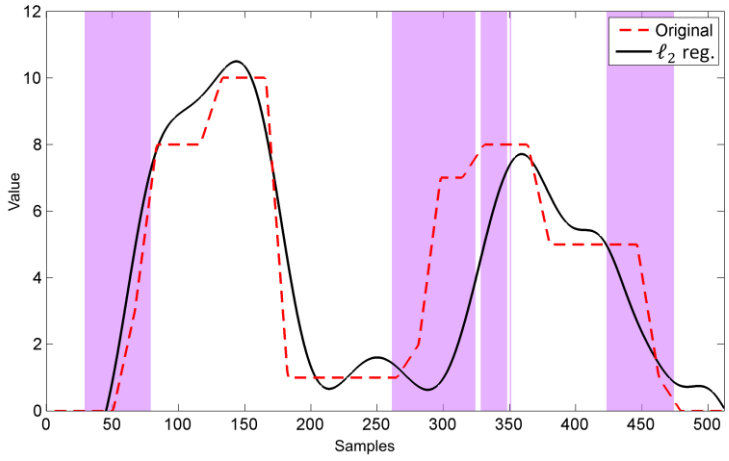

Figure 5.4. The reconstruction obtained with the ℓ_2 regularization (black) and the original signal (red) is shown in this figure. The regions with no data coverage are indicated in purple.

The ℓ_1 regularization reconstruction is illustrated in Figure 5.5. It was generated from two different basis functions: DB4 (black line) and Haar (green line).

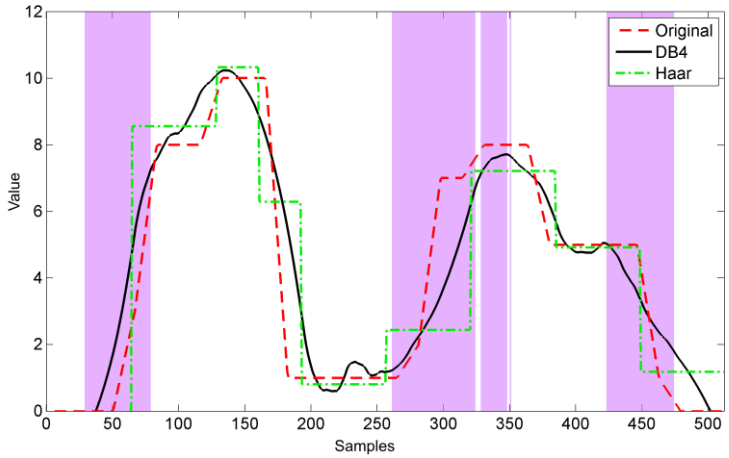

Figure 5.5. The reconstruction obtained with the ℓ_1 regularization using DB4 (black) and Haar (green) and the original signal (red) is shown in this figure. The regions with no data coverage are indicated in purple.

It can be observed that some of the variations are better recovered, especially in regions where data is missing. Furthermore, the position of the second structure is also better reconstructed. Thus, sparse regularization seems to better exploit the information than the ℓ_2 regularization, even for a small region such as the range 348-350 of the x-axis. This is due to the localization property of the wavelets and explains why the edge of the second structure is better recovered. As we can be seen, HA produces a more staircase-like reconstruction than by using DB4.

The same case study was tested with TV regularization. The TV algorithm depends strongly on the initial estimation of the solution $\boldsymbol{x}_0$ that is passed as input. Therefore, earlier estimations obtained from the ℓ_1 and ℓ_2 cases were used as input for TV. Figure 5.6 and Figure 5.7 show the reconstructions obtained with TV. The same threshold was used for α as with the previous reconstructions.

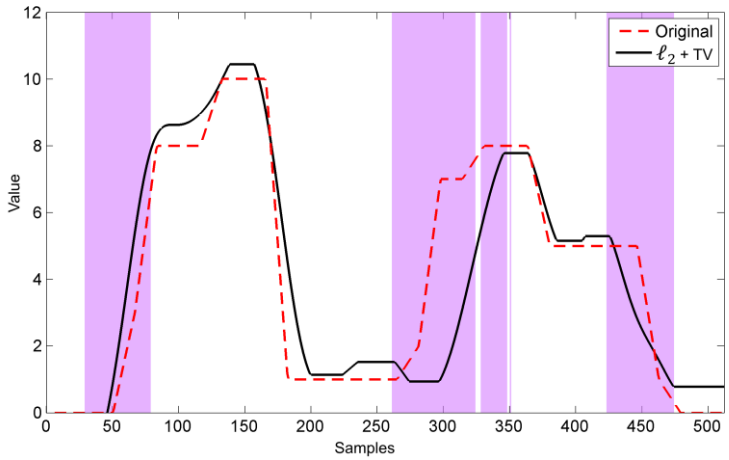

Figure 5.6. The reconstruction obtained with TV regularization (black) and the original signal (red) is shown in this figure. The estimation from the ℓ_2-regularized solution was used as input for the TV algorithm. The regions with no data coverage are indicated in purple.

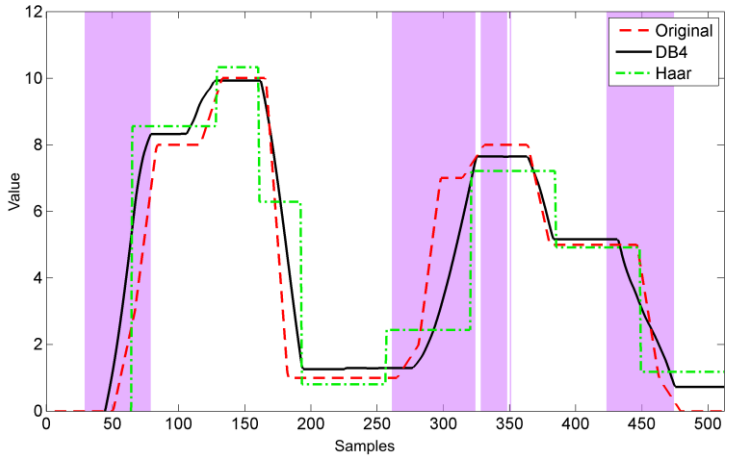

Figure 5.7. The reconstruction obtained with TV regularization (black and green) and the original signal (red) is shown in this figure. The estimations from the ℓ_1-regularized solution using DB4 (black) and Haar (green) were used as input for the TV algorithm. The regions with no data coverage are indicated in purple.

It can be seen from the figures that the small oscillations are removed and the gradients are better recovered. Figure 5.6 shows an incorrect estimation of the position of the second structure. This is due to the ℓ_2-regularized solution that did not estimate the same position correctly.

In general, TV regularization works particularly well when the signal to be recovered has strong gradients. However, it tends to create a staircase-like reconstruction. The effect of TV regularization in the case of Haar basis function is minimum. Indeed, the reconstruction using Haar can, in some ways, be seen as a TV regularization.

It must be noted that, the inverse problem in a real ionospheric scenario is actually more complicated that the one described in this simple one-dimensional case study. The aim of the comparison was to show the advantages of sparse and TV regularization over the standard Tikhonov regularization. In a real ionospheric scenario, many limitations are imposed by the geometry of the rays and ground receiver distribution. The observations are themselves an estimation of the real quantities that we aim to measure. Furthermore, the need of a priori information to estimate the vertical profiles of the electron density in regions with no data coverage makes also the real CIT problem very difficult to solve. Nevertheless, knowing the performance of sparse and TV regularization in a controlled environment can help the interpretation of the results shown in a real ionospheric case study.

5.5 Summary

This chapter presented the mathematical notations used to describe different regularization terms that were used to stabilize the inversion. In particular, the sparse and Total Variation (TV) regularizations were described as well as the solution to calibrate sTEC observations within the operation of inversion. The algorithms and their implementations were also shown,

together with a comparison on a simple one dimensional case. The results indicate sparse regularization as a valid alternative to the standard ℓ_2 regularization. Furthermore, the TV regularization showed the potential to better recover the gradients of the structures. The implementation of sparse regularization on CIT is discussed in Chapter 6 using simulated data, and in Chapter 7 using real data. TV regularization is discussed in Chapter 6.

Chapter 6
Simulation and first results

Introduction

Computerized Ionospheric tomography (CIT) has become in last decades an important tool for understanding and studying the ionosphere, its behaviour and its effects on radio propagation. Plasma structures can disrupt radiowave signals and, in the case of GNSS can reduce the accuracy of the pseudorange measurements. The plasma density is generally measured in terms of sTEC from double frequency measurements of the pseudoranges (see Chapter 2).

A series of limitations (e.g. limited angle geometry of the observations) and the uneven and sparse distribution of the observations on the Earth make CIT an underdetermined problem. These limitations are described, for example, in [*Yeh and Raymund*, 1991] and [*H Na and Lee*, 1992], while the uneven distribution of the observations is addressed in [*Michael Schmidt*, 2007]. Therefore, the regularization is needed to reduce artefacts and noise in the reconstruction due to lack of data.

In this chapter two methods for the imaging of the ionosphere are described. They are based on two different regularization techniques. In particular, the benefits of the multi-resolution analysis, implemented through wavelets with the sparse regularization, are illustrated in comparison with a standard approach based on Tikhonov regularization. The two different approaches are in some way tied with the basis functions that are used to describe the horizontal variation of the ionospheric structures. In particular, wavelet basis functions are used for sparse regularization, whilst spherical harmonics are used for Tikhonov regularization. Spherical harmonics are a standard choice in ionospheric tomography (see for example [*C.N. Mitchell and Spencer*, 2003]). More details about the regularization terms can be found in Chapter 5. The vertical distribution of the electron density needs, instead, a different representation. For this purpose, EOFs are employed to represent the vertical profile of the electron density. EOFs are used to compensate the lack of vertical information due to the geometry of the observations and to produce a physically meaningful electron density profile in the reconstruction (see Chapter 4 for further details).

At the time of this thesis, sparse regularization has not been used in ionospheric tomography yet. Nevertheless, results from other fields make this technique particularly interesting (e.g.[*Loris et al.*, 2010; *Simons et al.*, 2011]). Instead, wavelet basis functions like Haar

[*Amerian et al.*, 2010] and B-splines [*Durmaz and Karslioğlu*, 2011; *M. Schmidt et al.*, 2008; *Zeilhofer et al.*, 2009] have been used to describe the horizontal variation in space of the electron density. Although those basis functions can provide a compact representation of the ionosphere, the concept of sparsity has not been explicitly used to regularize the solution. Therefore, wavelets were used more for their ability to localize structures, leaving the sparseness that they may lead to completely unexploited.

Indeed, sparseness seems to be the key point of all the multi-resolution analysis using wavelet basis functions. The efficacy of the multi-resolution analysis with sparse regularization depends on the assumption that the horizontal variation in the ionosphere can be compactly represented with wavelets. It can be a difficult task to prove as a real global picture of the ionosphere cannot be taken, but through simulation of the process with a realistic ionospheric model, the algorithm can be demonstrated to work efficiently.

A sparse representation leads to a compact representation. Therefore, it has obvious advantages in terms of data space as few coefficients need to be stored. Instead, in the following experiments, the noise removal [*Tsaig and Donoho*, 2006] and the ability to handle better the uneven data distribution [*Michael Schmidt*, 2007] are the main aspects that will be investigated.

Sections 6.1 and 6.2 recall and comment on the main theoretical results used in the case studies of this chapter about inversion and basis functions. Section 6.3 shows the results from a simulated ionosphere. The sparse regularization is compared with a standard method (Tikhonov regularization) and the concepts of robustness, multi-resolution, and offsets retrieval are also addressed. A real case study is presented in Section 6.4. Section 6.5, instead, extends the method to High Resolution Computerized Ionospheric Tomography (HRCIT) and proposes a method based on TV in order to reduce artefacts in the reconstruction. The same section discuss also about limitation on doing HRCIT. Finally, conclusions are in Section 6.6.

The results discussed in this chapter were published in [*Panicciari et al.*, 2014a; *Panicciari et al.*, 2014b; *Panicciari et al.*, 2015].

6.1 Inversion

In this section the main results of inversion theory are recalled. For more information see Chapter 4 and Chapter 5.

In ionospheric tomography the observations for CIT are collected from ground based receivers. Those observations are in the form of sTEC and can be generally considered noise-free from the point of view of the instrument. Instead, an offset term, due to the ambiguity to

determine an absolute measure for the delay, needs to be taken into account. Given the vector $\boldsymbol{z}$ containing the collection of the whole observations within a time window, the forward problem can be defined as

$$\boldsymbol{z} = \mathbf{A}\boldsymbol{n} + \mathbf{B}\boldsymbol{b} \tag{6.1}$$

The problem of Equation (6.1) is defined on a 3-dimensional grid spacing in altitude, latitude and longitude and it is known as a forward-problem where **A** is the projection matrix that maps the electron content $\boldsymbol{n}$ into measurements $\boldsymbol{z}$, and depends on the geometry of the problem. The offset vector $\boldsymbol{b}$ is also included, together with its projection matrix **B**.

The inversion of Equation (6.1) is defined as the minimization of the following functional cost

$$f(\boldsymbol{n}) = \|\boldsymbol{z} - \mathbf{A}\boldsymbol{n}\|_{\mathbf{C}}^2 + \alpha\, \mathcal{P}(\boldsymbol{n}) \tag{6.2}$$

where $\mathcal{P}(\boldsymbol{n})$ is the regularization term, which is used to stabilize the inversion, and $\mathbf{C}$ is the matrix that is used to calibrate it, and it is defined as

$$\mathbf{C} = \mathbf{I} - \mathbf{B}\left(\mathbf{B}^{\mathbf{T}}\mathbf{B}\right)^{-\mathbf{1}}\mathbf{B}^{\mathbf{T}} \tag{6.3}$$

The parameter α is used to set the threshold between the best quality smoothed approximation and the best data-fitting. More details can be found in Chapter 5. The calibration matrix $\mathbf{C}$ is applied similarly to [*C.N. Mitchell and Spencer*, 2003; *Spencer and Mitchell*, 2011], with the difference that in this case an explicit relationship between the estimated offsets $\widehat{\boldsymbol{b}}$ and the reconstruction $\widehat{\boldsymbol{n}}$ is also provided and is given by

$$\widehat{\boldsymbol{b}} = \left(\mathbf{B}^{\mathrm{T}}\mathbf{B}\right)^{-1}\mathbf{B}^{\mathrm{T}}(\boldsymbol{z} - \mathbf{A}\widehat{\boldsymbol{n}}) \tag{6.4}$$

6.2 Basis functions in ionospheric tomography

In CIT the variation of the electron density is usually decomposed in vertical and horizontal. This is due to the different processes that describe the dynamics of the vertical and the horizontal spatial distribution of the electron density. Furthermore, the sampling of the ionosphere is mainly along the horizontal space, whilst vertically, the sampling is restricted by the satellite-receiver geometry. Therefore, two different basis function sets are used to describe a-priori information for the electron density profile and to represent the horizontal structures of the ionosphere.

The vertical profile is constrained with Empirical Orthogonal Functions (EOFs). They are obtained by applying the Singular Value Decomposition (SVD) to a series of Chapman

profiles [*Hargreaves*, 1995]. The EOFs become the set of basis functions that describe the vertical variation.

The horizontal variation of the ionosphere is described, instead, through another set of basis functions. Therefore, the electron density $\boldsymbol{n}$ is expressed as

$$\boldsymbol{n} = \mathbf{K}\boldsymbol{x} \tag{6.5}$$

where $\boldsymbol{x}$ contains the coefficients of the basis functions described through the matrix **K**. In this case the matrix **K** contains already the combination of horizontal and vertical basis functions.

Two different horizontal basis function sets were considered: spherical harmonics and wavelets. Spherical harmonics can be used to describe the periodicities of the ionosphere and are a standard choice due to the possibility to describe them over a sphere. Wavelets, instead, are used for their ability to represent localized structures at different scales and, in particular, for their property of sparse representation. Many families of wavelets exist. For the purpose of this experiment only the orthogonal wavelets were considered. Therefore, two different families were chosen, that is Daubechies 4 and discrete Meyer. Those wavelets are a good example of two wavelets with different support and sharpness/smoothness. These basis functions are defined into a grid spanning in latitude and longitude and, while spherical harmonics can be represented on a sphere, wavelets cannot and must be defined into a grid.

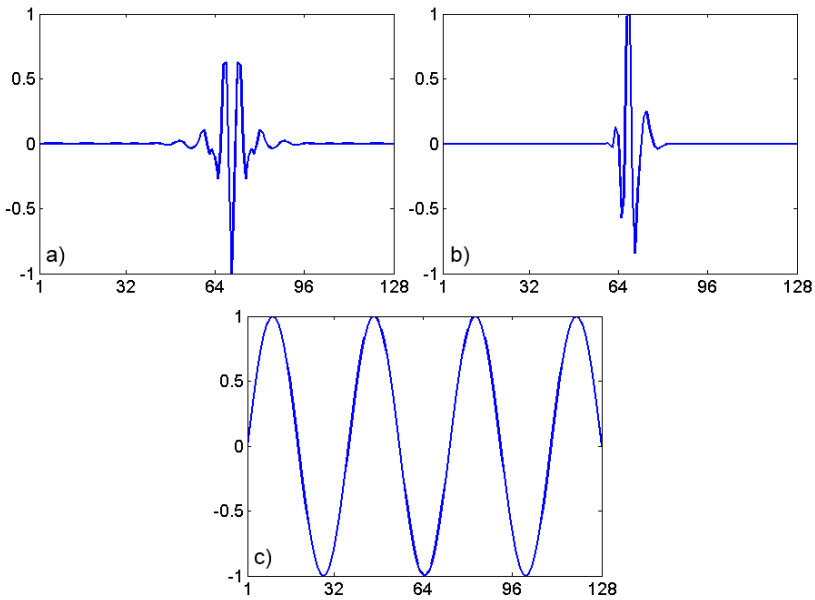

Figure 6.1. a) discretised Meyer basis function for a particular scale and translation; b) Daubechies 4 basis function with the same scale and translation as a); c) a Fourier sinusoid component of the spherical harmonic basis functions. Basis functions are shown normalized to one and are interpolated for ease of viewing.

Figure 6.1 shows a one dimensional example of the basis functions (normalized to one) that will be used in the experiment proposed in this section. Figure 6.1a and Figure 6.1b illustrate two wavelets at the same scale and position for discrete Meyer (DM) and Daubechies 4 (DB4). They have a spatial compact support that makes them particularly useful to resolve localized structures. Figure 6.1c shows a single harmonic (normalized to one) that has to be multiplied with the Lagrange polynomial (along latitude) to produce a Spherical Harmonic (SH). They have a longer spatial support and work well to describe periodicities in the ionosphere.

6.2.1 Two regularization techniques

The regularization term is the main focus in this chapter. It allows defining the uniqueness and stability of the solution.

Different regularizations exist to stabilize Equation (6.2) and make the solution unique and physically meaningful. The main goal is finding the best representation of the ionosphere that matches the observations and at the same time obviates the lack of data, which is common in CIT (e.g. in the oceans between continents). CIT is a difficult problem due to the nonlinear processes that describe the ionospheric evolution and its dependence with the Sun and geomagnetic activity. The observations used to solve the CIT problem are affected by uncertainties due to approximations in determining the sTEC (see Chapter 2). Furthermore, the observations are also affected by offsets that make them a relative measurement rather than absolute. Therefore, the regularization needs to be sufficiently strong to stabilize and make the solution unique but, at the same time, it needs to reproduce a physically meaningful ionosphere as well as a good estimate of the offsets. If all those aspects are satisfied, the regularization and its implementation can be considered suitable for the ionospheric scenario.

The minimiser of Equation (6.2) describes the best representation that can be obtained, and its properties will strongly depend on the chosen regularization term. In this chapter two different regularizations based on sparsity (or ℓ_1 regularization) and Tikhonov (or ℓ_2 regularization) will be used. They both aim to create a sufficiently detailed solution by maintaining as much information as possible from the observations. The difference lies in the information that can be extracted from the observations through basis functions and, therefore, on the efficiency on resolving different scale structures. As stated in the previous section, wavelets are good to localize structures, while spherical harmonics work well with periodicities.

The inverse problem of Equation (6.2) is solved by using two different regularizations. The ℓ_2 regularization term, referred to as the classical approach, is described as

$$\mathcal{P}_2(\boldsymbol{x}) = \|\mathbf{K}\boldsymbol{x}\|_2^2 \tag{6.6}$$

where $\mathbf{K}$ is the matrix containing the basis functions and $\boldsymbol{x}$ is the column vector of the coefficients of the basis functions.

The sparse regularization term, instead, is written as

$$\mathcal{P}_1(\boldsymbol{x}) = \|\boldsymbol{x}\|_1 \tag{6.7}$$

Another kind of regularization, namely spectral filtering, and not explicitly expressed in the equations above, comes from the selection of the number of basis functions. It is possible to select a smaller set of the whole basis function set in order to exclude the coefficients that contribute to the finer representation of the structures in the ionosphere. This has the advantage to remove the coefficients that may contribute with oscillation or high frequency artefact (especially for high resolution tomography) but also to reduce the dimensionality of the problem given by the minimization of Equation (6.2). In fact, it allows reducing the number of coefficients to estimate and, therefore, to introduce a further stabilization. This approach can work with SH, and permits to obtain a good estimation of the smooth structures of the ionosphere. The exclusion of the high frequency basis functions, in fact, reduces the finer variations that can be resolved and, therefore, reduces also the resolution of the reconstruction. With wavelets this approach is also used in order to have a comparison with SH.

6.3 A simulated case study

This section illustrates the performances of the sparse regularization applied to a simulated case study. The sparse regularization is compared with the Tikhonov regularization. Both methods rely on basis functions to describe the horizontal variation of the ionospheric electron density. In particular, wavelet basis functions are used for the sparse regularization, whilst SH are used for the Tikhonov regularization. The nature of SH allows describing better the periodicities rather than the localized structures. Instead, wavelets are basis functions that can better represent localized structures at different scales.

The grid of Figure 6.2 was selected, which spans from North America to Europe. This is a good example to show the limitation imposed, in this case by the ocean, on the density of the receivers. A grid of dimension 64x64 voxels in longitude and latitude, and 22 voxels in altitude was selected. It produces a voxel of dimension around 1x2 degrees in latitude and longitude and 50km in altitude.

6.3.1 Data

CIT relies on good data coverage, which is obtained from ground based receivers. They provide a measurement of the delay between the satellite and the receiver which is then converted in sTEC. The geometry of the observations is generally described through a geometry matrix **A**, which depends on the receiver and satellite positions.

For the experiment, the observations $\boldsymbol{z}$ were obtained from the electron density vector $\boldsymbol{n}$, which was simulated with the latest version of the International Reference Ionosphere (IRI2012) model [*Bilitza et al.*, 2014]

$$\boldsymbol{z} = \mathbf{A}\boldsymbol{n} \tag{6.8}$$

Some structures were then added in order to test the efficiency of the algorithm to resolve them. The observations were considered uncalibrated. This was obtained by adding a different constant offset to each receiver-satellite pair and the observations were collected within a time window of 8 minutes with a sample rate of 30 seconds. The representativity error was taken into account by adding a term to the observations distributed as a Gaussian noise. This term also takes into account any non-dispersive residual term (see Chapter 2). The assumption of static ionosphere was used; hence the simulated ionosphere was not changed within the 8 minutes when data were collected.

Figure 6.2 shows the Vertical Total Electron Content (VTEC) map that was used as truth while Figure 6.3 illustrates the number of rays that were used in the reconstruction (black dots are the ground stations). The number of rays is obtained by summing the intersections along the altitude within voxels of the grid. The VTEC is calculated by integrating the electron content in a certain latitude and longitude location along the altitude. The ray coverage strictly depends on the density of ground stations, data (sTEC) sampling rate and, in our case, the time window within which the reconstruction has been run. The selection of the grid is also important as a finer grid will increase the number of voxels that are not intercepted by a ray and the number of coefficients to estimate.

Some structures were located where data coverage is particularly low. In those locations the reconstruction will struggle to recover the actual value independently from the regularization that has been used. The behaviour of the algorithm in those zones will strongly depend on the regularization term.

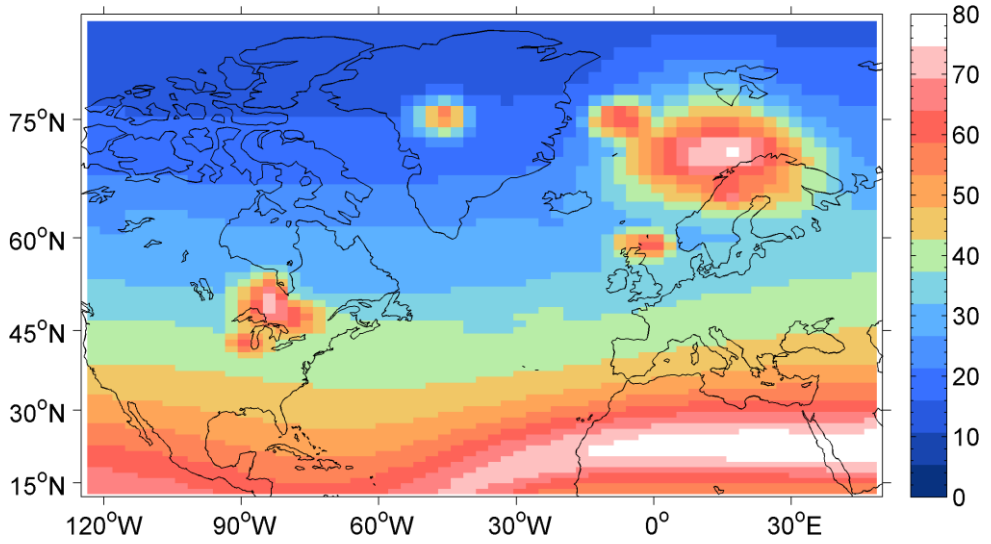

Figure 6.2. Simulated ionosphere with structures added to IRI2012. Values are in TECU (10^{16} electrons per m^{-2}).

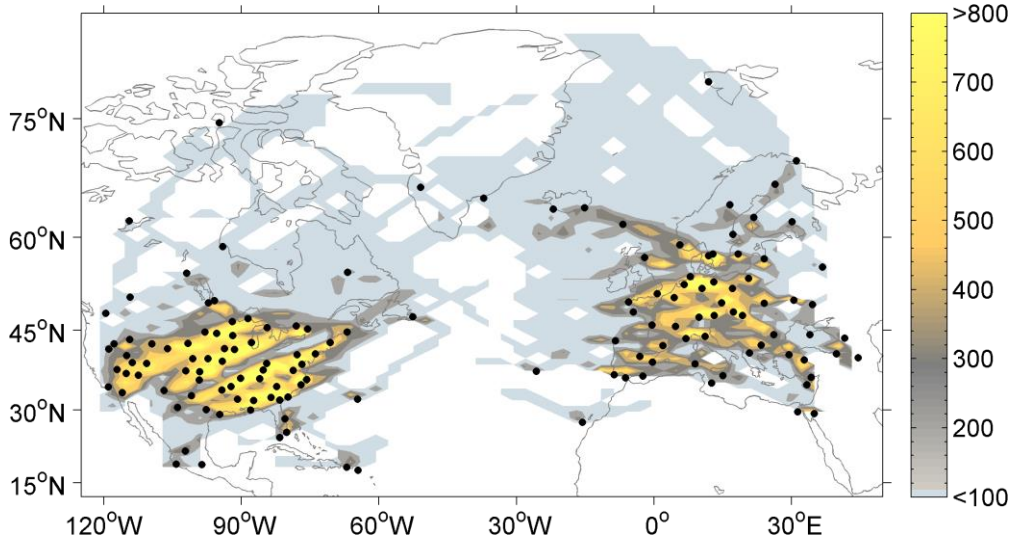

Figure 6.3. Number of rays with ground stations (black dots).

6.3.2 Results from the simulated ionosphere

The reconstructions are shown in Figure 6.4 for low resolution and Figure 6.5 for high resolution. Each figure shows the behaviour of the algorithm using different basis functions: SH (top), DM (middle) and DB4 (bottom). In order to highlight the regularization effects where only data coverage was present, a mask (left) to the reconstruction (right) was applied. In fact, each regularization technique will handle the absence of data in different ways and the mask helps to focus on other aspects more related to the ability to extract the information of the ionospheric structures from the observations where data are available.

At low resolution the reconstruction looks reasonable for both methods. The structures appear smoothed and with little detail (Figure 6.4a-c). SH seems to produce some oscillations outside the data coverage (Figure 6.4d), mainly in the Atlantic Ocean. This is due to the sinusoidal

nature of SH that makes the representation of localized structures problematic. Wavelets do not produce oscillations and the reconstruction looks reasonably smoothed for this resolution (Figure 6.4e-f), but there are some edge effects, especially for DM, between Canada and Greenland. Furthermore, DB4 unlike DM tends to fill the data gap in the Ocean (Figure 6.4f).

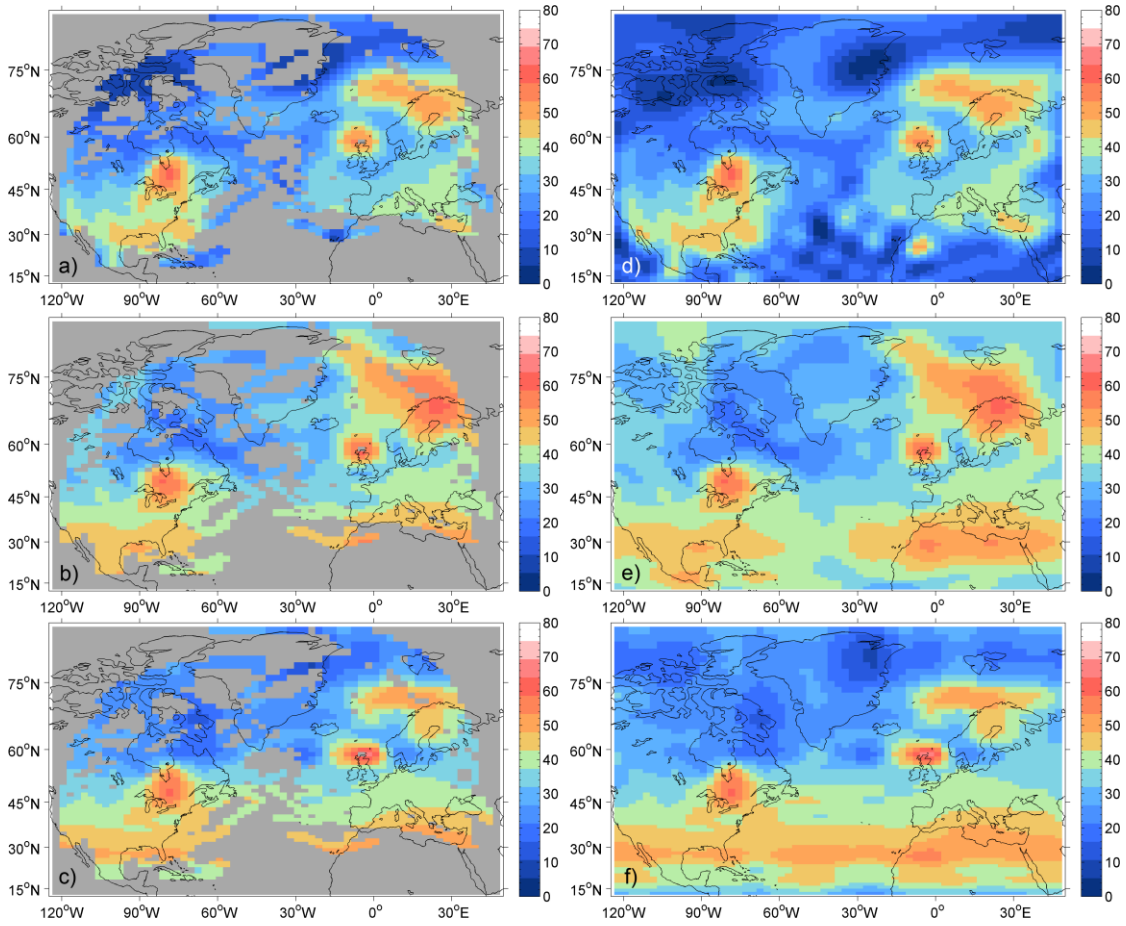

Figure 6.4. Reconstructions obtained at low resolution with masked out VTEC values where there is no ray coverage for: a) spherical harmonics; b) discrete Meyer; c) Daubechies 4; and without the mask for d) spherical harmonics; e) discrete Meyer; f) Daubechies 4. Values are in TECU (10^{16} electrons per m^{-2}).

As the resolution increases (and therefore the number of coefficients to estimate) the inversion needs in general a stronger regularization. This is shown in Figure 6.5. With SH the regularization damps many coefficients down but it seems to resolve well some of the structures (North UK and US) where good data coverage is present (Figure 6.5a). However the reconstruction presents the ring oscillation phenomenon that is an indication of the limitation of the method when a high number of basis functions are used (Figure 6.5d). The stronger regularization has reduced most of the coefficients, and the VTEC is in general underestimated. In fact, a VTEC of 40 in central Europe was expected but the reconstruction shows a VTEC less than 30. Where data are not available the regularization forces the VTEC to go rapidly toward zero. With wavelets the sparse regularization aims to minimize the number of non-zero coefficients. Therefore the smallest basis functions are contributing with the largest (smoother) ones to add detail to the reconstruction only where good data coverage

is available (this concept is regarded as multi-resolution, which will be explained later). Where data is not enough to resolve a small structure the solution will be approximated with a bigger and smoothed wavelet. By looking at the VTEC values, wavelets are recovering the value of 40 VTEC units in Europe (Figure 6.5b-c). This is mainly due to the fact that the regularization term, by exploiting the localization properties of wavelets, is adding the smallest basis only if they detect a significant enhancement over the threshold α of Equation (6.2). In general the VTEC variation is well recovered with wavelets and they seem to produce the best estimation of the ionosphere.

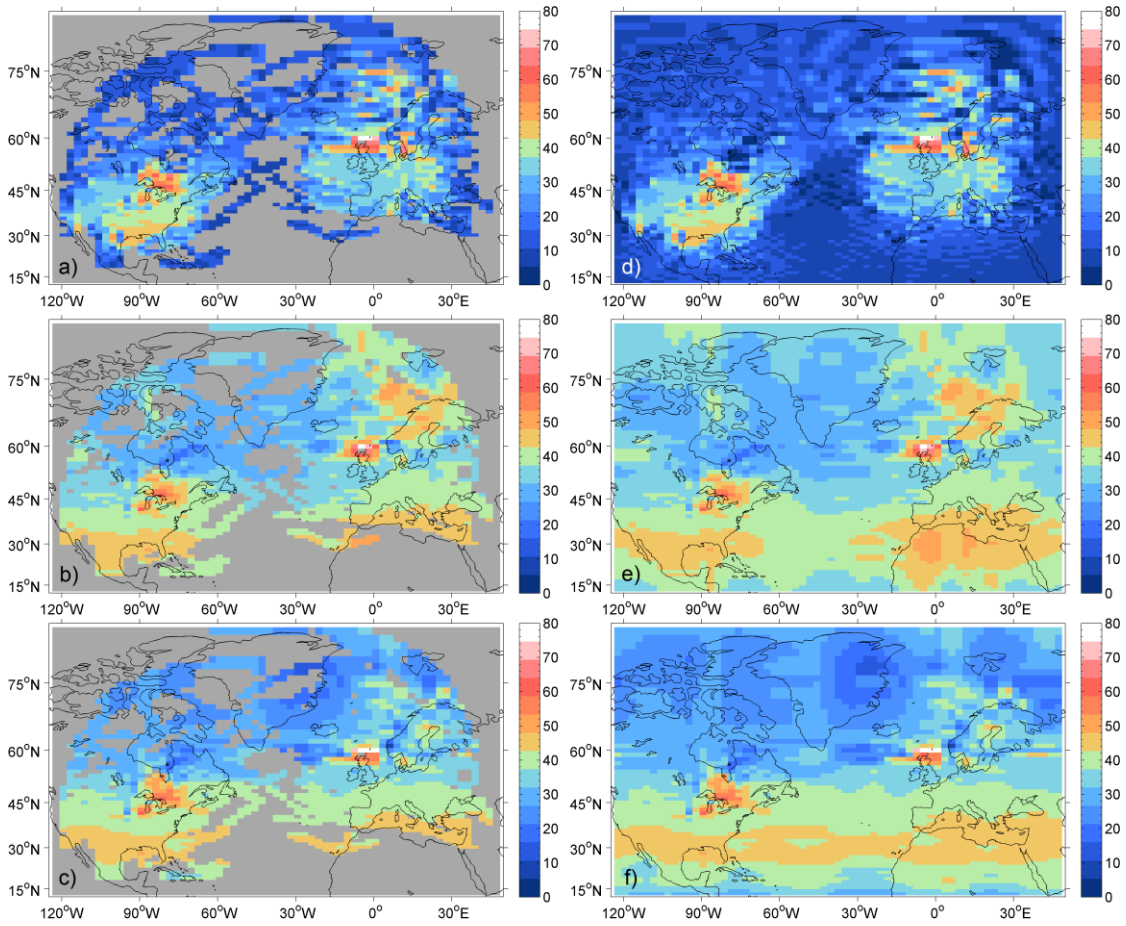

Figure 6.5. Reconstructions obtained at high resolution with masked out VTEC values where there is no ray coverage for: a) spherical harmonics; b) discrete Meyer; c) Daubechies 4; and without the mask for d) spherical harmonics; e) discrete Meyer; f) Daubechies 4. Values are in TECU (10^{16} electrons per m^{-2}).

For each reconstruction the Root Mean Square (RMS) error of the VTEC between the true and the reconstructed ionosphere was calculated. The RMS error is taking into account only the VTEC values where there is ray coverage. Values where there is no ray are, in fact, less meaningful for this statistic.

The RMS error and the number of basis functions for each reconstruction is shown in Table 6-1 for low resolution and

Table 6-2 for high resolution. The number of basis functions is shown in percentage and in absolute values within the brackets. The increasing of RMS error with resolution is caused by

the attempt of the basis functions to describe the small variations in sTEC due to non-uniform data coverage (especially in north Norway). Wavelets need less than 50% of basis functions at low resolution and even less at high resolution. This is due to the small scale wavelets that contribute only where there are enhancements with comparable scale in the ionosphere. This concept is in agreement with the asymptotical sparsity concept from [*Adcock et al.*, 2013], which states that as the resolution increases the sparsity in real signals increases as well. Furthermore, the small number of basis functions help to stabilize the inversion as only fewer coefficients have to be estimated.

Table 6-1. RMS error (values are in TECU) of the VTEC map obtained with spherical harmonics and wavelets at low resolutions. Only the VTEC coefficients where there is ray coverage were considered. The percentage of basis functions with non-zero coefficients is also shown and, within brackets, the number in absolute value.

Low Resolution					
Spherical Harmonics		Discrete Meyer		Daubechies 4	
RMS error (TECU)	Basis number	RMS error (TECU)	Basis number	RMS error (TECU)	Basis number
10.87	100% (1089)	6.66	36% (92)	7.48	41% (106)

Table 6-2. RMS error (values are in TECU) of the VTEC map obtained with spherical harmonics and wavelets at high resolutions. Only the VTEC coefficients where there is ray coverage were considered. The percentage of basis functions with non-zero coefficients is also shown and, within brackets, the number in absolute value.

High Resolution					
Spherical Harmonics		Discrete Meyer		Daubechies 4	
RMS error (TECU)	Basis number	RMS error (TECU)	Basis number	RMS error (TECU)	Basis number
19.32	100% (16641)	8.5	8% (317)	8.59	11% (467)

In this case study, Discrete Mayer basis functions seem to behave better than Daubechies 4 in terms of RMS error. Other case studies may have a different wavelet working better in terms

of reconstruction. In this case discrete Meyer will be considered as wavelet basis functions from here on.

6.3.2.1 Robustness of sparse regularization technique

Wavelets can remove better the noisy terms in the reconstruction thanks to the multi-resolution properties which allows representing the structures of the ionosphere according to their scale and position. This noisy term is given by the non-dispersive residual term described in Chapter 2 and, in particular, by the representativity error. This last source of error comes from the fact that the ionosphere is a dynamic medium, where different scale structures evolve continuously with time according to complicated physics laws in a complex environment. Therefore, a discretization of the ionosphere into a grid forces different TEC observations to represent the same value of electron density within the same voxel. Furthermore, the collection of the observations within a time window (in order to increase the data coverage) emphasizes this problem. A simple sketch of this concept is illustrated in Figure 6.6 where a voxel is crossed by three rays giving three different TEC measurements. This difference can be seen as noisy variation when a single electron density value needs to be estimated within that voxel.

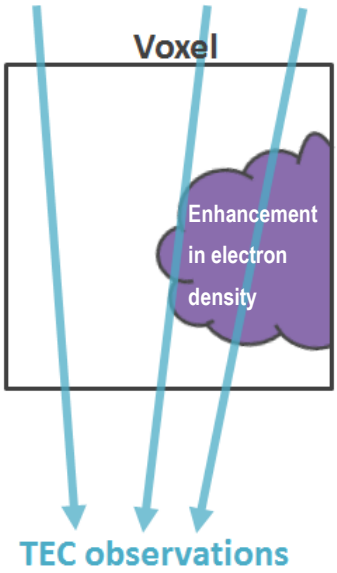

Figure 6.6. A pictorial representation of the representativity error. The sketch shows the ambiguity given by the discretization of the ionosphere over a grid. Three different rays cross the voxel and bring three different information from the ionosphere (an ionospheric structure is represented in purple). Since only one value of electron density can be estimated per voxel, the discrepancy of the measurement from those three rays is called representativity error.

As a consequence each voxel will contain an averaged value of the real electron density obtained by the observations crossing the same voxel. The ambiguity could be reduced if the resolution of the grid is increased. Unfortunately, the data coverage would put its limitation as the number of intersections of the rays with each voxel decreases as the resolution increases. This would make the problem more unstable and difficult to solve.

Figure 6.7 shows low resolution reconstructions using SH (Figure 6.7a-b) and DM (Figure 6.7c-d). TEC values are shown only where data coverage is defined. In this way, the issue of robustness is better examined rather than the issue of data coverage. It has been shown as both the methods successfully reconstructed a smooth ionosphere without the Gaussian noise term although SH (Figure 6.7a) tended to underestimate TEC value in Europe and America. On the contrary, DM (Figure 6.7c) better estimated the TEC value and produced the better reconstruction. With the Gaussian noise term the reconstruction is similar to the previous case for both SH (Figure 6.7b) and DM (Figure 6.7d) but DM produces a better image in terms of RMS error. The RMS error obtained from Figure 6.7 is shown in Table 6-3 together with the percentage of number of basis functions with non-zero coefficients.

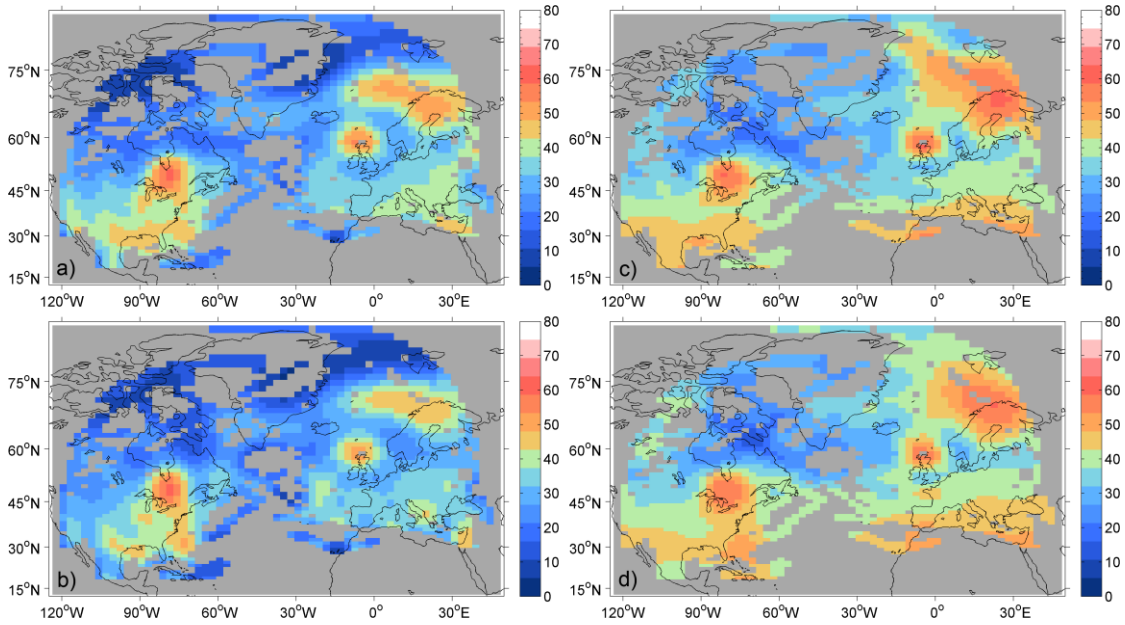

Figure 6.7. Low resolution reconstructions without Gaussian noise term for: a) spherical harmonics; c) discrete Meyer; and with Gaussian noise term for: b) spherical harmonics; d) discrete Meyer. Vertical TEC values where there is no ray coverage are masked out.

At high resolution (Figure 6.8) the number of coefficients increases and a stronger regularization is generally needed. The small structures in America are better resolved at this resolution in the case without the Gaussian noise term for both SH (Figure 6.8a) and DM (Figure 6.8c). Nevertheless, the reconstruction with SH looks noisy and strongly underestimates TEC values, mainly in Europe. DM (Figure 6.8c) resolves well the small scale structures and still produces the best reconstruction in terms of RMS error (Table 6-3). The Gaussian noise term introduces an error in the reconstructions which reduces the amount of information that can be extracted from the observations. It can be noticed as at this resolution SH (Figure 6.8b) struggles to reconstruct the image of the ionosphere and what can be seen is mainly noise. Instead, DM (Figure 6.8c) still produces a good approximation even though a small part of information is lost (see the RMS error in Table 6-3).

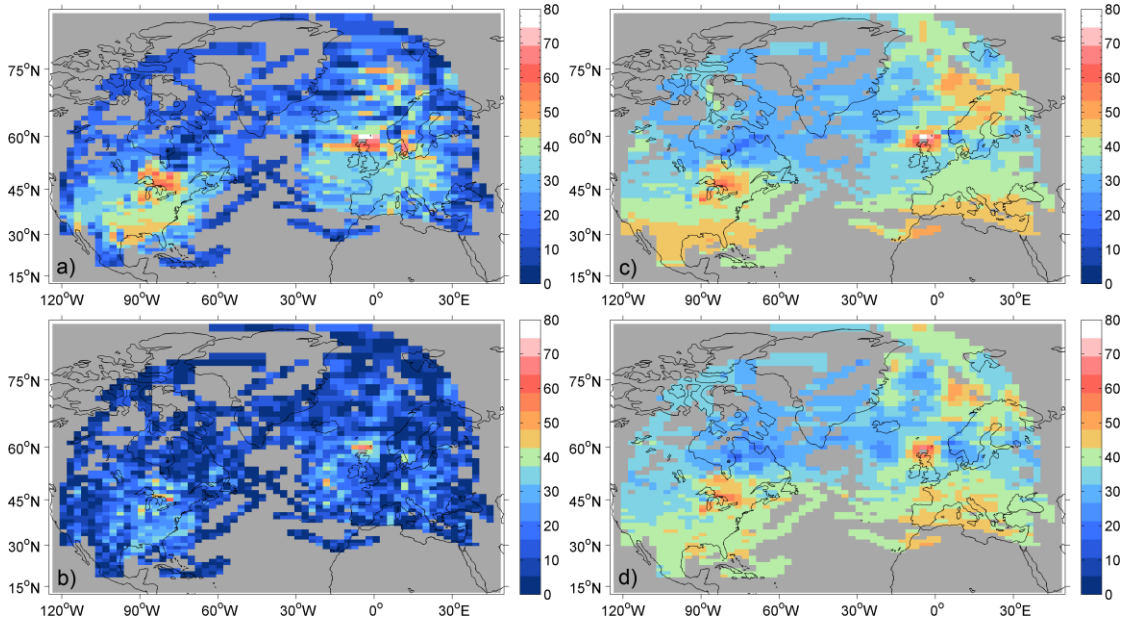

Figure 6.8. High resolution reconstructions without Gaussian noise term for: a) spherical harmonics; c) discrete Meyer; and with Gaussian noise term for: b) spherical harmonics; d) discrete Meyer. Vertical TEC values where there is no ray coverage are masked out.

In conclusion, DM reconstruction shows a better robustness to noise. This is mainly due to the sparse regularization which aims to minimize the number of nonzero coefficients. The sparse regularization, in fact, selects a subset of the most significant coefficients. Those coefficients will contain the most important part of the information (or energy) [*D. L. Donoho and Johnstone*, 1994b]. In general, it would not be possible to make the same considerations if the energy was evenly distributed among all the coefficients, like in the case of SH.

Table 6-3. RMS error of reconstructed ionosphere for spherical harmonics and discrete Meyer with and without the Gaussian noise term. The percentage of basis functions with non-zero coefficients is shown and, within brackets, the number in absolute value.

	Low Resolution				High Resolution			
	Discrete Meyer		Spherical Harmonics		Discrete Meyer		Spherical Harmonics	
Noise	RMS error	Basis number	RMS error	Basis number	RMS error	Basis number	RMS error	Basis number
No	6.6	36% (92)	10.9	100% (1089)	8.4	9% (355)	19.3	100% (16641)
Yes	7.7	30.8% (79)	12.5	100% (1089)	9.5	7.4% (304)	28.7	100% (16641)

6.3.2.2 Offsets estimation using ionospheric tomography

As described in Chapter 2, the estimation of the offsets is another important aspect for CIT. The sTEC observation is obtained from a differential phase measurement of the GPS signal delay. This measurement contains offsets that are not known a priori and that need to be estimated. Two solutions can be considered. One of them consists to estimate the offsets before the inversion. The second solution is to implement the calibration within the inversion process and let the minimization algorithm find the best approximation of the offsets that produce the best representation of the ionosphere. The method that involves the calibration within the inversion has been implemented through Equation (6.2).

For each satellite-receiver pair an offset was introduced in the sTEC measurement. The offsets were then calculated with Equation (6.4) from the reconstruction. The offsets are shown in Figure 6.9 for the low resolution reconstructions (Figure 6.4). The figure shows the scatter plot of the original offsets (x-axis) and the estimated ones (y-axis), for each receiver, obtained from the reconstructions using SH (Figure 6.9a) and DM (Figure 6.9b). The plots show that the estimated offsets of each receiver are very well recovered at low resolution.

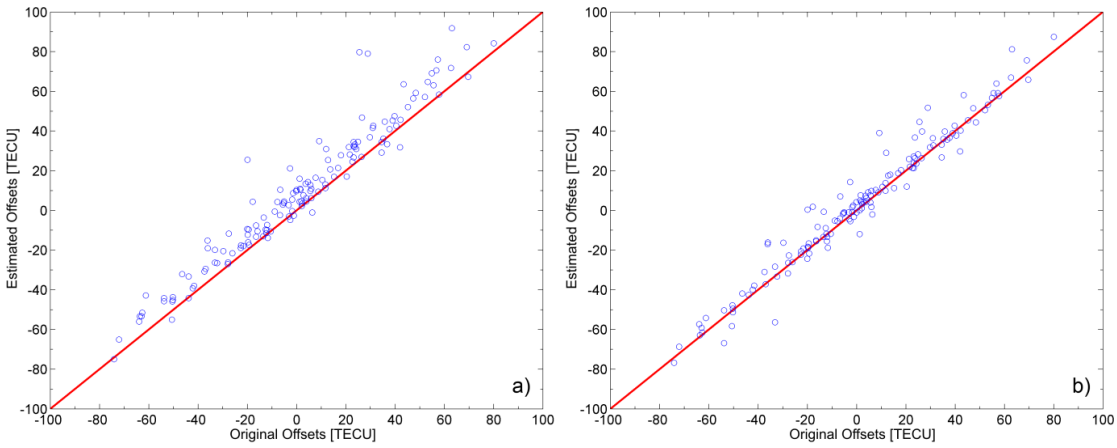

Figure 6.9. Scatter plot of the estimated offsets (y-axis) versus the true offsets (x-axis) with a) spherical harmonics and b) discrete Meyer at low resolution.

Figure 6.10 shows the scatter plot of the original offsets (x-axis) and the estimated ones (y-axis) for each receiver, obtained from the high resolution reconstruction using SH and DM basis functions. At high resolution the offsets are still well estimated from wavelets (Figure 6.10b) while they seem to be biased with SH (Figure 6.10a). There is in general an overestimation of the offsets that increases as the regularization coefficient α increases (Equation (6.2)). This is due to the fact that when α increases, the difference between the observations and the estimation in Equation (6.4) increases as well making the offsets bigger. This is probably due to the absence of regularization into the estimated offsets.

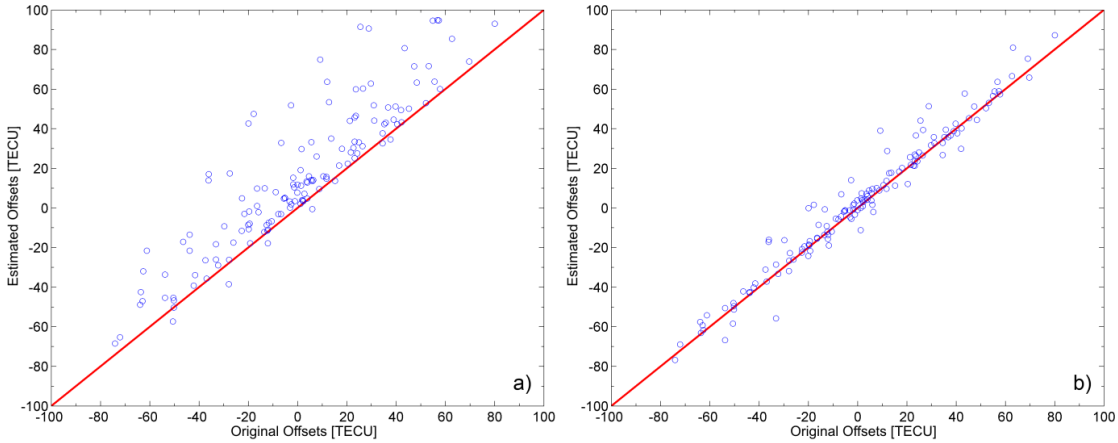

Figure 6.10. Scatter plot of the estimated offsets (y-axis) versus the true offsets (x-axis) with a) spherical harmonics and b) discrete Meyer at high resolution.

Previous works [*Chartier et al.*, 2012; *Dear and Mitchell*, 2006] implicitly used a calibration matrix similar to the one defined in Equation (6.2) to estimate the offsets from regional ionospheric reconstructions. They compared estimated offsets from tomographic reconstructions using an independent data set from the Center for Orbit Determination in Europe (CODE). Both the previous works showed a good match of the estimated offsets with CODE and did not exhibit any bias. The main differences from the work in this paper lie in the aim to reproduce high resolution maps from global ionospheric reconstructions. This involves the estimation of a higher number of coefficients and dealing with more uneven data distribution and data gaps (e.g. oceans). Consequently, a stronger regularization is needed. Therefore, the choice of the regularization becomes of primary importance if good data fitting is desired in the reconstruction with minimum artefacts.

6.3.2.3 Multi-resolution map from wavelet decomposition

As introduced earlier, another concept that can be exploited with wavelets is the multi-resolution analysis. A similar concept but with B-spline basis functions was already used in [*Michael Schmidt*, 2007]. Wavelets allow the detection of structures according to their scale and position. Small scale basis functions are therefore selected to represent small variations, otherwise only the basis functions with bigger scales are used. The ability of the algorithm to recognize small variations depends on the data availability and, therefore, the resolution (here intended as the smallest scale that can be resolved in a certain position in the map) will depend on data.

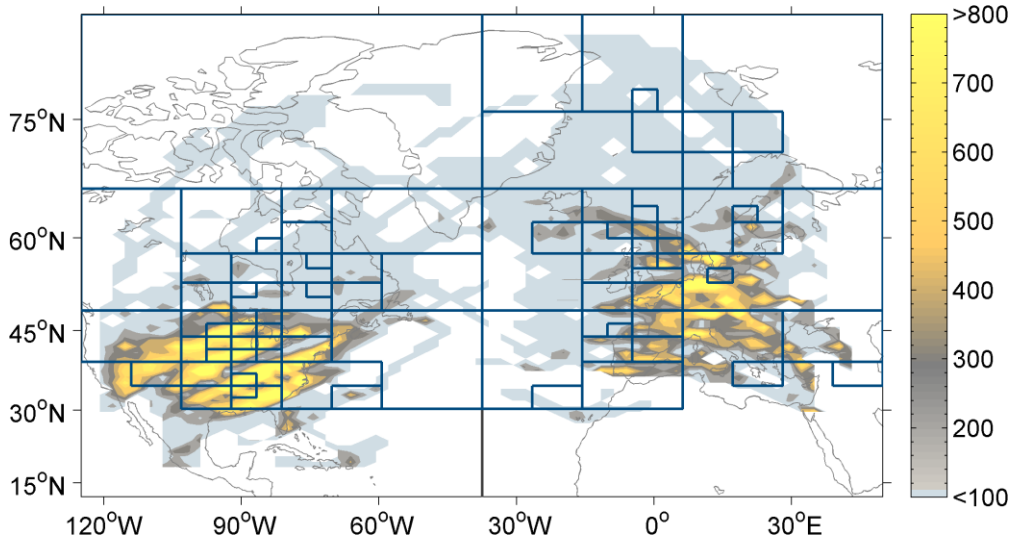

Figure 6.11. Multi-Resolution (MR) Map for the high resolution case with discrete Meyer basis functions. Each box represents the scale of the basis function and its position.

Figure 6.11 aims to explain multi-resolution with DM basis functions and gives also a visual interpretation of sparseness. Each square box indicates where the wavelet is centred in the map and the size that the wavelet is contributing with (i.e. the scale of the wavelet, which was selected of the same level for each box). This is valid only in principle as a wavelet can be defined in a longer domain than the one defined by the square. The algorithm selects smaller scale basis functions where data coverage is good, trying, as a consequence, to match better the observations. In regions where data are not available or not enough, only the biggest scale wavelets are selected and therefore the solution will look smoother. This is not possible to obtain with SH as they are longer functions and are defined over the whole globe. It is interesting to notice how small scale wavelets are not used if there is not a comparable (to the scale of the wavelet) enhancement from the data. This is the case in East and South Europe where, even if good data coverage is provided, only big scale wavelets are used.

6.3.2.4 Model-aided inversion, the best scenario

A model-aided inversion was implemented by imaging the residual after removing from the observations a background model of the ionosphere. This is called Three-Dimensional Variational (3DVar) data assimilation and assumes the knowledge of a priori information about the state of the ionosphere. This is generally obtained with an empirical model (like IRI2012) or a first principle physics model. The 3DVar was implemented with the assumption that an almost perfect knowledge of the ionosphere was provided. This was obtained by setting the background model n_0 to IRI2012 (without the added structures) and considered the residual δn

$$\boldsymbol{\delta n = n - n_0} \tag{6.9}$$

This residual is associated with a residual $\delta \boldsymbol{z}$ in the measurements $\boldsymbol{z}$ calculated as

$$\delta \boldsymbol{z} = \boldsymbol{z} - \mathbf{A}\boldsymbol{n_0} \tag{6.10}$$

Therefore, the problem in Equation (6.2) becomes

$$F(\delta \boldsymbol{x}) = \|\delta \boldsymbol{z} - \mathbf{A}\mathbf{K}\delta \boldsymbol{x}\|_{\mathbf{C}}^2 + \alpha\, \mathcal{P}(\delta \boldsymbol{x}) \tag{6.11}$$

where $\delta \boldsymbol{x} = \left(\mathbf{K}^{\mathrm{T}}\mathbf{K}\right)^{-1}\mathbf{K}^{\mathrm{T}}\delta \boldsymbol{n}$. Hence, the inverse problem is applied to Equation (6.11), which will calculate the residual information that the a-priori model could not reproduce (in this case the structures added to IRI2012). The final reconstruction is obtained by summing the estimated $\delta \widehat{\boldsymbol{n}}$ to the background model n_0. To make the problem more difficult a Gaussian noise term was added to $\boldsymbol{z}$ as in the previous section. The reconstruction (plus background model) is shown in Figure 6.12a and Figure 6.12ab, for SH and DM.

As expected both methods work well. The only remarkable difference is that SH basis functions (Figure 6.12a) are picking up some noisy coefficients which result in a noisier reconstruction than with DM (Figure 6.12b). Table 3 summarizes the RMS error obtained for these reconstructions.

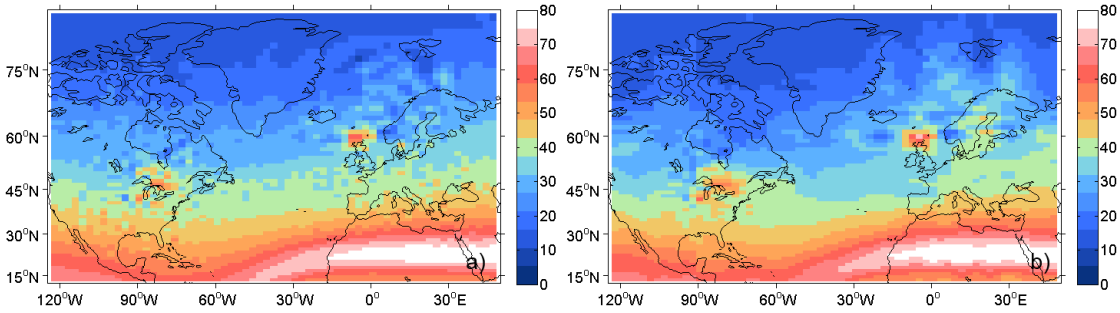

Figure 6.12. Model-aided reconstruction obtained with a) spherical harmonics and b) discrete Meyer at high resolution. A noise term (zero-mean Gaussian with 1TEC unit of standard deviation) was added to the observations.

By perfectly removing the background the algorithms need to resolve only few structures at different scale. This scenario can be considered as the best case in comparison with the worst case of the previous subsection where the model was not used. Actually, a perfect knowledge of the ionosphere cannot be ever obtained and, therefore, a background model cannot aid the reconstruction as in the above example. This mismatching with the truth means that the algorithm with an approximated background model will have performances between the worst and best case.

Table 6-4. RMS error (values are in TECU) of the VTEC map obtained with a 3DVar scheme using spherical harmonics and discrete Meyer with a noise term added to the observations. Only the VTEC coefficients where there is ray coverage were considered. The percentage of basis functions with non-zero coefficients is also shown and within brackets the number in absolute value.

Spherical Harmonics		Discrete Meyer	
RMS error (TECU)	Basis number	RMS error (TECU)	Basis number
9.26	100% (16641)	8.66	7.3% (300)

6.4 A real data case study

The aim of this section is to apply the method used in Section 6.3 to a real case scenario. A more detailed analysis will be presented in Chapter 7 with 4D reconstructions where also the time will be taken into account.

The same grid described in the simulation section was used in this real case scenario using ground based receivers. In addition to the standard GPS receiver network (from UNAVCO and IGS) Navy Navigation Satellite System (NNSS) data was also included. This data set [*Watermann et al.*, 2002] consists of three Coherent Ionospheric Doppler Receivers (CIDRs), developed at Applied Research Laboratories at the University of Texas (Austin) and capable of observing the signal from TRANSIT system[3]. The reconstruction is then validated by means of the Incoherent Scatter Radar (ISR) that was operating in Sondrestrom (Greenland) during the same day of the experiment[4].

6.4.1 Data

CIT uses sTEC measurements from ground based receivers to reconstruct the state of the ionosphere. In this experiment, data is based on a chain of TRANSIT receivers across Greenland together with GPS receivers. They recorded data during the day of the 30th September 2000, where data were collected within a time window of 9 minutes, with a sample rate of 30 seconds.

Figure 6.13 shows the receivers, TRANSIT (red) and GPS (blue), used for the reconstruction together with the ray coverage. The TRANSIT satellite (ID18362) pass (purple) is also shown. The location of the ISR is illustrated with a black circle, and the scan path is indicated with a black solid line.

[3] TRANSIT data were originally collected as part of a NSF project (ATM# 9813864).
[4] ISR observations were supported by the NSF Cooperative Agreement AGS-0836152.

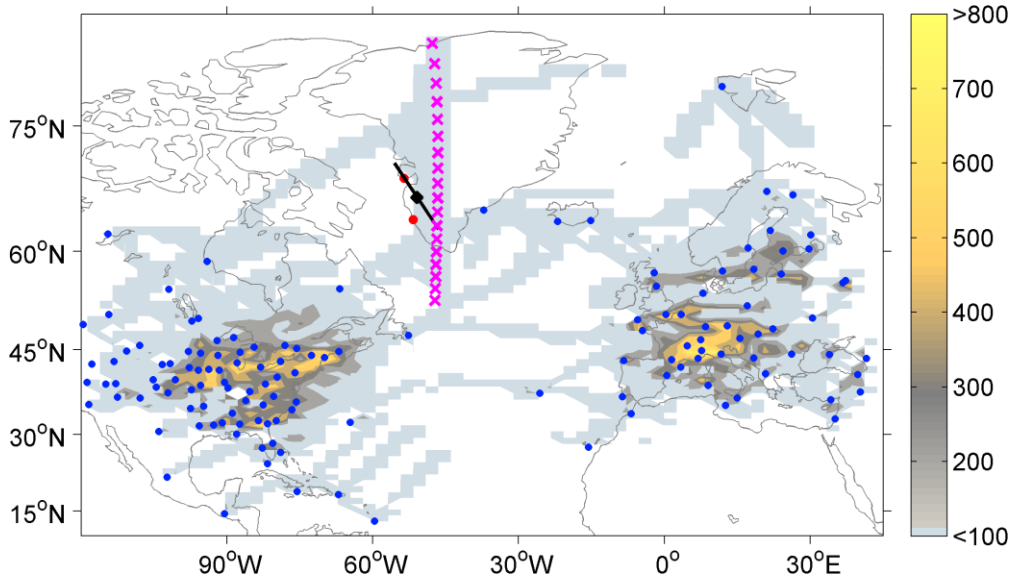

Figure 6.13. Number of rays and GPS ground stations (blue), TRANSIT ground stations (red), TRANSIT satellite pass (purple) and radar scan path (black).

6.4.2 Results from real data case study

A grid of dimension 64x64x22 voxels was selected in longitude, latitude and altitude. It corresponds to a maximum resolution of about 1x2 degrees in latitude and longitude and 50km in altitude.

EOFs from Chapman profiles [*Chapman*, 1931] were used to constrain the vertical profile to be physically meaningful. In contrast, DM wavelets and SH basis functions were used to describe the horizontal variation of ionospheric structures. The results obtained with DM were compared with SH basis functions at two different resolutions (by selecting subsets of horizontal basis functions). An ISR scan was also used as validation.

Figure 6.14 shows the reconstruction obtained with SH (top) and DM (bottom) at low resolutions. Values are in 10^{16} electrons/m^2. The resolution has been set by selecting a subset of basis functions and discarding the high frequency components. For each reconstruction a mask was also applied (Figure 6.14 a-b) in order to focus the attention only on the regions where there was data coverage.

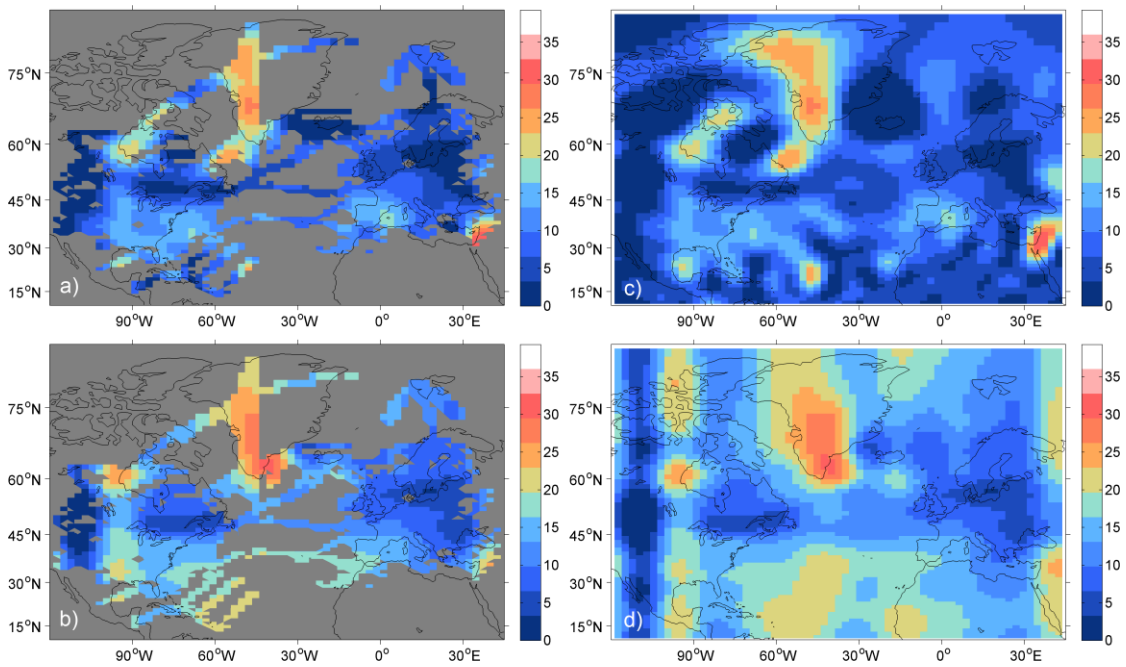

Figure 6.14. Reconstructions obtained at low resolution with masked out VTEC values where there is no ray coverage for: a) spherical harmonics; b) discrete Meyer; and without the mask for c) spherical harmonics; d) discrete Meyer. Values are in TECU (10^{16} electrons per m^{-2}).

The low resolution reconstruction is shown in Figure 6.14 a-b where there is data coverage for SH and DM, respectively. They both show a reasonable reconstruction with structures that appear smooth and with little detail. Figure 6.14c-d shows the reconstruction in the whole grid. As can be noticed SH (Figure 6.14c) produces some oscillations that generate false structures (e.g. in the ocean gap between Africa and US) where there is not data coverage. Generally, the structures tend to be smoothed and with edges that are poorly defined (as will be shown later). DM (Figure 6.14d), instead, produces some edge effects which create artefact in regions with no data coverage as with SH.

Figure 6.15 shows the reconstruction for SH and DM at higher resolution. The number of coefficients is significantly increased and SH needs a stronger regularization. The stronger regularization damps many coefficients down and the reconstruction loses its smoothness (Figure 6.15c) in comparison to the low resolution one (Figure 6.14c). A ring oscillation phenomenon is also present due to the high number of basis functions to be estimated. This does not happen for DM. There are some lengthened structures (Figure 6.14d) by using DM (between Greenland and Norway), which are mainly due to the particular data coverage (Figure 6.13). In general, better performances can be obtained with a higher number of ground stations. The correctness of the results cannot be easily verified using real data due to the limited number of instruments that can be used for validation. For this reason, the reliability of the methods described here has been tested in a controlled environment using simulated data in the previous Sections. It is also interesting to notice that DM reconstructs a structure located at south of Greenland, which is not present using SH.

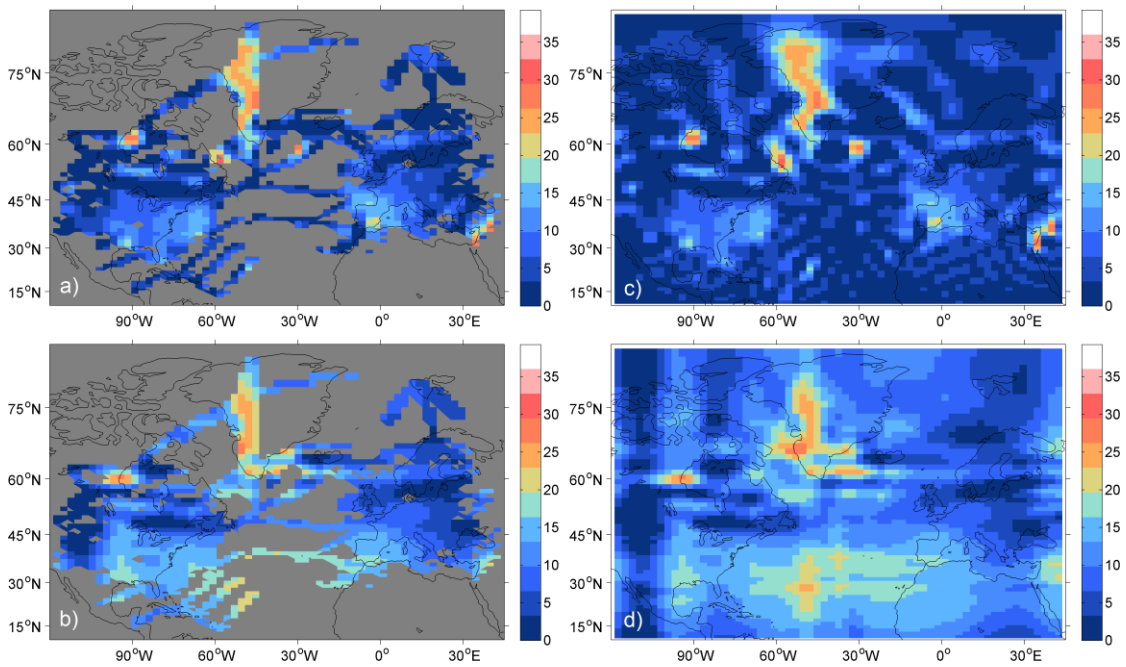

Figure 6.15. Reconstructions obtained at high resolution with masked out VTEC values where there is no ray coverage for: a) spherical harmonics; b) discrete Meyer; and without the mask for c) spherical harmonics; d) discrete Meyer. Values are in TECU (10^{16} electrons per m^{-2}).

The sparse regularization aims to reconstruct the state of the ionosphere with the minimum number of basis functions. This makes the inversion stable, maintaining most of the information that was available at low resolution but better defining the edges of the reconstruction. This can be shown by comparing the reconstruction with the ISR scan that was available during the same time interval.

Figure 6.16 shows a southward longitudinal ISR scan starting at 03:21:20UT on 30th September 2000. The scan has a duration of less than 4 minutes and values are shown in 10^{11} particles/m^3. An enhancement can be seen towards the North while a depletion is evident in the South. In particular, a trough is present at the latitude of about 64 degrees.

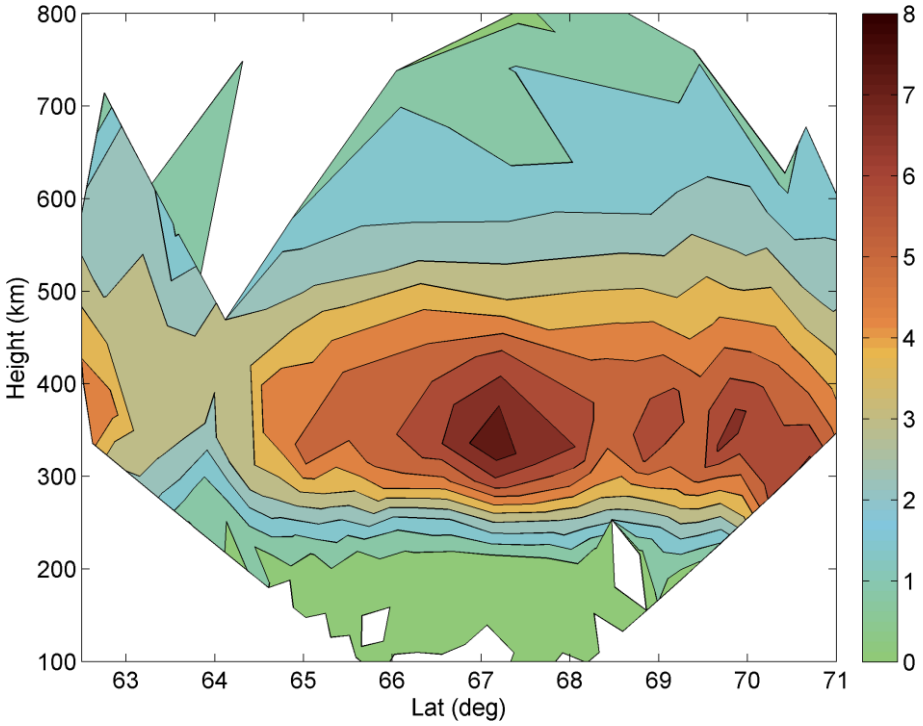

Figure 6.16. Southward longitudinal Incoherent Scatter (IS) radar scan starting at 03:21:20UT on 30th September 2000. Values of electron density are in 10^{11} particles/m^3.

Figure 6.17 a-d show the electron density (10^{11} particles/m^3) of the reconstruction along the radar scan path for SH (top) and DM (bottom), and for low (left) and high resolution (right). The radar scan plane is aligned with the geomagnetic field lines and is tilted by about 27 degrees in the anticlockwise direction with respect to the reconstruction plane. The latter reconstruction is instead aligned to the geographic coordinate system.

At low resolution SH produces a smooth profile (Figure 6.17a), as DM similarly does (Figure 6.17b). At higher resolution the depletion starts to be better defined and visible for SH (Figure 6.17c). Both SH (Figure 6.17c) and DM (Figure 6.17d) show the same trough as well as the tail of another structure (south) as indicated in the radar (Figure 6.16) although the trough edges are more well-defined for DM (Figure 6.17d). Furthermore, DM did better in resolving the structure at south of Greenland (Figure 6.15d).

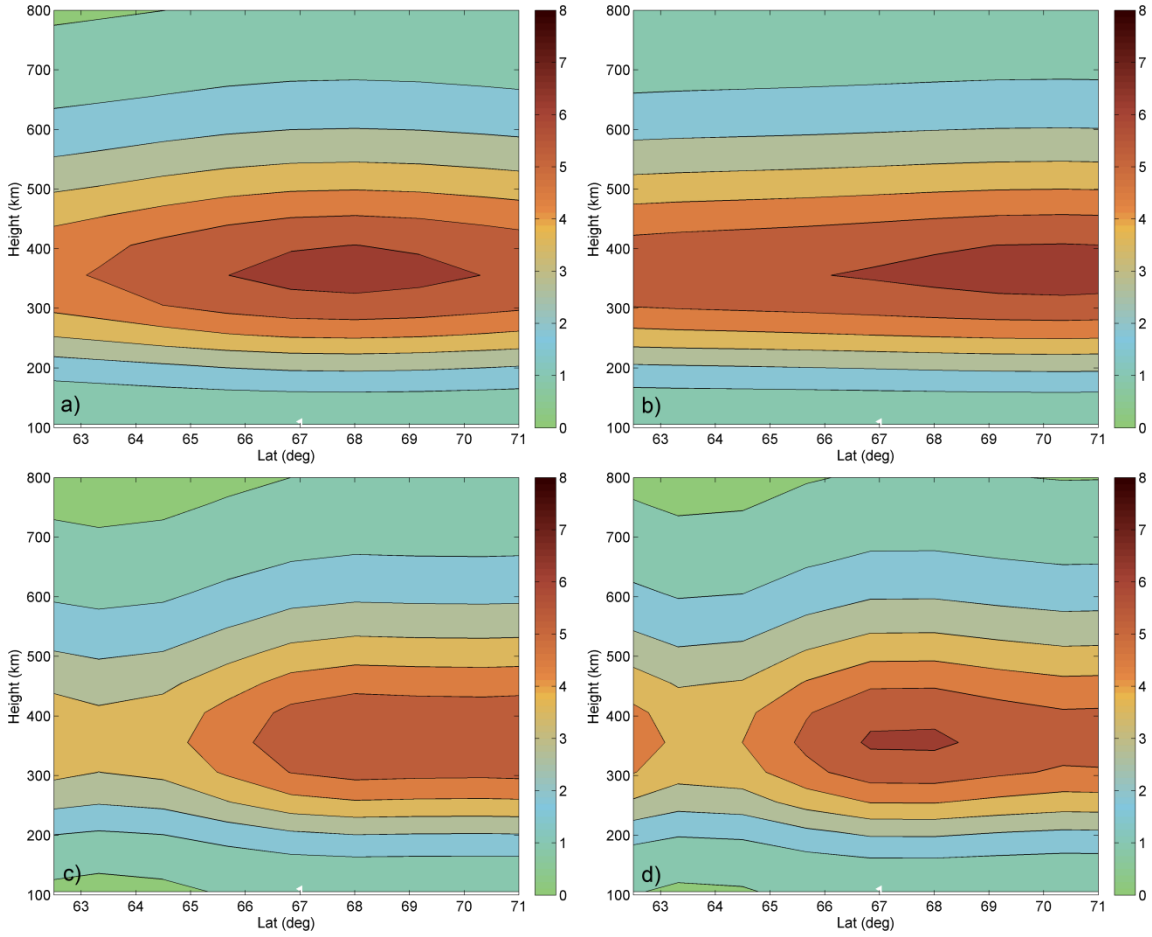

Figure 6.17. Cross sections from low resolution tomographic reconstructions for: a) spherical harmonics; b) discrete Meyer; and from high resolution tomographic reconstructions for: c) spherical harmonics; d) discrete Meyer. Values of electron density are in 10^{11} particles/m^3.

6.5 High resolution with wavelets

The maps reconstructed in the previous sections used a grid with a resolution of 1-2 degrees in latitude and longitude. This resolution can be considered as the lowest bound of a compromise (between artefacts and reliability in the reconstruction), generally used in MIDAS, of 2-4 degrees. Sparse regularization with wavelets still suffers from lack of data when the resolution increases. This is illustrated in the following experiment with real data during the day of 02 January 2010 at 00:50UT and with a resolution of 0.5 degrees in latitude and longitude. A method to reduce the appearance of artefacts is then proposed based on TV regularization (see Chapter 5).

The experiment used GPS data from ground based receivers over a grid centred in South America (Figure 6.18) and used data collected from CHAMP to validate the reconstruction. The reconstruction collected data within a time window of 8 minutes with a data rate of 30 seconds.

Figure 6.18 shows the number of observations (normalized to 1) together with the satellite CHAMP pass (purple).

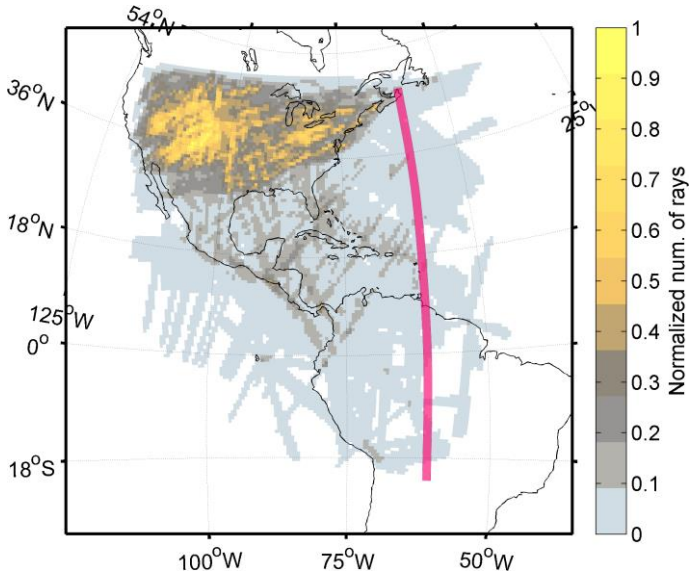

Figure 6.18. Normalized number of rays from GPS ground stations. It is also shown in purple the satellite CHAMP pass.

The reconstruction is illustrated in Figure 6.19 calculated with sparse regularization and DM wavelets. The map shows TEC values only where data was available.

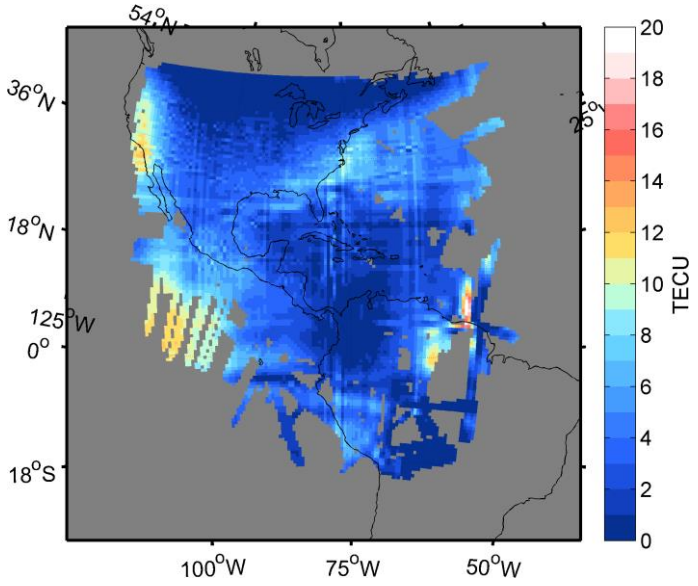

Figure 6.19. Reconstruction during the day of 02 January 2010 at 00:50UT. The map shows TEC calculated with sparse regularization and DM wavelets. TEC is shown only where there is data coverage.

The reconstruction looks very noisy. The presence of artefacts can be noticed all over the map. The reconstruction was validated by selecting the reconstructed electron density at the same altitude and position of CHAMP. The satellite CHAMP provides a localized measurement that is not easy to compare with the reconstruction. In fact, CHAMP is able to detect finer variations that would not be possible to calculate with CIT. Furthermore, the limitation given by the poor vertical resolution and the vertical resolution of 50 km may cause dissimilarities in the comparison. For those reasons, the average value was removed from CHAMP and the reconstructed profiles. This seemed to be sufficient in order to make the two

methods comparable. Figure 6.20 shows the normalized reconstructed profile (blue) calculated at the same CHAMP position (red).

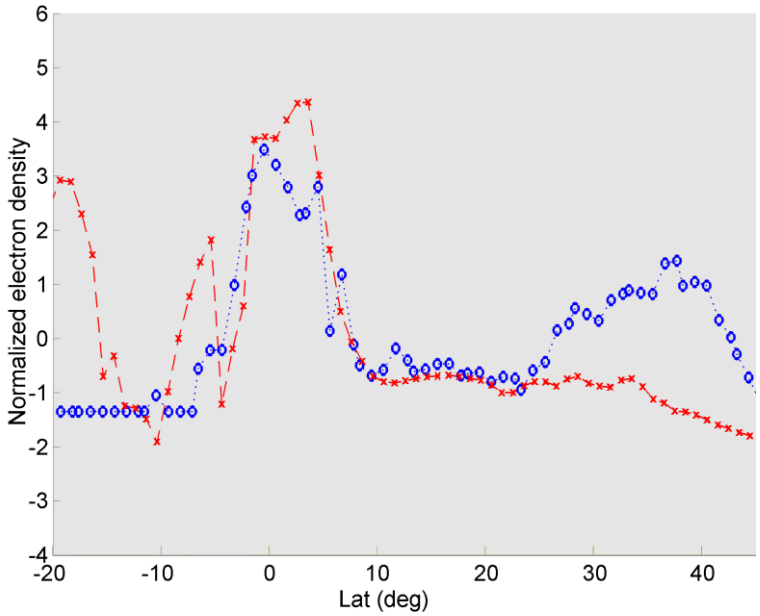

Figure 6.20. Normalized electron density obtained from the reconstructed map (blue/circle) at the same altitude and position of CHAMP (red/cross) with sparse regularization and DM wavelets.

The image shows reconstructed profile with some noisy variation in terms of normalized electron density. Not all the structures have been reconstructed. This was caused by the little data coverage, which was particularly low between -20 and -8 degrees in latitude. The reconstructed profile shows also an enhancement between 20 and 43 degrees in latitude that CHAMP did not detected. The presence of noise (or artefacts) can make the detection of sharp edges of the structures difficult. Sharp gradients, in fact, may cause disturbances in the signal propagation and their correct location is an important aspect when a high reliability in the navigation and communications system service is required.

TV techniques can help in preserving important details in the reconstruction such as edges of ionospheric structures. TV is used in combination with sparse regularization that uses wavelet decomposition to promote sparsity in the reconstruction. Sparse regularization was used, in fact, to compensate for the non-uniform data coverage and provided the input (initial solution) for the TV algorithm. This implementation is new in CIT. The reconstructed map with TV is illustrated in Figure 6.21.

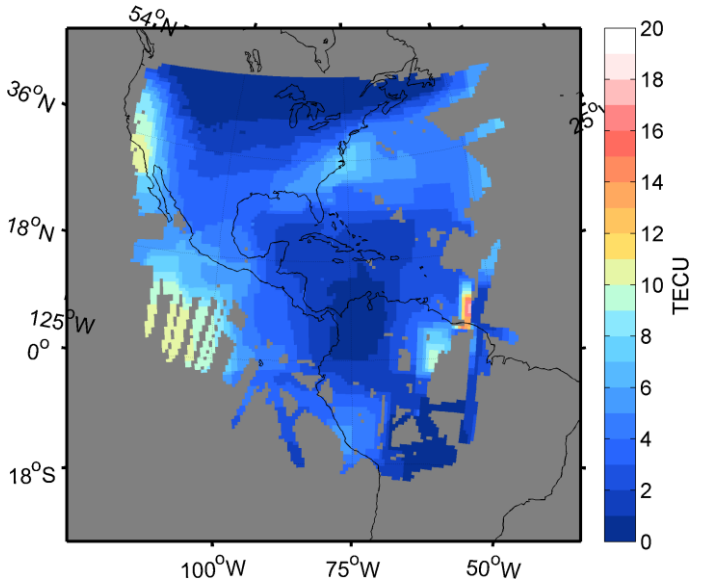

Figure 6.21. Reconstruction during the day of 02 January 2010 at 00:50UT. The map shows TEC calculated with TV regularization using the sparse regularization reconstruction as input. TEC is shown only where there is data coverage.

It can be noticed as the image looks clearer and the noisy artefacts have disappeared. However, the main information seems to be preserved. The following picture shows the validation with CHAMP.

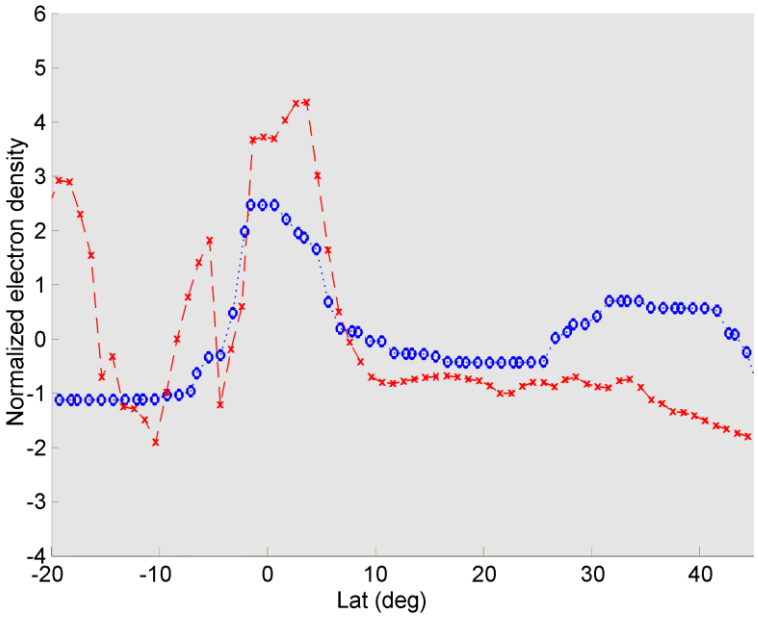

Figure 6.22. Normalized electron density obtained from the reconstructed map (blue/circle) at the same altitude and position of CHAMP (red/cross) with TV regularization.

The image shows a profile with fewer artefacts than in Figure 6.20. There is also a decrease of the peak at 0 degrees in latitude, which was caused by the TV regularization. In order to remove the artefacts, in fact, part of the information needed to be removed. However, the technique managed to efficiently separate the noise from the actual structures.

6.6 Summary

This chapter showed the first application of sparse regularization to CIT. The method was applied on both simulated and real data and compared with a standard method. Sparse regularization has been shown to be a valid alternative to the more standard method based on Tikhonov regularization and is particularly suitable with wavelets.

Sparse regularization allows better noise removal and is a more stable regularization when the number of coefficients to estimate increases considerably. Tomographic reconstructions obtained with Spherical Harmonics (SH) and two different wavelets, Daubechies 4 (DB4) and Discrete Meyer (DM) were shown in a worst and best case. The best case was obtained by selecting a background model which exactly represented the smoothed ionosphere, whilst the worst case was without any background model. In both cases wavelets were shown to produce the best reconstruction in terms of the Root Mean Square (RMS) error (where data were available) and lack of oscillations (artefacts). An important characteristic in this new approach is the ability of wavelets to handle the uneven distribution of the observations and produce a sparse solution. This was explained through the multiresolution map showing how the resolution is adapted to the data coverage and the ionospheric structures observed by the measurements.

The method has been tested to estimate the offsets of the observations and, even though it was applied to the specific case of the CIT, it can be used for general inverse problems where unknown offsets must be estimated. The method gave good performances in recovering the offsets, but a useful remark is that there is a tendency to overestimate them as the regularization parameter α increases. This effect is stronger with SH than DM and especially at high resolution. In fact, as the resolution increases the number of coefficients to estimate grows considerably. Therefore, CIT has to deal with instabilities in the reconstruction. For those reasons the choice of the regularization becomes of primary importance.

Sparse regularization aims to minimize the number of basis functions that are needed for the reconstruction. Therefore, the algorithm estimates only the coefficients of a smaller subset of the entire set of basis functions; which in an underdetermined problem like in CIT becomes of particular attractiveness. The reconstructions illustrated with real data demonstrate sparse regularization as a valid alternative to Tikhonov regularization. Furthermore, sparse regularization seemed to preserve the information when the total number of coefficients to estimate increases. A further comparison of the reconstructions with the IS radar, showed the ability of wavelets to better resolve structures in the ionosphere in comparison with the standard method. This is a good result that confirms wavelets and sparse regularization as a promising approach to detect the different scale structures of the ionosphere.

Further improvements can be achieved with different wavelet constructions. In addition, prior knowledge of the scales of structures that can be expected at different locations might help in the case of non-uniform or a small number of observations, to produce a smoother ionospheric reconstruction.

Sparse regularization was used to reconstruct high resolution ionospheric maps. The results showed the limitation of this method for HRCIT caused by the limited number of observations. It also proposed a method to overcome such limitation, which was based on TV regularization. This regularization allowed the removal of the artefacts and, at the same time, the enhancement of the structure edges. Nevertheless, some limitations of such a high resolution must be considered. Firstly, there is the possibility that time variations of sTEC in the observations could be mapped into a space variation. This can be particularly true for fast-moving ionospheric structures. On the other side, the presence of inter-crossing rays can be sufficient to reduce that risk, especially during quiet ionospheric conditions. However, the reliability of high resolution maps requires more data and a better knowledge of the time dependencies of the structures in the ionosphere.

Chapter 7
New time-dependent algorithm using sparse regularization

Introduction

The overall aim in this chapter is to demonstrate the application of different time dependent wavelet-based approaches using experimental observations. For this, a case study is chosen over the region of Europe, where there are sufficient observations to implement the technique. A major feature of the mid-to-high latitude ionosphere is the main trough; a large-scale depletion in electron density that is limited in latitudinal extent but is elongated in longitude. The trough is routinely observed from Europe in the evening and nighttime and therefore it was possible to select several suitable test cases. Independent validations of the results were possible using observations from the CHAMP satellite.

Since the 1960s, the mid-latitude trough has been analysed and the extensive observational evidence makes it a good candidate to test the sparse regularization method. The capability of the current version of MIDAS to retrieve the trough is well established. In this research here, sparse regularization is implemented in MIDAS [*C.N. Mitchell and Spencer*, 2003]. Therefore, the aim of this chapter is to establish the viability of the new method described in Chapter 6. This method relies on the wavelet decomposition and for this reason it will be referred as the wavelets method. The method described in Chapter 6 is extended over time and two different solutions are investigated. The first one uses a smooth regularization over time (but still sparse over the spatial scale), whilst the second one implements 3-Dimensional wavelets in order to fully exploit the sparsity (and MRA) also in time. This would potentially allow having a sharper representation of the TEC variation over time.

Those new enhanced methods are also compared with the previous method available at the University of Bath contained with the MIDAS software algorithms (e.g. see [*C.N. Mitchell and Spencer*, 2003; *Spencer and Mitchell*, 2011]).

In Section 7.1 a brief review of the mid-latitude trough is presented. The new reconstruction methods, based on time-sparse and time-smooth regularization, are explained in Section 7.2. There, the modification to the regularization term (in order to take into account the time), an essential element for real data, is also explained in Section 7.2.1. Due to the differences in the measurements provided by CHAMP (electron density) and CIT (TEC), a normalization formulation has been proposed in Section 7.2.2 that allows a meaningful comparison of the

results. Data and ray coverage are briefly discussed in Section 7.3. Results and analysis are in Section 7.5, whilst the conclusions are in Section 7.6.

7.1 Review of imaging the mid-latitude trough

The mid latitude trough is one of the major features of the ionosphere. It originates from the polar circulation [*Moffett and Quegan*, 1983], which causes the ionosphere to appear depleted in regions between mid-latitude and high-latitude. The mid-latitude trough, also referred to as the main trough, has been studied since the 1960's where it was known as the *Canadian border effect* since it appeared in many cases between the Canadian and USA border [*Hargreaves*, 1995]. It has an effect on the propagation of HF radio waves because of the strong gradients in the electron densities and the strong irregularities, which are often present at the edges of the trough.

The trough has been primarily observed as a night-time phenomenon and typically can develop at latitudes between 60 and 70 degrees geomagnetic. It can be identified through its depletion width (usually of some degrees) and its poleward and equatorward walls. The trough can appear in both hemispheres and in any season, even though it is observed mainly in winter [*Rodger*, 2008]. It is less common during summer where it is most likely to show around midnight [*Moffett and Quegan*, 1983].

The trough is a dynamic feature of the ionosphere that appears to move gradually towards lower latitudes [*Jones et al.*, 1997; *Moffett and Quegan*, 1983; *S. E. Pryse et al.*, 1998] throughout the evening. During an ionospheric storm, the trough is displaced strongly equatorward. It is, generally, a quite broad structure that tends to be sharper on the poleward side. The position of the trough can be roughly estimated with the following equation [*Hargreaves*, 1995].

$$\Delta = 65.2° - 2.1\mathrm{K_p} - 0.5\mathrm{t} \tag{7.1}$$

where Δ corresponds to the latitude of the minimum, $\mathrm{K_p}$ the planetary 3-hour-range index of magnetic activity, and t is the time from the local midnight (positive after and negative before) in hours.

An early attempt at measuring the trough using tomography was performed by [*S.E. Pryse et al.*, 1993] from the University of Wales. They showed 2D electron density reconstructions obtained through tomographic techniques from eight consecutive satellites passes. By using a comparison with the electron density peak, measured from the ionosonde, they demonstrated the expected behaviour of the mid-latitude trough and confirmed the potential of CIT to image large-scale ionospheric structures.

Another experiment [*C.N. Mitchell et al.*, 1995] used the EISCAT radar as independent instrument to verify the electron density estimated with 2D tomographic reconstructions. The reconstructions showed a narrow trough under very quiet geomagnetic conditions with a structure in its poleward wall. They demonstrated the accuracy of the method on retrieving the trough and the applicability of tomographic techniques to develop an ionospheric model of the main trough. Similarly, in Kersley et al. (1997) the capability of tomography for investigating and modelling the trough was shown by a series of images from tomographic reconstructions at both UK and Scandinavian latitudes.

In [*C. N. Mitchell et al.*, 1999] a large database of 2D images obtained from tomographic reconstructions showed the improvement obtained by including a seasonal term in modelling the trough position.

A more recent investigation of the trough was conducted by *Voiculescu et al.* [2006]. They produced a statistical analysis of the trough from 700 images obtained from 2D tomographic reconstructions. They found that the Interplanetary Magnetic Field (IMF) plays a role in the occurrence of the trough at different geomagnetic activities and, therefore, confirmed once again the ability of tomography on imaging the trough.

Here, the interest in the trough is to use its presence in CIT images to validate the new observation techniques. Given the new advancements introduced by the spare regularization method (Chapter 6) the aim is to test those new capabilities on resolving the trough with 4D tomographic reconstructions.

7.2 Method

Chapter 2 contains a review of the processing of GNSS data to obtain slant TEC. The main limitation is that the data are essentially relative, along a particular satellite to receiver arc. That is to say that there is very little error in the relative slant TEC observations, whereby the error is essentially due to the calibration of the offsets.

There are two approaches to the calibration. The first is to calibrate the differential code bias estimates by making use of external information. This in turn then adds an error to the TEC and could result in artificial structures being imposed on the image. An alternative approach is to calibrate within the imaging process and this necessitates the move to a 4D inversion with regularization, as is implemented in MIDAS. This is what is outlined here, with the further modification of the incorporation of wavelets in the space and time domain.

Given a dynamic linear system described through the matrix **A**, inversion methods aim to retrieve the best estimation of the states $\boldsymbol{n}$ of the system by using a set of observations $\boldsymbol{z}$. In

practice, the states are estimated by knowing only their perturbation caused by the system **A**. Generally, the observations are smaller than the states, i.e. the information is little and the system underdetermined. In CIT, the states $\boldsymbol{n}$ are the electron densities and the observations $\boldsymbol{z}$ the slant TECs. They are related through the following equation, also called the forward equation:

$$\boldsymbol{z} = \mathbf{A}\boldsymbol{n} = \mathbf{AK}\boldsymbol{x} \tag{7.2}$$

where **A** is the matrix that defines the effects of the ionosphere into the observation, and depends on the geometry of the problem defined by the satellite and receiver positions, and **K** the matrix containing the basis functions corresponding to the coefficients of the transformed states $\boldsymbol{x}$

$$\boldsymbol{n} = \mathbf{K}\boldsymbol{x} \tag{7.3}$$

As stated before, the number of observations are considerably smaller when compared to the number of the variables that need to be estimated. Therefore, the inverse form of Equation (7.2) needs to be regularized with the term $Q(\boldsymbol{x})$. The estimated solution $\hat{\boldsymbol{x}}$ (to be distinguished from the real solution $\boldsymbol{x}$) of the inverse problem can be written as the minimization of the functional cost $F(\boldsymbol{x})$

$$\hat{\boldsymbol{x}} = \min_{\boldsymbol{x}} f(\boldsymbol{x}) \tag{7.4}$$

where

$$f(\boldsymbol{x}) = \|\boldsymbol{z} - \mathbf{AK}\boldsymbol{x}\|_{\mathbf{C}}^2 + Q(\boldsymbol{x}) \tag{7.5}$$

where $Q(\boldsymbol{x})$ represents the regularization terms used to make the inversion operation more stable and $\mathbf{C}$ the calibration matrix. A standard implementation of $Q(\boldsymbol{x})$ is given by the Hessian matrix (see [*Spencer and Mitchell*, 2011]), which will encourage the solution to be smooth. A more detailed description of the inverse problem can be found in Chapters 3 and 4.

7.2.1 Sparsity meets time

As seen in Chapter 5, the concept of sparsity is exploited through a regularization factor $Q(\boldsymbol{x})$, which appears as a term in the functional $F(\boldsymbol{x})$ of Equation (7.5). This term can be described in different ways. The formulation that is of most relevance for this thesis corresponds to the ℓ_1 or sparse regularization:

$$Q(\boldsymbol{x}) = \alpha\|\boldsymbol{x}\|_1 \tag{7.6}$$

where the term α corresponds to the trade-off between a regularized dependent solution and a more data dependent solution. Equation (7.6) is used to promote sparsity in the solution, i.e. the solution $\hat{\boldsymbol{x}}$ will have many coefficients equal to zero and most of the energy will be contained in a few coefficients.

This formulation was used in Chapter 6 to analyse the performance of this new method on simulated data. In that analysis the observations were collected inside a time window of 10 minutes. Within this duration the ionosphere was assumed to be static, i.e. to not change with time. This assumption is held when simulated data are used since a perfect simulation of all the dynamics of the ionosphere cannot be expected and, moreover, there is also a better control of the environment. With real data the conditions change.

Chapter 6 demonstrated that inconsistencies between observations, within a particular time window, can be partially handled by using wavelet basis functions and the multi-resolution analysis. Whilst from the simulation this shows the stability and reliability of the method, from the physical point of view this is not realistic enough. In reality the ionosphere is not static with time – it evolves with time. Therefore, the dynamics of the real ionosphere must be described in terms of time and spatial evolution. Ignoring the time dependency of the dynamics can introduce limitations (in terms of accuracy) in the results, especially since observations need to be collected within a large time window in order to compensate for the lack of data coverage. Thus, the inversion needs to account also for the time variation. In this case, the solution $\hat{\boldsymbol{x}}$ will be referred as a 4-Dimensional vector, where the fourth dimension is the time.

In practice, the observations are collected by GNSS receivers, at typically 30 second intervals. The concept of a window of time (for the complete set of data used in an inversion) and its individual element (i.e. frame) within, is introduced here. This is shown in Figure 7.1, where a time window made up of three different frames is shown. Each frame collects some measurements (in yellow) within a defined time period. Each frame is referred to the central time t_0 for the current time and t_{-1} and t_{+1} for the previous and future time respectively. The measurements define the coverage (in each frame) over a hypothetical grid covering the continents: Europe, Africa and Asia. The single frame has usually a length of 10 or 15 minutes and the coverage that it provides is typically unevenly distributed in space. Therefore, assuming the window is made up of three frames, the total time will be 30 or 45 minutes. The advantage of this approach is that more observations can be included in the inversion, improving the overall data coverage. This is shown on the bottom picture of Figure 7.1. As a consequence, observations cannot be considered to be static anymore and a time evolution needs to be provided.

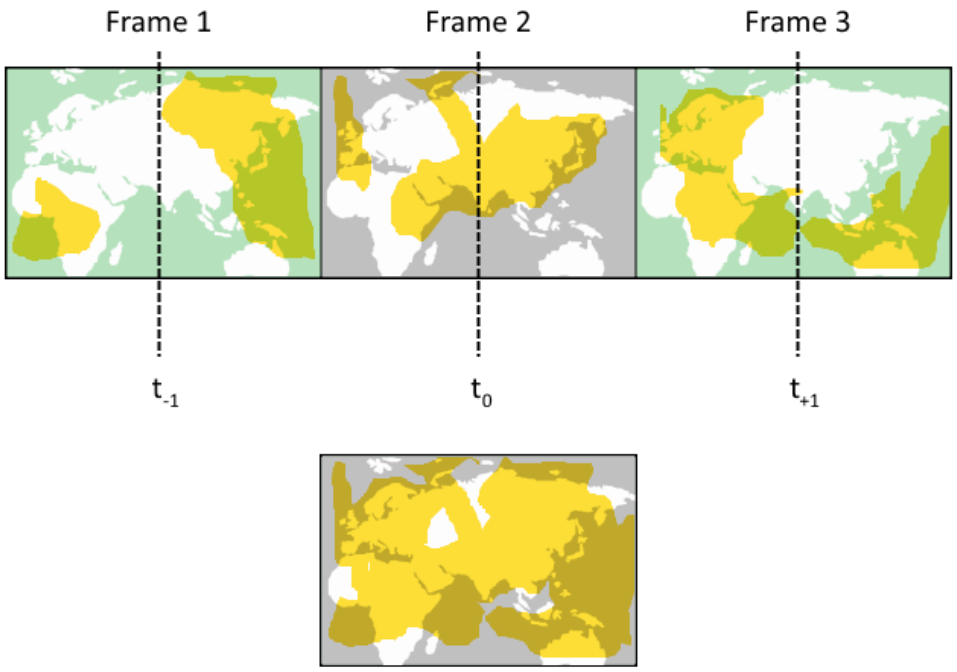

Figure 7.1: Pictorial representation of a time window consisting of three frames. Each frame contains observations (in yellow) which are distributed only partially on a hypothetical grid covering the continents Europe, Africa and Asia. The combination of the frames produces a better coverage, but the description of the time variation of the measurement between one frame to the other must be provided.

If the time evolution of the ionosphere cannot be expressed with a physical model, a sensible constraint of the time dependency of the ionospheric dynamics at each time must be provided.

Three solutions were considered based on ℓ_1 and ℓ_2 regularization:

- *Standard MIDAS (SM)*. It uses a smoothing regularization based on the ℓ_2-norm, such as $\varrho\|\mathbf{P_t}\mathbf{K}\boldsymbol{x}\|^2$, where $\mathbf{P_t}$ corresponds to the Hessian matrix (see for example. [*Spencer and Mitchell*, 2011]).
- *Time-smooth with ℓ_1 regularization in space (TS)*. It implements Equation (7.6) to define the regularization in space and a time-smoothing regularization term based on $\varrho\|\mathbf{P_t}\mathbf{K}\boldsymbol{x}\|^2$.
- *Time-sparse and spatial-sparse ℓ_1 regularization (TP)*. It is implemented by creating 3-dimensional wavelet basis functions with the sparse regularization of Equation (7.6).

The time-smooth solution was implemented with the ℓ_2 regularization for the time variation and required a few modifications in the FISTA algorithm. In fact, it was necessary to merge the two ℓ_1 and ℓ_2 regularizations in the same algorithm. In this case, the ℓ_2 regularization consists in a second regularization factor that minimizes the rate of change of gradient of the electron density at each time frame. As a consequence high and rapid changes of electron density in time (at the same position) are unlikely. Therefore, Equation (7.6) becomes:

$$Q(\boldsymbol{x}) = \alpha\|\boldsymbol{x}\|_1 + \varrho\|\mathbf{P_t}\mathbf{K}\boldsymbol{x}\|_2^2 \tag{7.7}$$

where $\mathbf{P_t}$ corresponds to the Hessian matrix and is defined only for the dimension corresponding to the time. In this equation the regularization parameters have been included in order to distinguish between α for ℓ_1 regularization and ϱ for ℓ_2 regularization. This is believed to be the first combination of (spatial) ℓ_1 regularization with (time) ℓ_2 regularization for ionospheric tomography.

This TS regularization is implemented in the FISTA algorithm. Details about the algorithm can be found in Chapter 5. Instead, the main line of the FISTA algorithm is reported below:

$$\boldsymbol{x}_n = \eta_{\alpha\tau}\left(\boldsymbol{y}_n + \tau\mathbf{K}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}(\boldsymbol{z} - \mathbf{A}\mathbf{K}\boldsymbol{y}_n)\right) = \eta_{\alpha\tau}\left(\boldsymbol{y}_n + \tau\left(\mathbf{K}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\boldsymbol{z} - \mathbf{H}\boldsymbol{y}_n\right)\right) \tag{7.8}$$

where $\mathbf{H} = \mathbf{K}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{A}\mathbf{K}$. At the n-th iteration, the solution $\boldsymbol{x}_n$ is obtained by applying the soft-threshold operator $\eta_{\alpha\tau}$ to the estimation $\boldsymbol{y}_n$ plus the residual not yet exploited in the solution. The soft-threshold operator can be written as:

$$\eta_\gamma(\hat{\boldsymbol{x}}_i) = \begin{cases} 0 & if\ |\hat{\boldsymbol{x}}_i| \le \gamma \\ sgn(\hat{\boldsymbol{x}}_i)(|\hat{\boldsymbol{x}}_i| - \gamma) & if\ |\hat{\boldsymbol{x}}_i| > \gamma \end{cases} \tag{7.9}$$

where $\gamma = \alpha\tau$ and τ is equal to $1/\max\left(\mathrm{eig}(\mathbf{H})\right)$.

In order to include the time regularization, the matrix $\mathbf{H}$ was substituted with the matrix $\mathbf{H_t}$ (where the subscript t indicates the time-dependency of the matrix):

$$\mathbf{H_t} = \mathbf{H} + \varrho\mathbf{P_t^T}\mathbf{P}_t \tag{7.10}$$

In this way, the projection of the solution into the observation domain will be constrained by the smoothing operator $\mathbf{P_t}$ (supposing it corresponds to the Hessian matrix), which operates in the time domain.

In order to minimize the border effects due to the wavelets, in particular on the edges of the grid where there is little coverage, a further regularization was introduced to TS (included in the same matrix $\mathbf{P_t}$). It forces the solution to be smooth in regions where the data coverage has a total number of rays (summed along the vertical) equal to or less than 50. Generally, this regularization has a minimum influence on the solution and helps to reduce oscillations in particular at the border of the grid.

7.2.2 Comparing TEC maps with CHAMP electron density profiles

CIT provides a 4D reconstruction corresponding to the best estimation of the ionosphere at a certain time. This estimation depends strongly on the assumptions that are introduced to make the problem stable, and the algorithm efficient. One of the tools used to reduce the dimension of the problem and resolve the difficulties, due to angle-limited geometry of the rays, are the Empirical Orthonormal Functions (EOFs). They can provide a certain degree of freedom on

defining the vertical profile in the reconstruction. Nevertheless, their contribution will strongly depend on the set of vertical profiles used to create the EOFs through SVD decomposition (more information can be found in Chapter 4). Therefore, vertical profiles in the reconstruction cannot be considered completely reliable.

Observations from the CHAMP satellite provide a local measurement of electron density at an altitude close to the F2 peak. A direct comparison of CHAMP with CIT electron density maps cannot be used directly for the reasons previously explained in Chapter 6. However, the solution adopted for this analysis is different from the one in the previous chapter. The solution consists in using electron density values (from CHAMP) and TEC values (from CIT reconstructions) normalized with their standard deviation. In addition, the average value is also removed in order to take any bias out of the analysis. If v refers to either electron density or TEC, the normalization is obtained as:

$$v_n = \frac{v - \bar{v}}{std(v)} \tag{7.11}$$

where $\bar{v}$ represents the average value of v and $std(\cdot)$ the standard deviation calculated from the profile at the same spatial and time sampling of CHAMP..

Normalized TEC and electron density have, therefore, a unitary standard deviation. As a consequence the comparison will describe better the presence of a structure/gradient rather than its enhancement in terms of density.

7.3 Data

Measurements were collected in a grid over Europe from different repositories, such as UNAVCO, IGS, and CDDIS. The observations are initially analysed using the MIDAS pre-processing tool, as detailed in Chapter 2. One of the main purposes of this tool is to convert the pseudoranges (provided by RINEX files) into slant TEC observations.

The grid for CIT reconstructions is shown in Figure 7.2 together with the normalized number of available rays for two different times. The number of rays shown in the figure is obtained by summing along the vertical the number of rays that intercept the voxels of the grid, which are then divided by the maximum value in the grid.

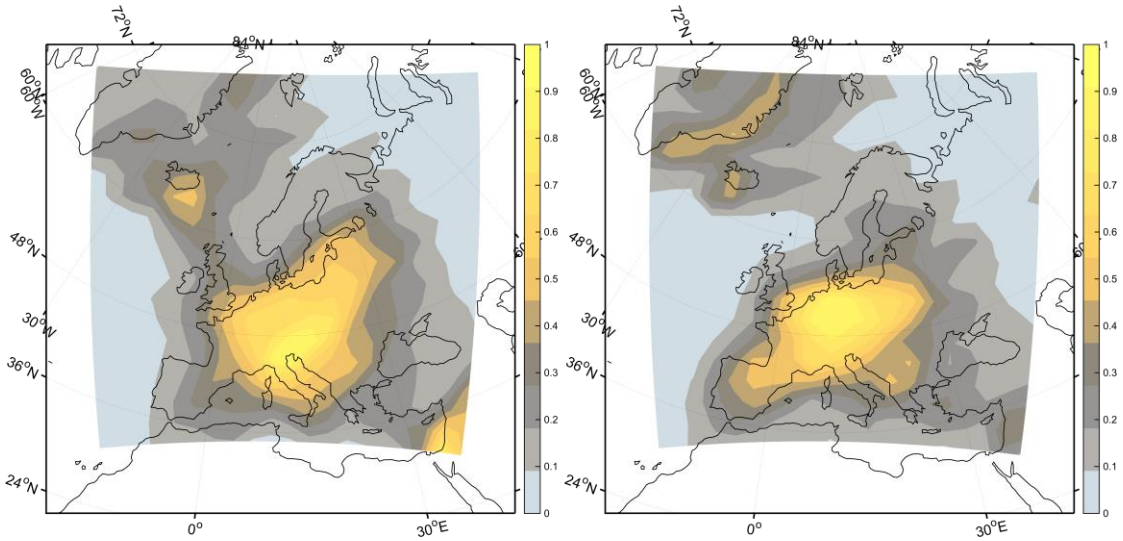

Figure 7.2. Normalized number of available rays for two different times. The coverage corresponds to a grid over Europe with 2.5 degree resolution in latitude and longitude.

Figure 7.2 is used to give an indication of the coverage over a grid with a resolution equal to 2.5 degrees in latitude and longitude. As expected there is a good coverage in Europe, Greenland and Iceland. However, little coverage is present on the edges of the grid and in the ocean..

TEC maps from CIT reconstructions are compared with the electron density measured from CHAMP data that were collected with a rate of 5 seconds and are provided by AF Research Lab (US).

7.4 Sensitivity to the regularization parameter

In this section, the effect of the regularization parameters α and ϱ on the ionospheric reconstructions is discussed. The parameters α and ϱ are chosen in order to have the best compromise between the most reliable reconstruction and the best match with the observations. A small value could produce the best match between data and reconstruction (Equation (7.5)) but with high probability of producing artefacts. On the other side, a high value reduces the probability of having artefacts in the reconstruction in favour of a more reliable reconstruction. The effects of this parameter strongly depend on the choice of the regularization term $Q(\boldsymbol{x})$ (Equation (7.5)). The following picture shows three different reconstructions for SM (Figure 7.3a-c) and wavelets with TS (Figure 7.3d-i) and TP (Figure 7.3j-o). The reconstruction was calculated for the day of 23 July 2009 at 23:40 UT with three different values of α and ϱ, each ten times bigger.

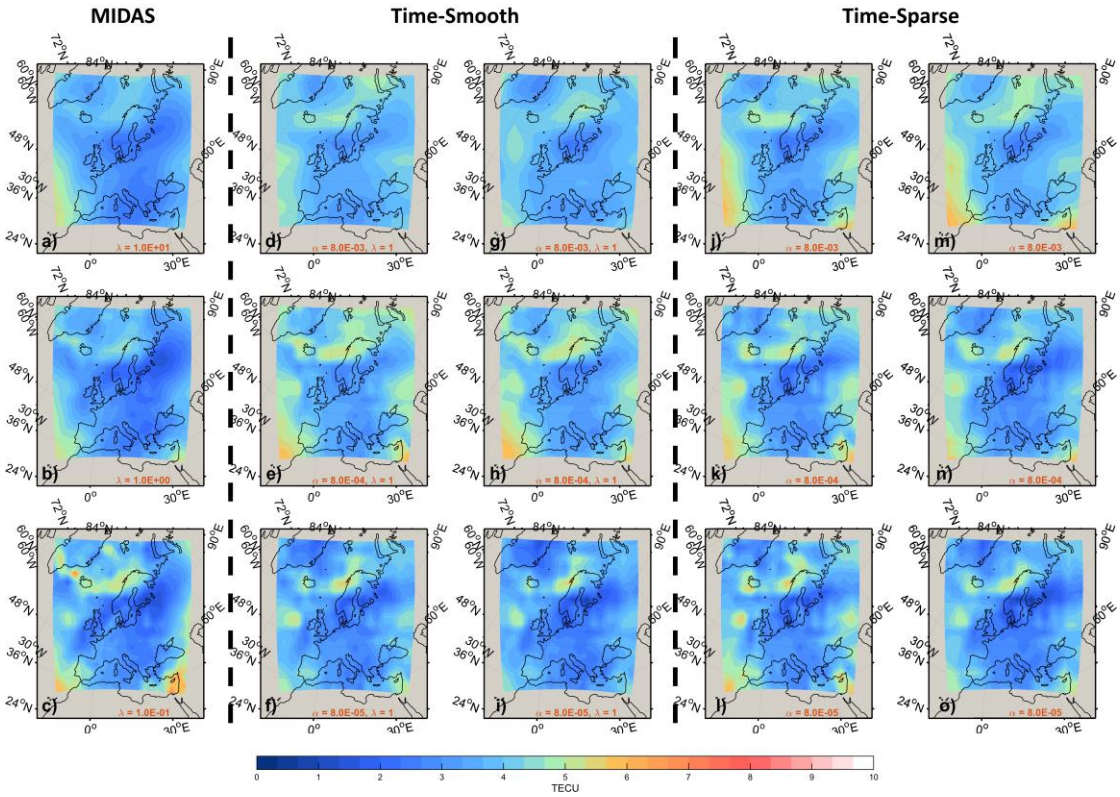

Figure 7.3. Reconstructions obtained with (a-c) standard MIDAS (left), time-smooth regularization with (d-f) Daubechies 4 and (g-i) discrete Meyer, and time-sparse regularization with (j-l) Daubechies 4 and m-o) discrete Meyer. The images show TEC maps of 23 July 2009 at 23:40 UT for three different values of α and ϱ.

It is shown that different results are produced in the reconstructions by varying the ϱ parameter in SM. A small value produces more variation (and potentially more artefacts) in the reconstructed ionosphere (Figure 7.3a) in comparison to a high value (Figure 7.3c). In general an optimal compromise was empirically found when ϱ was equal to one. It can be also noted that for ϱ equals to 1E-1, structures appear at 0 degrees longitude and 82 degrees latitude. This structure could be associated with an artefact since it is not present in the other reconstructions, even with a smaller regularization parameter. The situation is different for wavelets. An α equals to 8E-4 produces almost the same structures that were obtained with a higher value (α equals to 8E-3). This is due to the fact that wavelets are representing the ionosphere in a simplified form. This means that the information is contained within just a few basis functions. The coefficients associated with those basis functions will have high values in correspondence with the number of structures. Therefore, increasing the regularization parameter α will lessen the number of small ionospheric oscillations, due to artefacts or noise, from the reconstruction, but will still keep the main information extracted through the basis function. However, the reconstructions with TP seem to be less sensible due to the decreasing of the regularization parameter; in particular for α equals and less than to 8E-4. However, for higher values, the regularization is too strong and the wavelet reconstructions appear smooth.

7.5 Results and analysis

A set of five case studies were selected by using a grid over Europe during moderate geomagnetic activity (Kp index 3) during the second half of 2009. Moderate geomagnetic activity was selected to ensure a good chance of seeing a well defined trough over Europe. For each case study, TEC maps are shown using reconstructions from using three techniques; time-sparse regularization, time-smooth regularization and the standard version of MIDAS. For each case study a total of seven reconstructions, 10 minutes apart, were calculated. The reconstructed TEC maps are then normalized and compared with the normalized electron density as measured from CHAMP, which is used as reference. The normalization, explained in Section 7.2, is a necessary operation in order to make two comparable measurements, which are otherwise different. However, it must be remembered that the CHAMP satellite orbits at an altitude of about 400 km and provide a local measure of the electron density, whilst the reconstructions show vertical TEC integrated in the range 70-1000 km. Therefore, the operation of integration will obviously average local variations in the profile, which are in any case not reproducible because of the use of vertical basis functions (EOFs, see Chapter 4). This is not problematic because the comparison aims to test the reconstruction of horizontal variations of the ionosphere. Nevertheless, CHAMP remains a good and reliable measure that allows, in this chapter, validating of the reconstructed TEC maps over the full range of latitudes of the grid.

Two different wavelet families were used for the sparse regularization, Discrete Meyer (DM) and Daubechies 4 (DB4). The regularization parameter ϱ was chosen equal to 1 for SM and TS. The regularization parameter α was set to 0.008 for TS and TP. The maps are reconstructed over a grid that is then cropped in latitude and longitude in order to reduce the border effects.

7.5.1 Case Study 1, 22 October 2009

This first case study shows the evolution of the trough located at around 65 degrees north of the UK. The reconstructions are shown in Figure 7.4 for SM (a-c), DB4 (d-f) and DM (g-i) by using the time-smooth regularization. The time increments between each reconstruction correspond to 30 minutes starting from 01:10 UT of 22 October 2009. The time 01:40 UT was chosen at the closest proximity to the mid-latitude trough as seen from CHAMP.

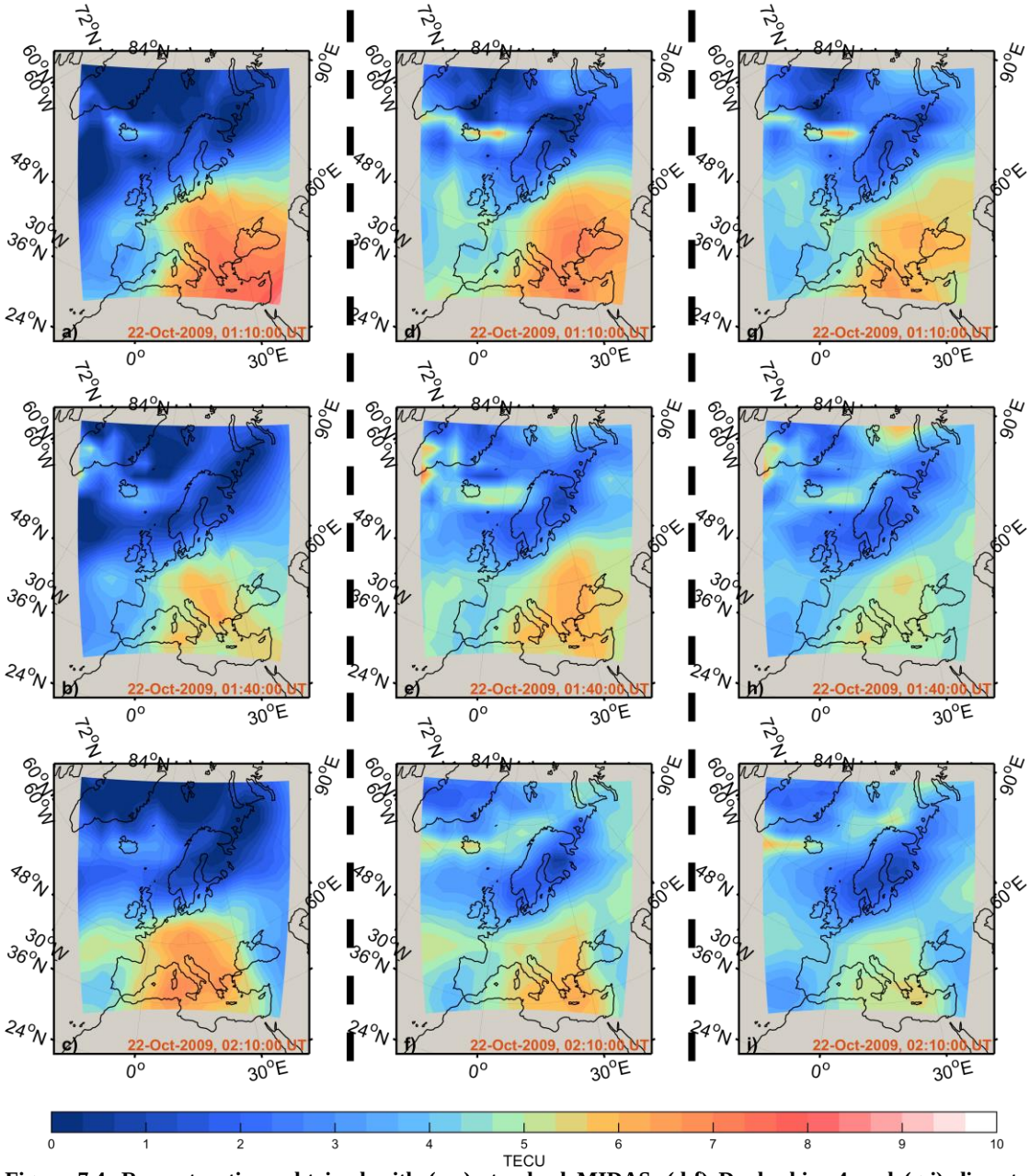

Figure 7.4. Reconstructions obtained with (a-c) standard MIDAS, (d-f) Daubechies 4 and (g-i) discrete Meyer with time-smooth regularization. The images show TEC maps every 30 minutes starting from 01:10 UT of 20 October 2009.

With reference to the reconstructions, the trough starts to evolve from the West-side of the Grid and reaches the maximum extension at 01:40 UT. Both MIDAS and wavelets reconstruct similar information even if the trough looks shallower with wavelets. There are substantial differences between the the three methods over North Europe, in particular at 02:10 UT where wavelets still reconstruct some structures. In fact, SM produces the smoothed image followed by DM and DB4. In conclusion, wavelets seem to reconstruct more structures than SM, which is as expected. Those differences are associated to the different spatial regularization because the same regularization in time was used. The bulk of TEC, however, is quite similar between SM and DB4, whilst DM produced a lower value in TEC.

The TEC maps for the time-sparse algorithm are shown in Figure 7.5.

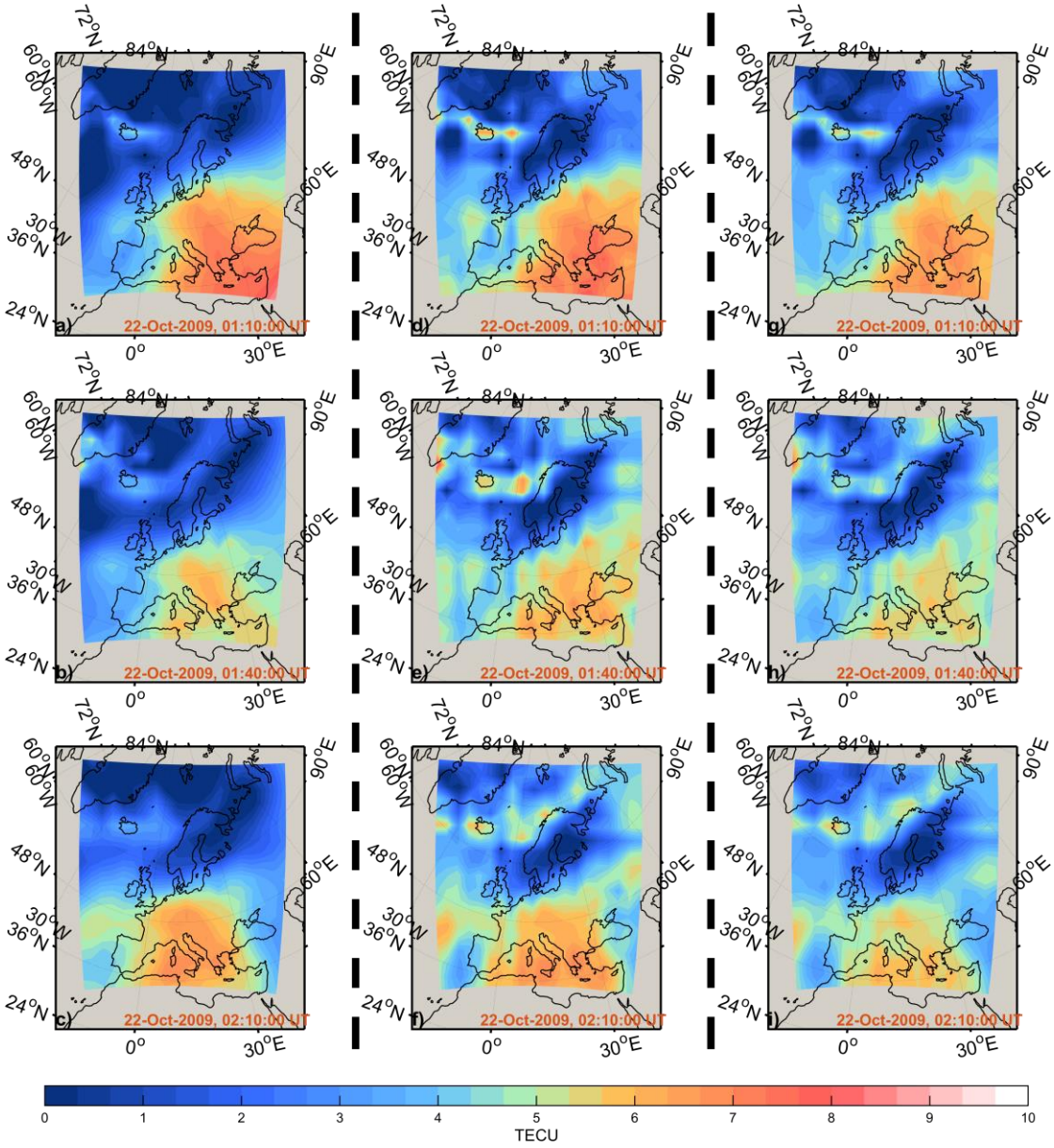

Figure 7.5. Reconstructions obtained with (a-c) standard MIDAS, (d-f) Daubechies 4 and (g-i) discrete Meyer with time-sparse regularization. The images show TEC maps every 30 minutes starting from 01:10 UT of 20 October 2009.

In this case the 3D wavelets produce a more structured ionosphere for both DM and DB4. This shows that the time-smooth regularization is a strong assumption that consistently smooths the TEC in the wavelet-based method and, therefore, SM. This clearly shows that the time regularization also has an effect over the spatial scales of the resolved structures. In this case the trough looks shallower for SM than for wavelets. This, probably, is caused by the time-sparse algorithm that could exploit the compact representation of wavelets more efficiently.

The following figures show the comparison between CHAMP electron density and the reconstructed TEC maps. TEC and electron density are normalized as explained in Section 7.2. The normalized CHAMP electron density is shown as coloured dots, whose colour is proportional to the normalized electron density. The plots on Figure 7.6 show the reconstructions at 01:40 UT, which corresponds to the closest time where CHAMP sees the trough. The reconstructions were done with DB4 (Figure 7.6a) and DM (Figure 7.6b) by using the time-smooth regularization and SM (Figure 7.6c). Figure 7.6d shows, instead, the normalized TEC obtained with DB4 (red), DM (orange) and SM (green) compared with CHAMP normalized electron density (black).

The bulk of electron density at low latitudes is well recovered, but the enhancement at 70 degrees in latitude and the trough at 64 degrees in latitude, have not been reconstructed well by SM (Figure 7.6c). Instead, both wavelets (Figure 7.6a-b) have recovered a structure nearby to where CHAMP measured an enhancement in electron density, even if DM seems to have the best reconstruction (Figure 7.6b).

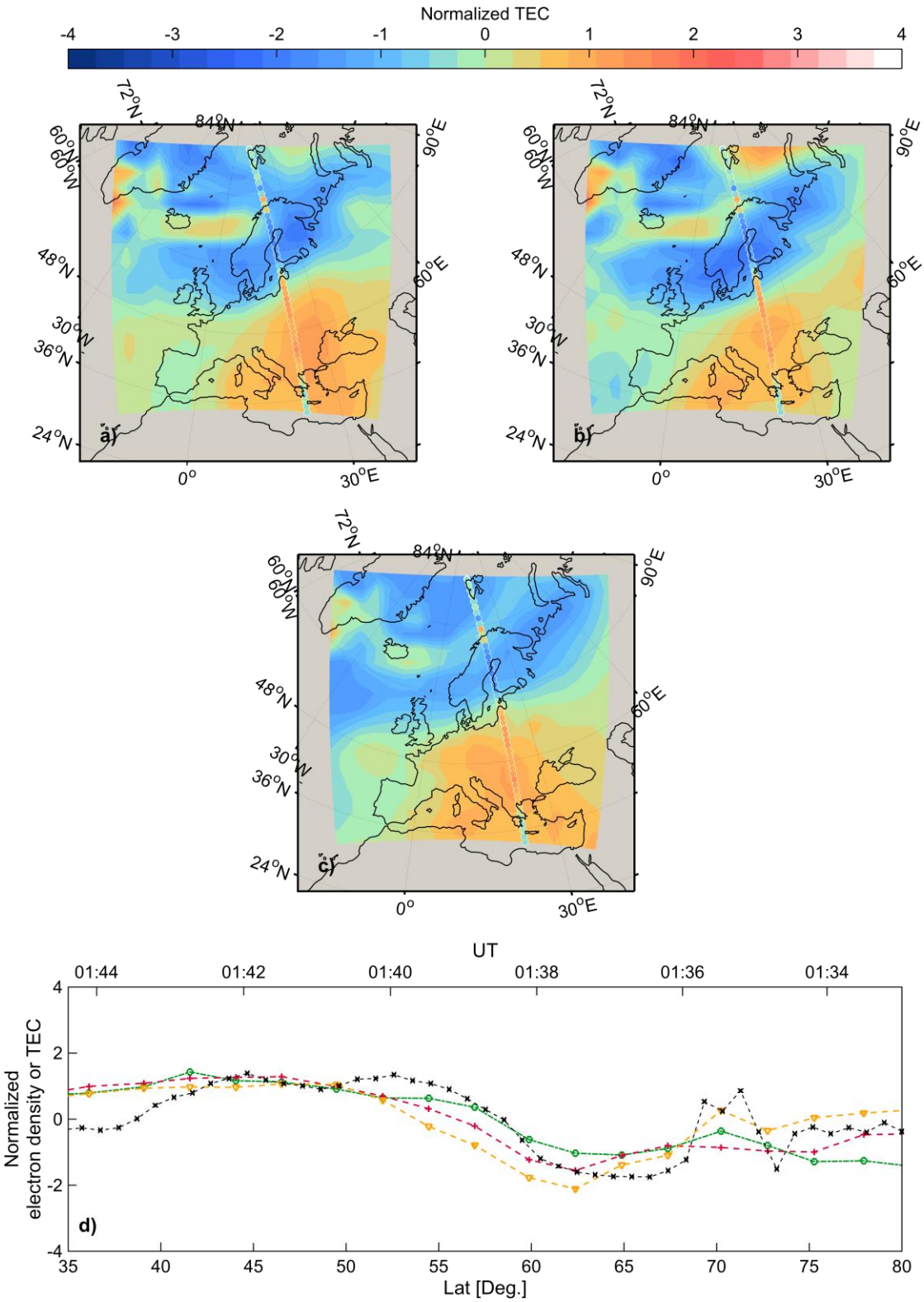

Figure 7.6. Comparison between normalized CHAMP and normalized TEC maps by using a) Daubechies 4; b) discrete Meyer; c) MIDAS with the time-smooth algorithm. CHAMP is shown in terms of normalized electron density as coloured circles over the reconstruction maps. The colour is proportional to the normalized electron density. The bottom plot shows, instead, the normalized electron density (black) and normalized TEC for Daubechies 4 (red) and discrete Meyer (orange) and MIDAS (green) along the satellite CHAMP pass. The plot shows two axes: the latitude (degrees) and the CHAMP time when the measurement was taken. TEC maps are calculated at 01:40 UT of 22 October 2009.

The disagreement between the MIDAS and wavelets methods is due to the different regularization. In MIDAS the solution tends to be smoothed, whereas with the wavelets method, the solution can exploit the multi-resolution analysis. The reconstruction with wavelets seems to be more structured, as opposed to the more smoothed reconstruction from MIDAS. In fact, if a strong correlation is found between a certain wavelet's scale and the corresponding ionospheric structure, a high coefficient is expected in the reconstruction. If this is related with the concept of sparsity, it can be found that not all the scales of the wavelets are used, but only the ones that have a strong similarity with the real structures detected from the observations. This motivates the use of a sparse-promoting algorithm (in this example applied in space only) to estimate the solution by using wavelets.

The following figure shows the same reconstructions but with the time-sparse algorithm for DB4 (Figure 7.7a) and DM (Figure 7.7b) with SM (Figure 7.7c) as comparison.

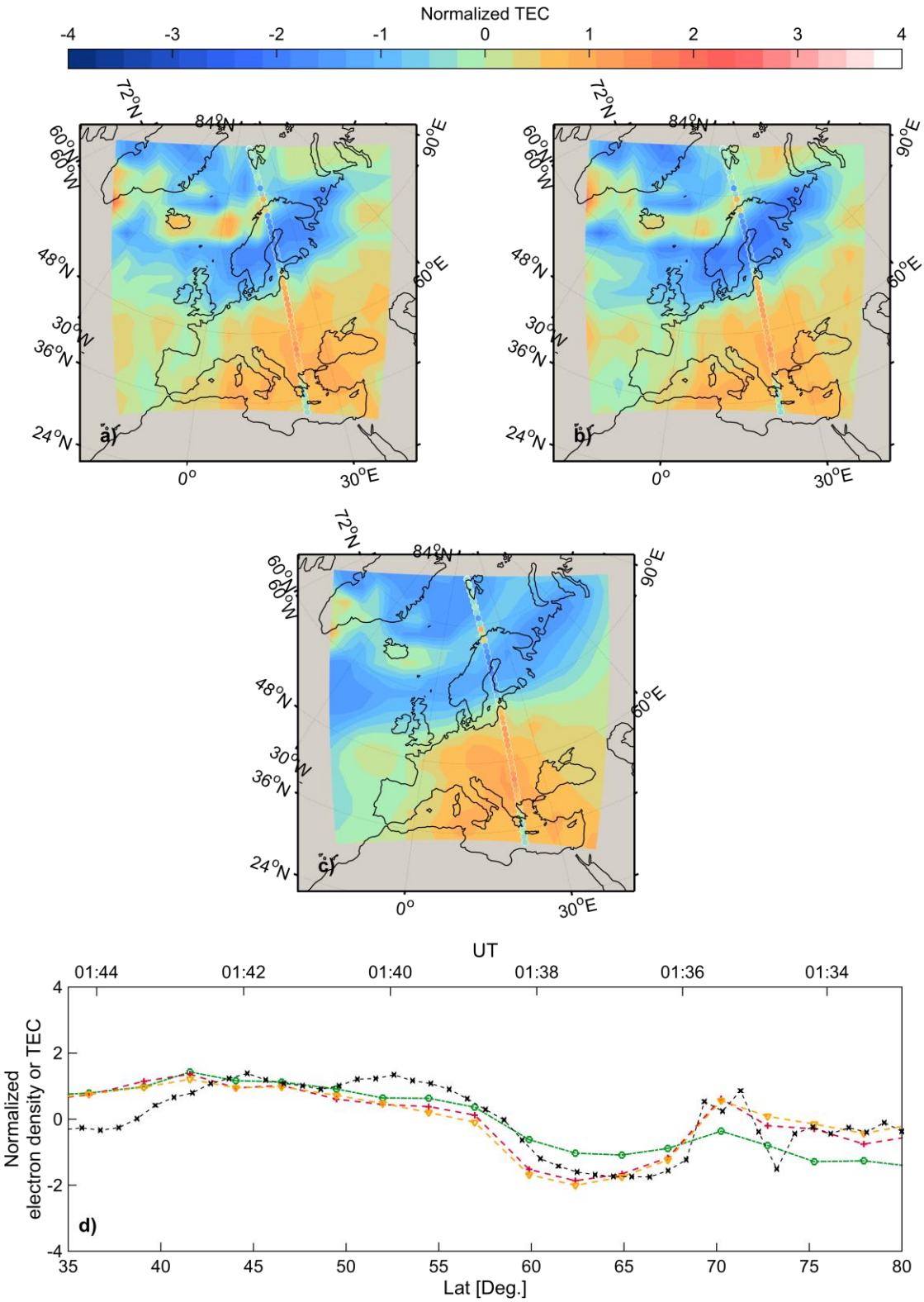

Figure 7.7. Comparison between normalized CHAMP and normalized TEC maps by using a) Daubechies 4; b) discrete Meyer; c) MIDAS with the time-sparse algorithm. CHAMP is shown in terms of normalized electron density as coloured circles over the reconstruction maps. The colour is proportional to the normalized electron density. The bottom plot shows, instead, the normalized electron density (black) and normalized TEC for Daubechies 4 (red) and discrete Meyer (orange) and MIDAS (green) along the satellite CHAMP pass. The plot shows two axes: the latitude (degrees) and the CHAMP time when the measurement was taken. TEC maps are calculated at 01:40 UT of 22 October 2009.

The reconstructions look similar to the previous case and both DB4 and DM seem to do the best reconstructions in comparison with CHAMP. The structures resolved by the time-sparse algorithm show the most significant differences that can be witnessed over North Europe in comparison with the time-smooth algorithm (Figure 7.6). This is due to the 3-dimensional wavelets that allow the solution to potentially better capture the structures. This could not be obtained with the time-smooth algorithm (Figure 7.6) where the regularization inhibited the solution to have sharp variations of TEC over time.

For each case study a total of 4 reconstructions were available every 10 minutes in order to cover the time window used from CHAMP to cross the grid. Those maps were used to extract the closest value in time and space to the one measured from CHAMP. In order to get the same sampled points, as in CHAMP, two interpolations were performed in space and time with a spline polynomial function. Those results are shown in Figure 7.8a for the normalized TEC, and Figure 7.8b shows the normalized TEC gradient for the time-smooth regularization.

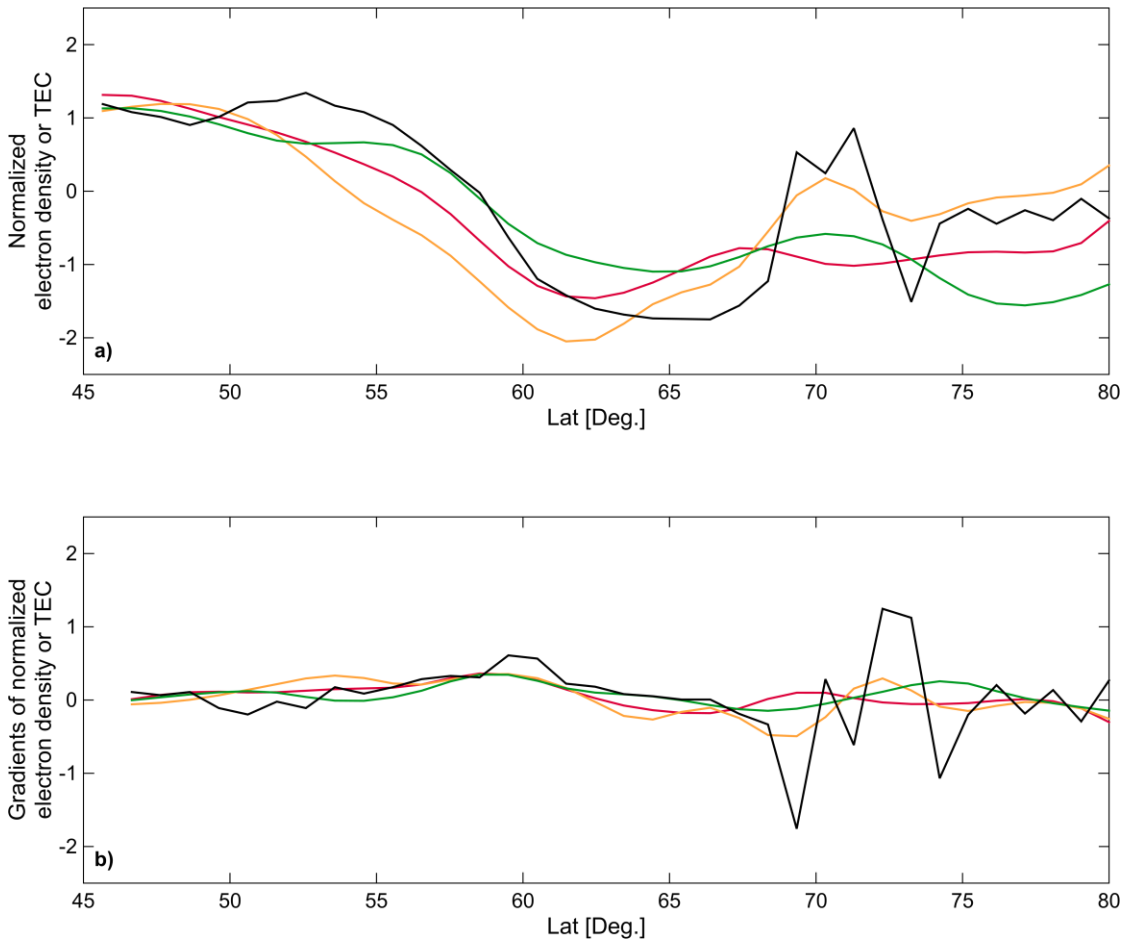

Figure 7.8. Normalized (a) TEC and (b) TEC gradients interpolated over time and space with a spline polynomial function for Daubechies 4 (red), discrete Meyer (orange) and MIDAS (green). The CHAMP normalized electron density (black) is also shown. The results were calculated for the 22nd October 2009 with the time-smooth regularization.

Figure 7.8a shows the normalized TEC for DB4 (red), DM (orange) and SM (green). The normalized electron density for CHAMP is shown with a black line. It can be noticed that wavelets produce the deepest trough, whilst DM produces the best representation of the enhancement at 70 degrees in latitude. This can also be seen in the gradients (Figure 7.8b).

The same results are shown for the time-sparse algorithm in Figure 7.9a and Figure 7.9b for the normalized TEC and normalized TEC gradient respectively. In this case there is an improvement of the normalized TEC, in particular for DB4. The two wavelets produce the best results with the deepest trough (Figure 7.9a) and strongest gradient (Figure 7.9b) as seen in CHAMP in comparison to SM.

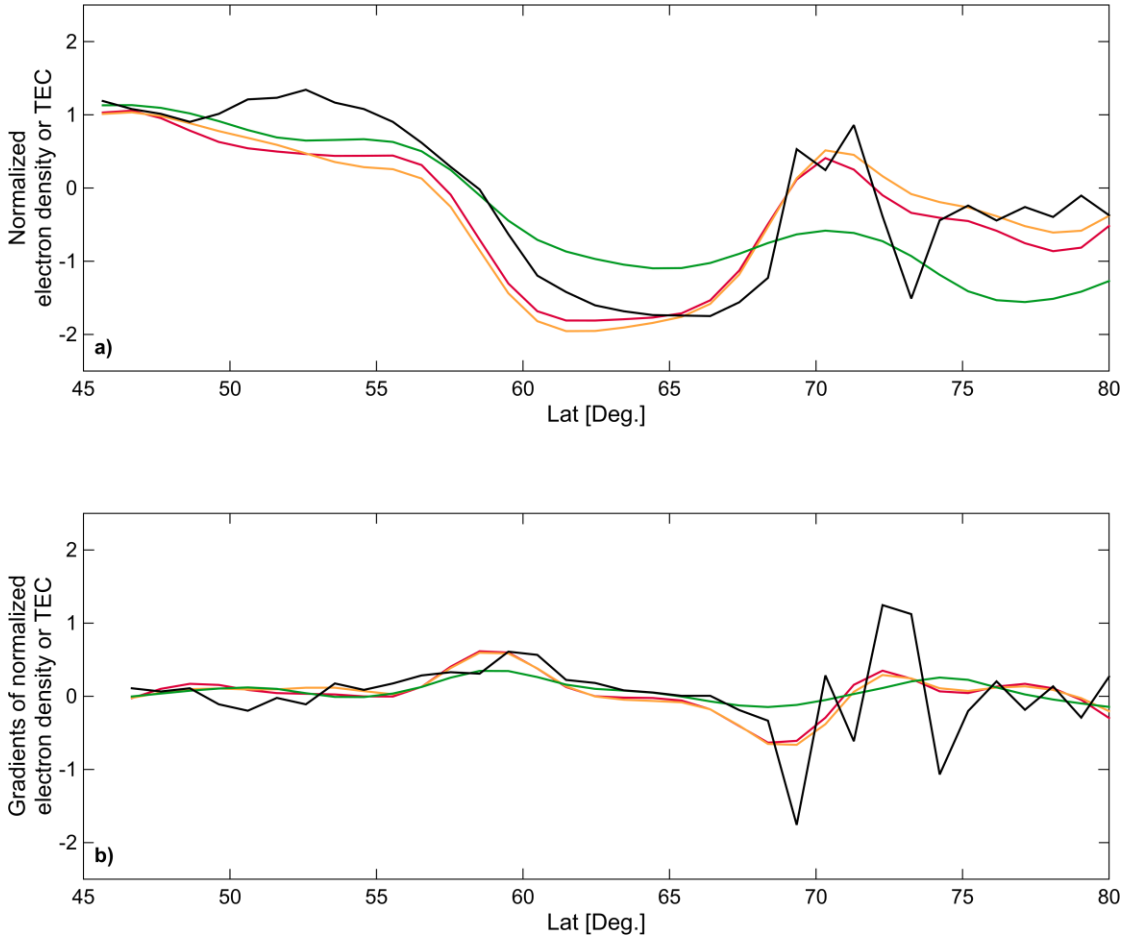

Figure 7.9. Normalized (a) TEC and (b) TEC gradients interpolated over time and space with a spline polynomial function for Daubechies 4 (red), discrete Meyer (orange) and MIDAS (green). The CHAMP normalized electron density (black) is also shown. The results were calculated for the 22nd October 2009 with the time-sparse regularization.

It is observed that the outer wall at 73 degrees in latitude of the structure centred at 70 degrees is not perfectly recovered. This problem is consistent with all the reconstructions produced in the proposed five case studies as they share a similar ray coverage map where there is lack of data coverage in that area (see Figure 7.2).

A summary of the results for this first case study is provided in Table 7-1. The ability to measure the advantages of a method over the other one is difficult to show with a single metric. For this reason, different metrics, where CHAMP was taken as the reference, were used in the table. The root mean squared (RMS) error was used to provide an error of the reconstructions in terms of TEC and TEC gradients.

Unfortunately, the RMS error is well known to have strong limitations to differentiate between different types of errors and also to be highly susceptible to large errors, even if they are much localized or sporadic. This may be is in favour of the ℓ_2 regularization used in SM as it aims to minimize the presence of strong variations. On the other hand, wavelets do not have this limitation and strong variations are allowed (if they are not detected as artefacts by wavelets) depending on data coverage.

Therefore, the RMS error is a good index that may favour smoothness instead of structured details. In the same way, the maximum value of the cross-correlation function (C_{MAX}, normalized to 100) was used, which is a good indicator of the similarity between CHAMP and the reconstructions but may fail for similar reasons. The Mean Structural Similarity (MSSIS) index was also used [*Zhou et al.*, 2004]. The MSSIS index was developed in order to have a measure that could better represent the quality of an image using one as reference. The index is calculated as an averaged multiplication between three different measures: the mean intensity, the signal contrast and the structure comparison. The two different measurements that were used for CHAMP and the reconstructions required a normalization of their values in order to make them comparable in terms of fluctuation. The normalization was done by removing the averaged value and dividing the result by the standard deviation. Therefore, the mean intensity and the contrast comparison, which are basically the average value and the standard deviation, would be meaningless. For this reason the only structure comparison $\boldsymbol{s}$ was used:

$$\boldsymbol{s}(\boldsymbol{x},\boldsymbol{y}) = \frac{\boldsymbol{\sigma_{xy}} + C_3}{\boldsymbol{\sigma_x}\boldsymbol{\sigma_y} + C_3} \tag{7.12}$$

where $\boldsymbol{y}$ is the reference signal, $\boldsymbol{x}$ the signal to be compared, $\boldsymbol{\sigma_{xy}}$ the covariance between $\boldsymbol{x}$ and $\boldsymbol{y}$, $\boldsymbol{\sigma_x}$ and $\boldsymbol{\sigma_y}$ the variance of $\boldsymbol{x}$ and $\boldsymbol{y}$ and C_3 a constant term that prevent the case when $\boldsymbol{\sigma_x}\boldsymbol{\sigma_y}$ is zero. The MSSIS index is calculated for each point of the normalized profiles (both the reconstructed and CHAMP profiles) by using also a weight based on a Gaussian-shaped function with standard deviation 1.5. The MSSIS index is then calculated as:

$$MSSIS = \frac{1}{M}\sum_{j=1}^{M} s_j \tag{7.13}$$

where M corresponds to the number of point in the profiles.

Table 7-1 shows that there are not particular differences between the methods in terms of RMS error. The cross-correlation index (normalized to 100), instead, indicates that the time-sparse algorithm performs better than the other two, i.e. MIDAS and the time-smooth algorithm. This is also confirmed by the MSSIS index for the normalized TEC gradients, whilst for the normalized TEC it shows better performances of MIDAS followed by the time-sparse algorithm. The main interesting point in the table is given by the percentage of basis functions used in the reconstruction, which is remarkably small for wavelets reconstructions.

Table 7-1. RMS error of the VTEC map obtained with MIDAS and wavelets using CHAMP as reference for the 22nd October 2009. The maximum cross-correlation values (C_{MAX}, normalized to 100) between the reconstructed values and CHAMP, and the MSSIS index are also included. The percentage of basis functions with non-zero coefficients is also shown. For each basis function the time-smooth (TS) and time-sparse (TP) results are illustrated.

	Type	Basis Functions	RMS error	CMAX	MSSIS	Percentage number of basis
TEC	SM	No	0.71	78.00	0.42	100.00
	TS	DB4	0.65	80.42	0.19	14.50
	TP		0.49	91.27	0.37	25.36
	TS	DM	0.67	85.14	0.31	10.27
	TP		0.52	92.60	0.34	14.45
DTEC	SM	No	0.48	2.63	0.18	100.00
	TS	DB4	0.51	1.31	0.19	14.50
	TP		0.43	7.72	0.38	25.36
	TS	DM	0.44	6.26	0.30	10.27
	TP		0.42	7.78	0.40	14.45

7.5.2 Other case studies

The following section illustrates the results from other four case studies for the days of:

- 20 November 2009
- 05 December 2009
- 14 July 2009
- 23 July 2009

For each day, TEC maps obtained using; SM, the time-smooth and time-sparse (with DB4 and DM) techniques are shown. Normalized TEC maps are not shown in this section but can be found in Appendix C. The comparison of CHAMP is instead discussed with the normalized TEC and normalized TEC gradient profile. Finally, the indices RMS error, C_{MAX}, MSSIS and number of basis functions are also reported as in the previous section.

The following figures show TEC maps calculated every 30 minutes for November 2009 (Figure 7.10), December 2009 (Figure 7.11) and July 2009 (Figure 7.12 and Figure 7.13). Each figure illustrates the reconstructions for SM (a-b), time-smooth regularization with DB4 (d-f) and DM (g-i) and time-sparse regularization with DB4 (j-l) and DM (m-o).

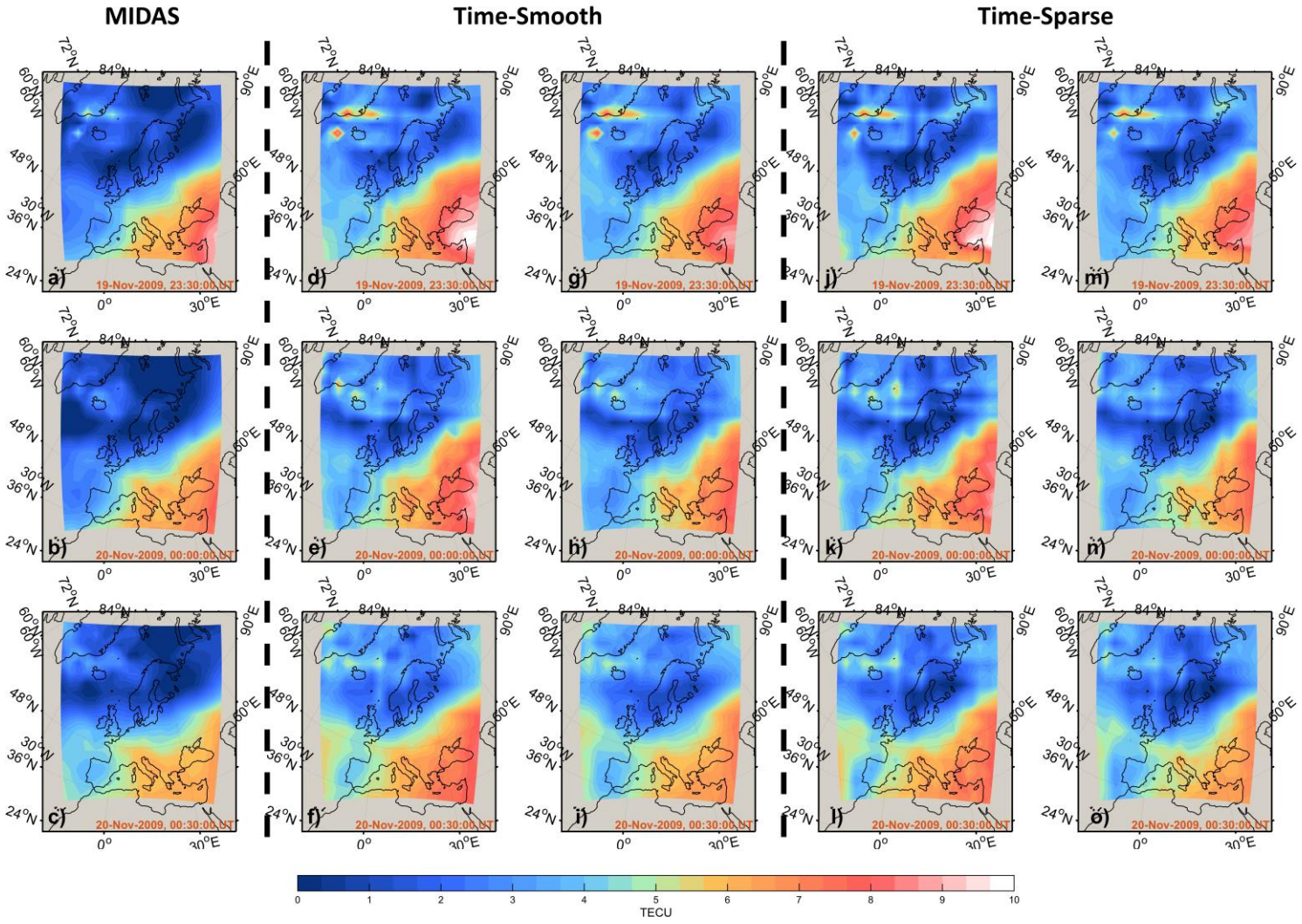

Figure 7.10. Reconstructions obtained with a-c) standard MIDAS (left), time-smooth regularization (centre) with d-f) Daubechies 4 and g-i) discrete Meyer and time-sparse regularization (right) with j-l) Daubechies 4 and m-o) discrete Meyer. The images show TEC maps every 30 minutes starting from 00:00 UT of 20 November 2009.

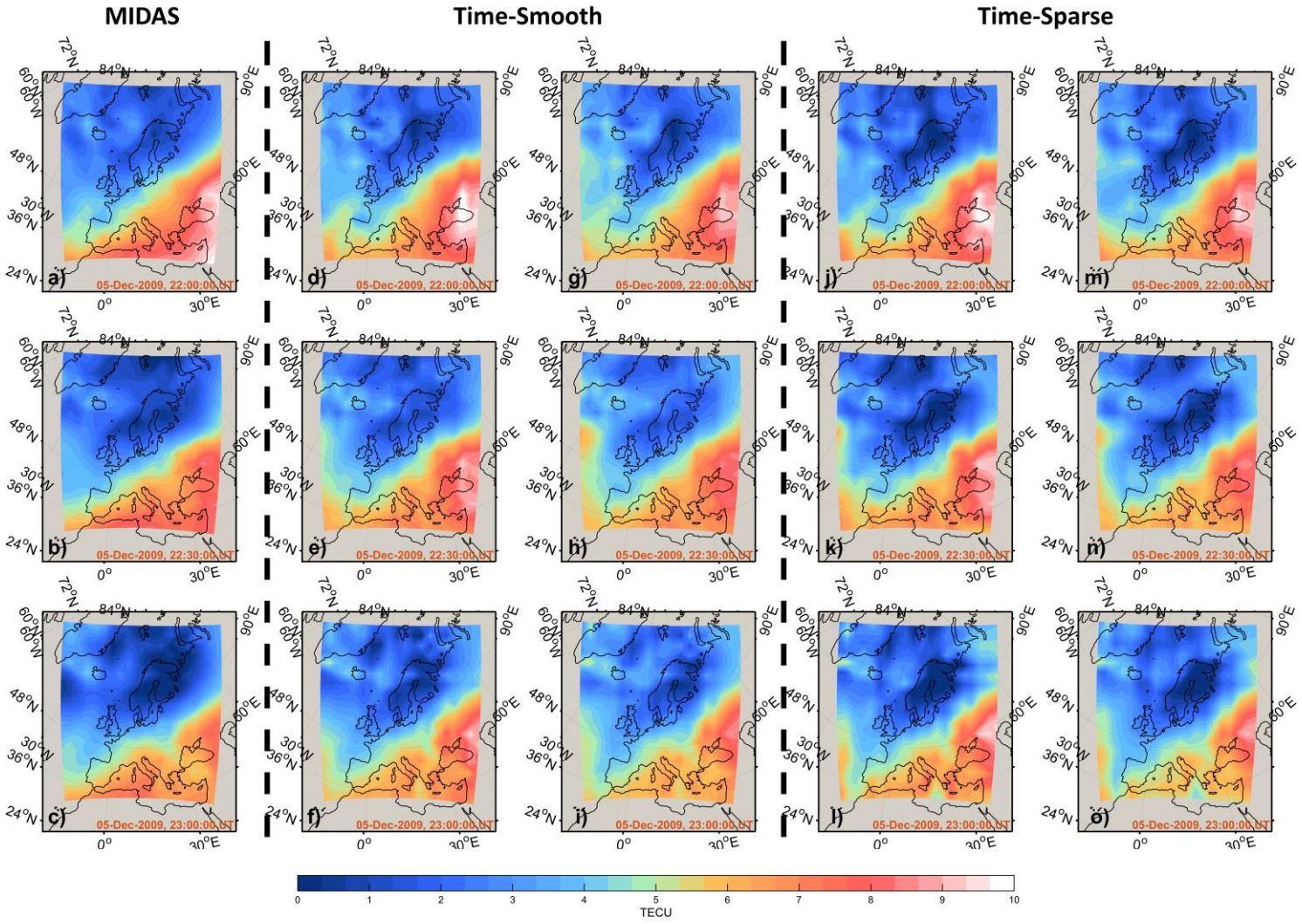

Figure 7.11. Reconstructions obtained with a-c) standard MIDAS (left), time-smooth regularization (centre) with d-f) Daubechies 4 and g-i) discrete Meyer and time-sparse regularization (right) with j-l) Daubechies 4 and m-o) discrete Meyer. The images show TEC maps every 30 minutes starting from 22:00 UT of 05 December 2009.

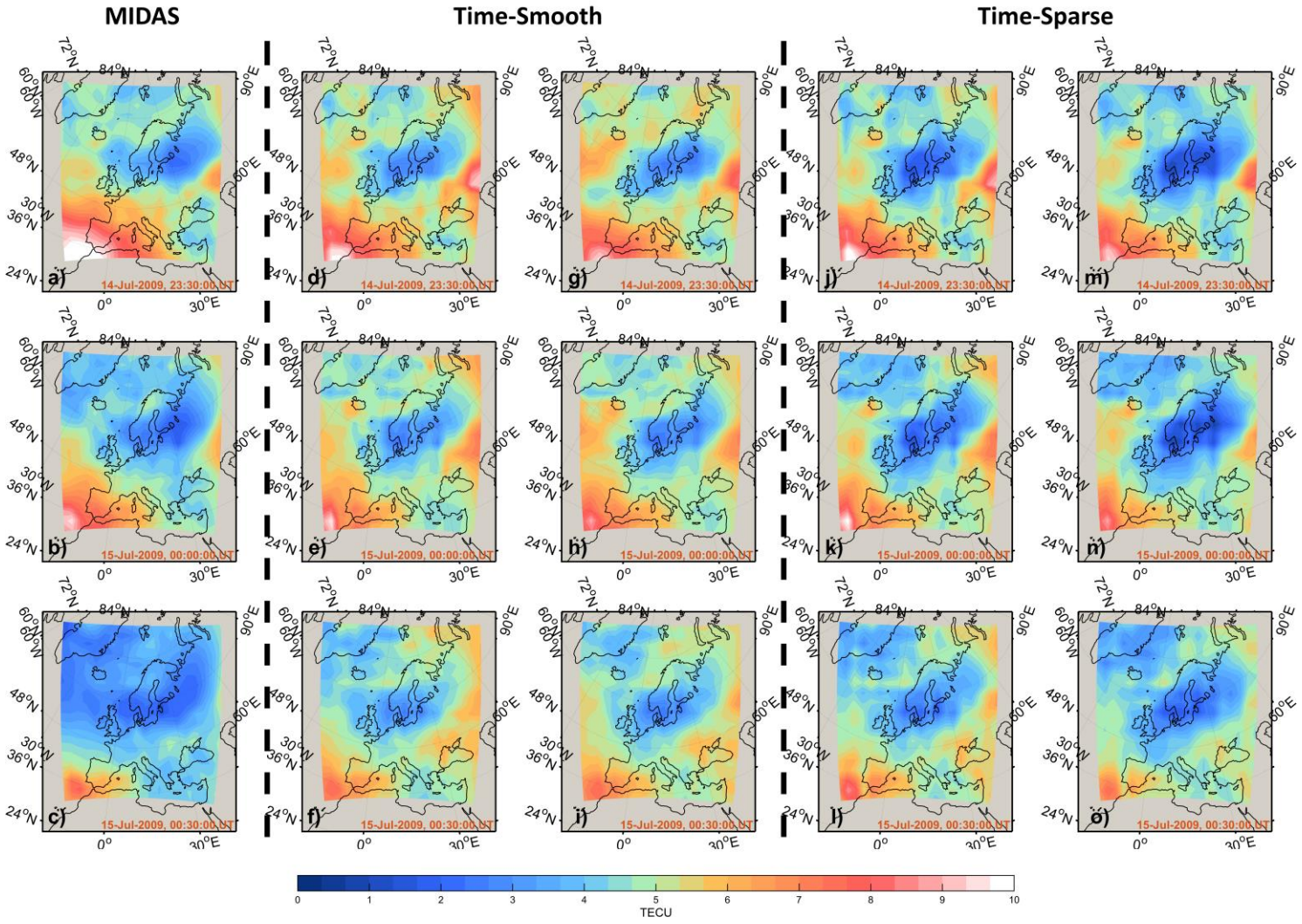

Figure 7.12. Reconstructions obtained with a-c) standard MIDAS (left), time-smooth regularization (centre) with d-f) Daubechies 4 and g-i) discrete Meyer and time-sparse regularization (right) with j-l) Daubechies 4 and m-o) discrete Meyer. The images show TEC maps every 30 minutes starting from 23:30 UT of 14 July 2009.

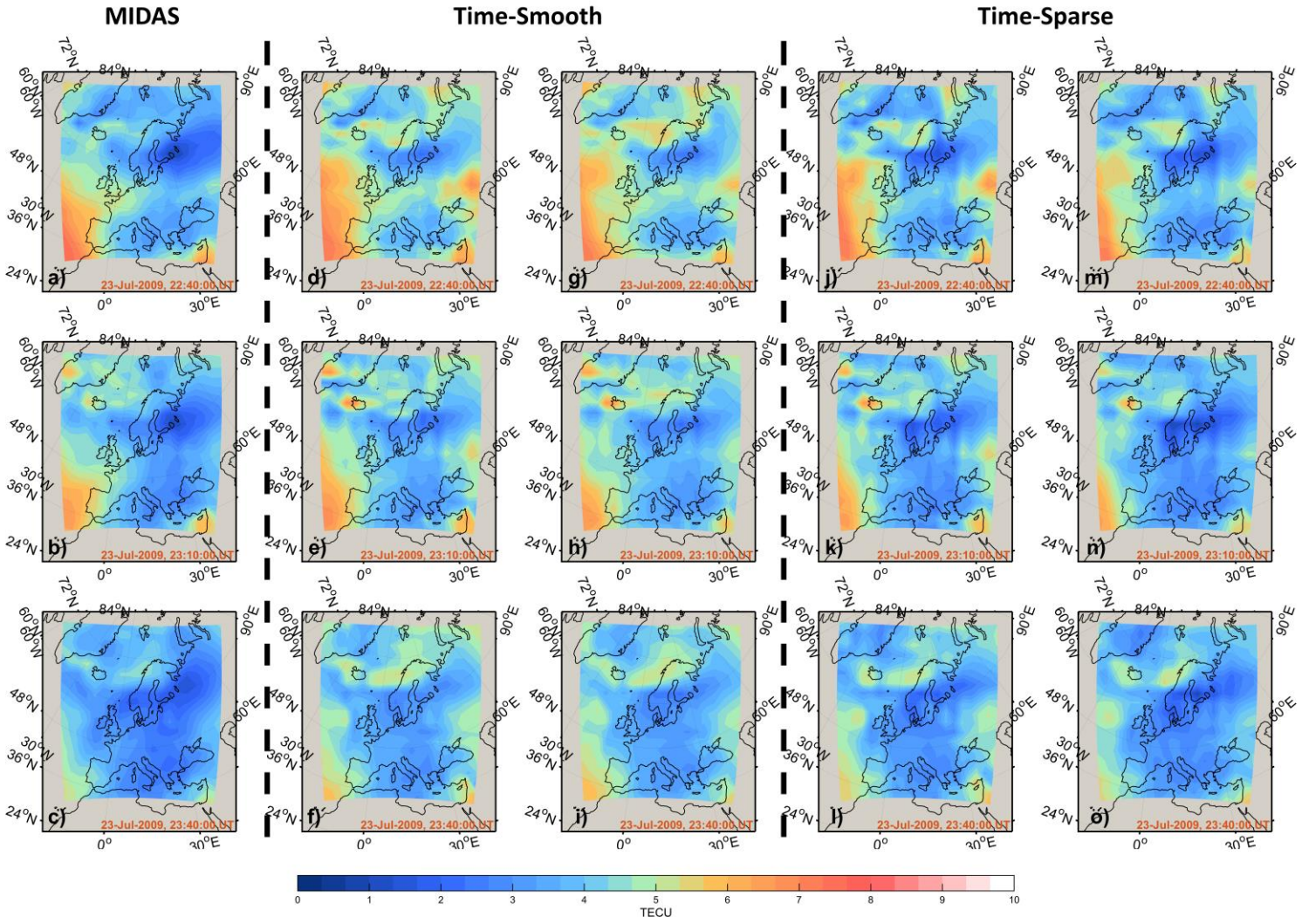

Figure 7.13. Reconstructions obtained with a-c) standard MIDAS (left), time-smooth regularization (centre) with d-f) Daubechies 4 and g-i) discrete Meyer and time-sparse regularization (right) with j-l) Daubechies 4 and m-o) discrete Meyer. The images show TEC maps every 30 minutes starting from 22:40 UT of 23 July 2009.

It can be seen that for the bulk of TEC in Southern Europe, the three methods generally produce similar reconstructions. In Figure 7.10 a TEC enhancement over Greenland can be recognized from the three methods. However, wavelets generate more distinctive structures close to Iceland and Greenland and a TEC enhancement on the north of Europe. Some structures over Greenland are more visible with DB4 although still present in DM, in particular at 00:00 UT. The reconstructions from SM exhibit a TEC enhancement at high latitudes of the grid that fades quickly in the following half-hour. Wavelets, however, indicate a more complex ionosphere with structures that become smoother over a time span of 1 hour. Also in this case, wavelets show two troughs; the first one at 62 degrees north and a second one located at 68 degrees north and 0 degrees west. This will be more visible on the normalized TEC plot in the following analysis.

Figure 7.11, instead, shows the closest similarity between MIDAS and wavelets. The reconstructed scenario is similar to the previous ones with a trough located around 65 degrees in close proximity of Norway. However, the evolution of the ionosphere is slower and there are no particular differences between the reconstructions at 22:00 UT and 23:00 UT.

Figure 7.12 and Figure 7.13 illustrate the cases where SM and wavelets reconstruct two different scenarios. In particular, Figure 7.12 shows an evolution of the ionosphere where wavelets estimate more TEC than SM. This enhancement persists until the end of the reconstruction, i.e. at 00:30 UT on 15 July 2009. At that time, in fact, SM has a minimum TEC over North Europe, whilst the wavelets reconstruct a TEC enhancement over the Northern Scandinavian regions. Similarly, Figure 7.13 shows an initial different scenario between SM and wavelets at 22:40 UT that evolves, then, similarly during the next hour. However, a trough located at 63 degrees in latitude is clearly more visible with wavelets.

A proper look at these maps also shows that TP always reconstructs a deeper trough than TS and SM. The reason is certainly caused by the different time regularization employed by the two wavelet-based methods. As it will be seen later, TS uses a significant higher number of basis functions (double or more) than TP. The little difference between TS and TP reconstructions, mean that more basis functions are needed by TS in order to reconstruct a similar variability in the ionosphere as in TP. Therefore, the smoothness in time has a strong impact in the spatial scale of the structures. This can be seen in the following picture. Figure 7.14 shows the reconstruction for the 23 July 2009 at 23:10 for TS with DB4 (a) and DM (c) and for TP with DB4 (b) and DM (d). The regularization parameter was increased to 5E-3 for TS to force the solution using the same number of basis functions of TP.

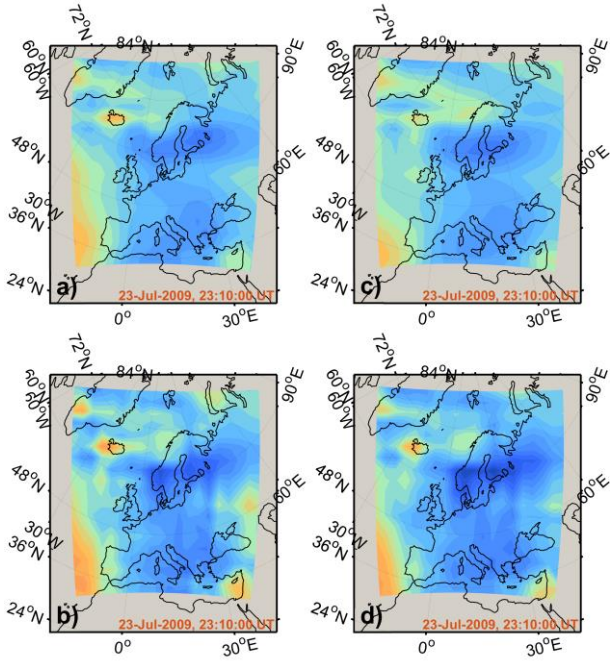

Figure 7.14. Reconstruction obtained with time-smooth regularization with a) Daubechies 4 and c) discrete Meyer and with time-sparse regularization with b) Daubechies 4 and d) discrete Meyer. The regularization parameter was selected differently between the two methods (TS and TP) in order to have a comparable number of basis functions used in the reconstruction.

The figure above shows that the reconstruction looks smoother for TS than TP, when a similar number of basis functions are used between the different solutions. This confirms the smoothing effect of a time-smooth regularization over the resolved spatial scales in the reconstruction. The percentage of basis functions were 11% for DB4 and 7% for DM with TS, whilst 18% for DB4 and 11% for DM with TP.

More basis functions allow the reconstruction of smaller scale structures but with the risk of more artefacts and noise. Furthermore, the sparse regularization algorithm relies on the compactness of the ionospheric decomposition through wavelets. Therefore, the efficiency of the algorithm to resolve different scale structures of the ionosphere can diminish if more basis functions are required for the reconstruction.

Figure 7.15 shows the normalized TEC (a-d) and normalized TEC gradients (e-h) obtained from a series of four reconstructions for the time-smooth regularization. Those values were then interpolated over time and latitude/longitude in order to get the estimated values at the same spatial and temporal sampling of CHAMP. The same information is shown in Figure 7.16 for the time-sparse regularization.

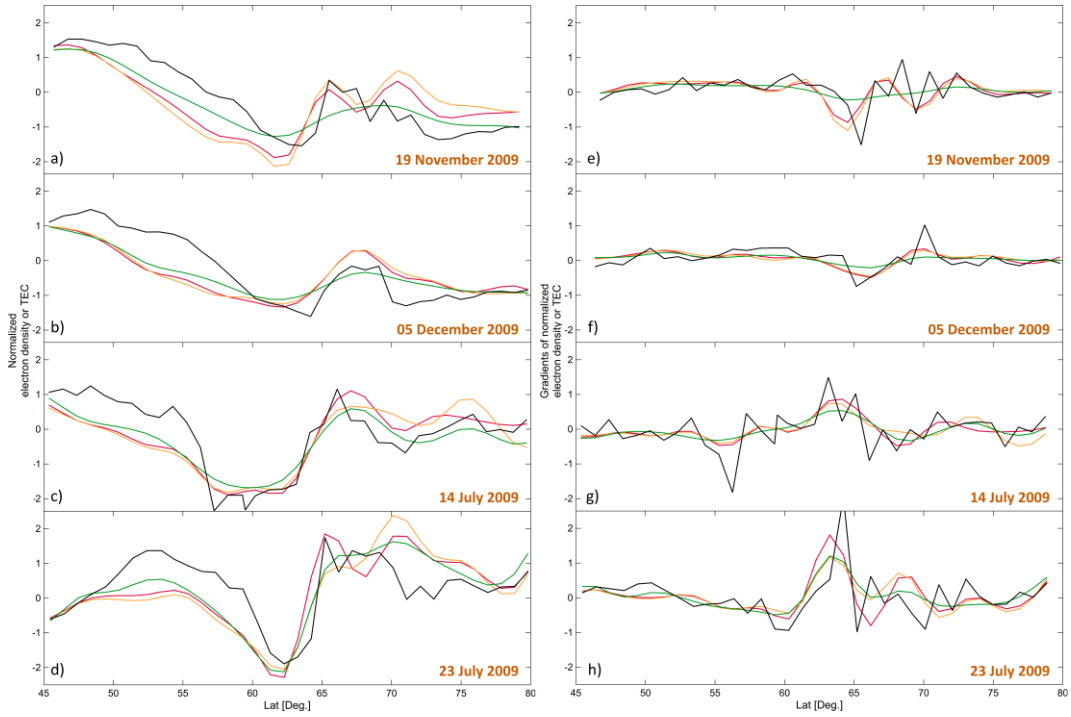

Figure 7.15. Normalized (a-d) TEC and (e-h) TEC gradients interpolated over time and space with a spline polynomial function for Daubechies 4 (red), discrete Meyer (orange) and MIDAS (green). The CHAMP normalized electron density (black) is also shown. The results were calculated for different days with the time-smooth regularization.

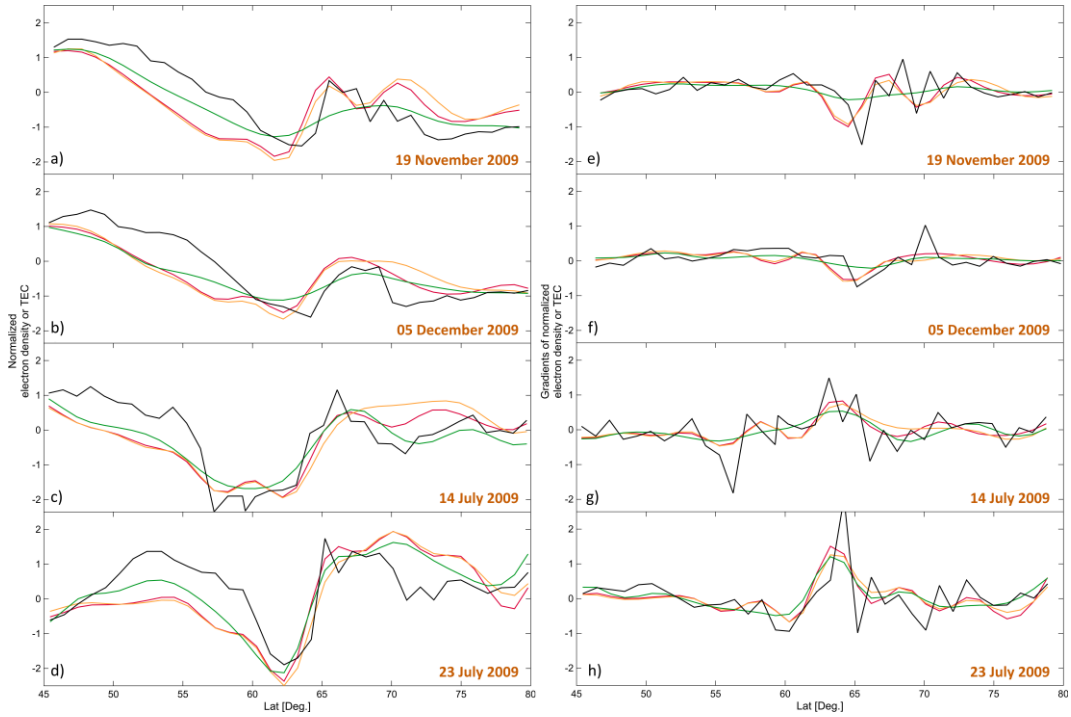

Figure 7.16. Normalized (a-d) TEC and (e-h) TEC gradients interpolated over time and space with a spline polynomial function for Daubechies 4 (red), discrete Meyer (orange) and MIDAS (green). The CHAMP normalized electron density (black) is also shown. The results were calculated for different days with the time-sparse regularization.

It can be seen from Figure 7.15 and Figure 7.16 that SM (green) produces, generally, good ionospheric reconstructions in comparison with CHAMP. However, the SM profiles tend to be smoother than wavelets (red and orange), and, occasionally, with less details than CHAMP. This is more visible in the normalized TEC gradient profiles in Figure 7.15 and Figure 7.16.

The reconstruction of 19 November 2009 shows the difficulty of SM to estimate the finer structures of the poleward wall of the trough (between 65 and 70 degrees in latitude). Wavelets (with TS and TP), instead, reconstructed a more structured profile and reproduced what could be considered (also according to CHAMP) a second trough at 68 degrees.

The profile from the reconstruction of the 5 December 2009 shows that SM and wavelets (with TS and TP) produce a smooth profile in comparison with CHAMP in terms of normalized TEC. Nevertheless, a gradient around 66 degrees in latitude is more visible with wavelets than with SM.

The last two profiles (14 and 23 July 2009) exhibit more similarities between SM and wavelets for both TS and TP in terms of normalized TEC and normalized TEC gradients. However, wavelets still produce a more structured ionosphere than SM, as shown from the normalized TEC gradients.

The differences between TP and TS are in general very small and are not obvious in the profiles. The most noticeable difference is in the last case study (23 July 2009) where TS (Figure 7.15d) produces a more structured profile than TP (Figure 7.16d). This is due to the highest number of basis functions (as it will be seen later) used by TS. The TS technique created a normalized TEC gradient profile whose values seem to overestimate the normalized gradients measured by CHAMP (Figure 7.15h), whilst they are still in a reasonable range for TP (Figure 7.16h).

The normalized TEC profile for November, December and 23 July 2009 show a good estimation of the gradient of the trough but located a few degrees apart from the position measured by CHAMP. This misalignment is present for both MIDAS and wavelets (with TS and TP). Therefore, it is probably caused by the discretization of the ionosphere into a grid. The resolution of the grid was about 2.5 degrees in both latitude and longitude. Therefore, it is expected that a maximum misalignment equal to 2.5 degrees could exists, which corresponds to what is shown in the profiles. The same limitation can be extended to wavelets, which are, in fact, a discrete representation of the continuous wavelet basis functions.

The differences between DM (orange) and DB4 (red) are minimal. However, Figure 7.15c and Figure 7.15g, and Figure 7.16c and Figure 7.16g illustrate the case where DM and DB4 are

dissimilar. Both images reconstruct the day of 14 July 2009. In this example, DM seems to produce a smoother reconstruction than DB4 with both TP and TS.

The following table summarizes the results obtained from all the case studies and the different methods. The table shows four indices calculated from the normalized TEC (or electron density) and normalized TEC (or electron density) gradient (DTEC) obtained with the reconstructions and CHAMP. The indices are the RMS error, the maximum cross-correlation function (C_{MAX}, normalized to 100) and the MSSIS. The percentage number of basis functions is also included. Each index represents the average value obtained from the case studies previously analysed. The indices for each case study can be found in Appendix D. An average of the four indices is also included in the table. The average is performed after a normalization of the RMS error and the MSSIS indices. In particular, RMS is normalized as $100/(1 + \mathrm{RMS})$, whilst the MSSIS index is multiplied by 100. This ensures that the average is performed on values that vary between 0 and 100, and higher is the value and better is the score.

Table 7-2. The table shows different indices calculated from normalized VTEC maps obtained with standard MIDAS (SM) and wavelets using CHAMP as reference. The indices are the RMS error (values are in TECU), the maximum cross-correlation values (CMAX, normalized to 100) between the reconstructed values and CHAMP and the MSSIS index. The percentage of basis functions with non-zero coefficients is also shown. The time-smooth (TS) and time-sparse (TP) results are illustrated for each basis function and for different case studies. Each index was calculated by averaging the values obtained for each case study (five in total). An average of those (normalized) four indices is also included.

	Type	Basis Functions	RMS error	C_{MAX}	MSSIS	Percentage number of basis	Average
TEC	SM	No	0.59	73.01	0.50	100.00	46.38
	TS	DB4	0.67	72.57	0.49	37.39	61.12
	TP		0.68	71.57	0.50	21.85	64.80
	TS	DM	0.72	70.48	0.45	31.04	60.61
	TP		0.74	70.29	0.42	12.28	64.35
DTEC	SM	No	0.42	5.64	0.35	100.00	27.71
	TS	DB4	0.42	8.00	0.33	37.39	43.50
	TP		0.41	8.41	0.35	21.85	48.13
	TS	DM	0.41	7.91	0.31	31.04	44.73
	TP		0.41	7.64	0.31	12.28	49.28

The above table shows that, as expected from Section 7.5.1, SM produces the best results in terms of RMS error, even if the difference with wavelets is relatively small. The same consideration can be made with the C_{MAX} index. The MSSIS index, instead, shows more similarities between wavelets and SD. On the other hand, it is worth noting the low number of basis functions used by wavelets and in particular by DM with TP. This confirms that wavelets can represent the same information as SM but with less basis functions. Therefore, CIT reconstructions can benefit from sparse regularization methods like FISTA.

The section of Table 7-2 regarding the gradients shows a completely different scenario. In fact, the RMS error is the same between SD and wavelets (with TS and TP) as well as MSSIS (with negligible small variations). However, the C_{MAX} index, shows that wavelets better reconstruct the gradients in the ionosphere, in particular DB4.

The average result is illustrated in the last column of Table 7-2. It combines the four (normalized) indices (RMS error, C_{MAX}, MSSIS and number of basis functions) and provides an averaged indication of the best algorithms in terms of sparsity and reliability (of the reconstruction). This average, therefore, will penalize SM as it does not account for sparsity in the reconstruction. It can be seen, in fact, as DM produces the best value in terms of average. The number of basis functions has been considered as a fundamental parameter in this analysis since a more compact representation has numerous advantages in terms of noise (or artefact) removal and efficiency of sparse regularization.

7.6 Summary

The trough is a well-known feature of the ionosphere that has been extensively studied since the 1960s. Because of this, it has been selected to test the new sparse regularization methods over the standard one. The new methods rely on the multi-resolution analysis applied through wavelets decomposition. Although sparse regularization methods were developed by using MIDAS, the name MIDAS or standard MIDAS (SM) has been kept as a reference to the current tomographic tools available within the University of Bath only. The new methods are instead referred to as time-smooth (TS) and time-sparse (TP) and are linked to the wavelet representation.

Five different experiments have been run. In each of them a comparison between wavelets (DB4 and DM) and SM was proposed. The electron density observed from CHAMP was used as independent measurements. Because of the different measurements between CHAMP (electron density) and tomography (TEC), a normalization formulation was used.

MIDAS showed a very smooth ionosphere and the enhancements in TEC were better visible with wavelets. And this was confirmed by the CHAMP measurements. Wavelets, in fact, were

able to reconstruct a more structured ionosphere, in particular over middle-high latitudes. The two different wavelets used in the case studies, DB4 and DM, produced generally similar results, in particular with the time-sparse algorithm that proved to have better performances than the time-smooth algorithm. The negative effect of the time-smoothing operator over the ability to resolve the spatial scale of ionospheric structures was also shown.

The translation of those results was attempted with the use of four indices: the RMS error, the C_{MAX}, the MSSIS and the percentage number of basis functions. However, the interpretation of those indices was sometime inconsistent or based on small differences between the methods. This caused difficulties in their interpretation and analysis. For this reason, the average of the normalized RMS error, the C_{MAX}, the normalized MSSIS and the percentage number of basis functions was proposed. However, the overall analysis with CHAMP showed that wavelets can reproduce an accurate reconstruction and, potentially, represent better the structures of the ionosphere.

According to the average value of Table 7-2, DM could be considered as the best basis function to represent ionospheric structures. It also used the smallest number of basis function in the reconstructions, which makes DM more coherent with the ionospheric structures than DB4. This makes the sparse regularization algorithm more efficient to detect and to recognize real structures from artefacts (or noise).

The results shown in this chapter gave evidence of the importance of the regularization in time. A better knowledge of the dynamics of the ionosphere could probably improve those aspects. On the contrary, when no a-priori information is available, the time regularization can only rely on the data. Therefore only simple assumptions can be done. And this was the case analysed in this chapter.

Another important aspect was given by the discretization of the grid and basis functions, which caused, occasionally, a misalignment between CHAMP and the reconstructions. This may suggest that a higher resolution could improve the detection of the edges of ionospheric structures. Unfortunately, as discussed in Chapter 6, the data coverage imposes a strong limitation on the maximum resolution that could be obtained in CIT.

Chapter 8
Conclusions and future work

In this project the overall aim was to study, implement and test new approaches to image the ionosphere with slant TEC measurements through Computerized Ionospheric Tomography (CIT). The choice to investigate the use of wavelets was because they had the inherent property to accommodate uneven distribution of observations. The underlying theory behind the ideas was outlined in Chapter 5.

Wavelets were compared with spherical harmonics in order to show the potential to resolve better ionospheric structures with wavelets. The method presented in Chapter 5 contained also the first implementation of sparse regularization with the ℓ_1 norm for ionospheric applications. It was tested with simulated data, at low and high resolution and the method used 2-dimensional wavelets to represent the horizontal variation of TEC. The robustness of the method was investigated with different level of representativity errors, which were introduced in the simulated data in order to emulate the discretization of the ionosphere on a grid. The results showed that the sensitivity of wavelets to this source of error was substantially lower than the sensitivity obtained with spherical harmonics, in particular at high resolution. At low resolution, in fact, both the methods produced a robust reconstruction. The best performances at high resolution were caused by the smaller set of coefficients that wavelets needed to estimate, which were around 9% of the total number of basis functions, in comparison to spherical harmonics that used 100% of the basis functions in the reconstructions.

GNSS measurements are uncalibrated and this required a development in line with other ionospheric tomography algorithms. What was new here was the approach to combining wavelets into a new inversion algorithm that could accommodate relative observations. It was important that offsets were estimated correctly - no matter with which method. This aspect was simulated in Chapter 6, by introducing a random offset to each satellite-receiver measurement pair. As expected the accuracy of the estimation reduced dramatically at high resolution with spherical harmonics. The better performances obtained with wavelets were caused by their compact representation achieved through sparse regularization, which made the inversion more stable than with spherical harmonics. A compact representation, in fact, uses only a little number of basis functions to represent the structures of the ionosphere. In this case, the stability of the inversion is helped with the employment of the sparse regularization, which selects the minimum number of basis functions whose shape has the

highest correlation with the ionospheric structures detected from the observations (through a mapping operator, see Chapter 4). The amount of basis function is chosen according to a threshold parameter.

This point was shown in Chapter 6 with the Multi-Resolution (MR) map. The MR map is a visual representation of the Multi-Resolution Analysis (MRA) property of wavelets, which denotes the ability of wavelets to decompose the ionosphere at different scales - each scale is associated with a wavelet and its position. This allows the identification of the most significant structures (according to a threshold) that compose the ionosphere. The MR map illustrated the smallest scale wavelet used for the reconstruction in each point of the grid. The map showed an evident correlation between the number of observations and the minimum scale of the wavelet that could be used in the reconstruction. The smallest scales, in fact, were used mainly in regions with high data coverage. This supported the motivation that wavelets can adapt according to data distribution. Furthermore, it was interesting to notice that small scale wavelets were not necessarily used where there was good data coverage. Indeed, it also depended on the presence of ionospheric structures with scale comparable to the wavelet scale.

Chapter 6 concluded with some demonstration of imaging high-resolution reconstructions. It showed the limitations of sparse regularization through wavelets with a resolution grid of only 0.5 degrees and the need of another method to reduce the artefacts in the reconstruction. In that scenario, Total Variation (TV) was used to reduce the artefacts but without losing the information about the edges of the structures and assumed the ionosphere to be static. However, it must be said that there may be strong limitations with such a high-resolution reconstruction. In fact, the boundaries of validity of the assumption of static ionosphere can be not clear and difficult to assess and, therefore, more data could be required to ensure a reliable reconstruction.

The ionosphere is a dynamic medium that changes continuously over time. This is an aspect that makes CIT considerably more difficult to solve than a standard tomography problem. The time-dependency was analysed in Chapter 7 where the method based on wavelets was extended to dynamic imaging of the ionosphere in a manner analogous to the MIDAS approach. A comparison was made with the current standard version of the Multi-Instrument Data Analysis System (MIDAS), developed at the University of Bath, by using real data. Results illustrated that a time-smoothing regularization had visible effects on the smoothness of the structures in the spatial scale. It also showed that better defined structures can be obtained when the smoothness in time is replaced with sparsity in time (i.e. time-sparse regularization). This was achieved by implementing 3-dimensional wavelets, where the third

dimension was time. The usage of wavelets in time had multiple advantages. Firstly, it allowed the usage of MRA also in time with the potential to represent the time-variation of TEC more sharply. Furthermore, wavelets in time can produce a more compact representation of the ionosphere. Therefore, the sparse regularization method, which relies on the ability to represent the ionosphere compactly, can be more robust and effective.

The reconstructions were validated against CHAMP but normalization was required in order to compare TEC with the localized electron density measurement of CHAMP. Unfortunately, it must be said that it is very difficult to find a way to compare the improvements given by very different methods with a single general index. For this reason, three indices were provided for the validation: the Root Mean Square (RMS) error, the maximum cross-correlation value (C_{max}, normalized to 100) and the Mean Structural SIMilarity (MSSIM) index. Those indices were used for the normalized TEC and normalized TEC gradient profiles but in some cases those indices were discordant. The results, anyway, showed the applicability of wavelets in CIT and their potential to better resolve ionospheric structures. The five case studies proposed in this chapter confirmed wavelets and sparse regularization as a valid alternative to more standard methods (MIDAS). However, there were cases where wavelets could reconstruct ionospheric structures better then MIDAS. The results also indicated that the Discrete Meyer (DM) wavelet family used the lowest number of basis functions to reconstruct the state of the ionosphere in comparison to Daubechies 4 (DB4). A better compression (i.e. low number of basis functions used in the reconstruction) can lead to better performances of the sparse regularization method. In fact, a good compression is similar to say that the shape of the wavelets matches the shape of ionospheric structures and, therefore, the information content (in this case represented by TEC) extracted through wavelet is more robust to noise (i.e. artefacts). For this reason, DM can be considered a good choice to represent the structures of the ionosphere.

In conclusion, the main contributions in this thesis work can be listed as:

- Implementation and testing of a wavelet reconstruction and sparse regularization method into CIT, which is believed to be the first in ionospheric applications.
- Implementation of a TV technique to preserve the edges of structures at high resolution, which is also new in CIT.
- Extension to sparse regularization in time with 3-dimensional wavelets.
- Identification of DM as the best wavelet family for ionospheric purposes, which was compared in terms of lowest number of basis functions used in the reconstructions.
- Incorporation of a ℓ_2 regularization term into the FISTA algorithm. It allowed the analysis in space caused by a time-smoothing term. It showed an evident spatial

smoothing effect of the resolved ionospheric structures, which was reduced with the time-sparse term.

Further research needs to be done in order to reveal the efficacy of the method based on, for example, geographical locations. In fact, it is well known that structures are more likely to appear in equatorial and high latitude regions. Furthermore, ionospheric perturbations can be associated to geomagnetic indices such as Ap and Kp, which can be used to assess and identify the conditions where the new wavelet-based methods may be more suitable. Those scenarios should be classified according to different ranges of the Kp index, for example 0-3, 3-6 and 6-9.

The sparse regularization method relies on basis functions to extract the information of ionospheric structures from sTEC observations. Therefore, basis functions play an essential role, and the tomographic reconstruction could benefit from different basis functions like wavelet packets, contourlets, and curvelets. Further, CIT is a time-dependent problem and better results could be obtained with basis functions that can adapt to specific instances of the data, which could replace the use of generic basis functions. An example of this is given by the K-SVD, a dictionary learning algorithm, based on SVD, for creating a set of basis functions for sparse representation [*Rubinstein et al.*, 2010].

In future the sparse regularization method can be improved under different aspects:

- Including vertical profiles using a better model. For example the A New Ionospheric Model (ANIMo) [*Da Dalt*, 2015] could be used to improve the vertical profiles in terms of electron density and height peak.
- Implementing a full Data Assimilation (DA) approach by using sparse regularization techniques. A DA approach combines the better definition of the time-dependencies of the ionosphere through a model (e.g. ANIMo) with the ability of sparse regularization methods to resolve ionospheric structures that a model cannot predict. DA relies on a covariance matrix that contains the information of the spatial correlation of the structures in the ionosphere.

Appendix A

This appendix is intended as an extension of Chapter 4, which provides more details about the derivation of the RLS algorithm equations.

A.1 Regularized Least Squares (RLS) Algorithm

The ionospheric tomography is an undetermined problem described by Equation (3.9) its matrix notation. One solution to the problem is obtained by minimizing the misfit functional defined as the ℓ_2 norm of the residual error between the predicted and observed data, i.e.

$$f(\boldsymbol{n}) = \|\tilde{\boldsymbol{z}} - \boldsymbol{z}\|^2 = \|\mathbf{A}\boldsymbol{n} - \boldsymbol{z}\|^2 = (\mathbf{A}\boldsymbol{n} - \boldsymbol{z},\ \ \mathbf{A}\boldsymbol{n} - \boldsymbol{z}) = (\mathbf{A}\boldsymbol{n} - \boldsymbol{z})^{\mathrm{T}}(\mathbf{A}\boldsymbol{n} - \boldsymbol{z}) \tag{A.1}$$

The value which minimizes the functional can be calculated by differentiating Equation (A.1) and putting it equals to zero

$$2(\mathbf{A}\boldsymbol{n} - \boldsymbol{z})^{\mathrm{T}}\mathbf{A} = \mathbf{0}^{\mathrm{T}} \tag{A.2}$$

obtaining the *pseudo-solution* of the system Equation (3.9)

$$\widehat{\boldsymbol{n}} = \left(\mathbf{A}^{\mathrm{T}}\mathbf{A}\right)^{-1}\mathbf{A}^{\mathrm{T}}\boldsymbol{z} \tag{A.3}$$

The squared matrix $\left(\mathbf{A}^{\mathrm{T}}\mathbf{A}\right)$ could be ill conditioned and then make the inversion instable. Andrei Tikhonov in 1977 proposed a solution for ill-posed problems. He introduced the following parametric functional [*Tikhonov and Arsenin*, 1977*Tikhonov and Arsenin*, 1977] that has been here extended to the case were an a-priori information $\boldsymbol{n_0}$ was provided

$$f^{\varrho}(\boldsymbol{n}, \boldsymbol{z}) = \|\mathbf{W}_{\mathrm{d}}\mathbf{A}\boldsymbol{n} - \mathbf{W}_{\mathrm{d}}\boldsymbol{z}\|^2 + \varrho\|\mathbf{W}_{\mathrm{m}}\boldsymbol{n} - \mathbf{W}_{\mathrm{m}}\boldsymbol{n_0}\|^2 \tag{A.4}$$

where $\mathbf{W}_{\mathrm{d}}$ and $\mathbf{W}_{\mathrm{m}}$ are two weighted matrices and the vector $\boldsymbol{n_0}$ can be considered as an initial guess of the data source vector $\boldsymbol{n}$. The solution can be found by minimizing the functional of Equation (A.4) i.e. by taking the derivative and sets it equal to zero

$$\left(\mathbf{A}^{\mathrm{T}}\mathbf{W}_{\mathrm{d}}^{\mathrm{T}}\mathbf{W}_{\mathrm{d}}\mathbf{A} + \varrho\ \mathbf{W}_{\mathrm{m}}^{\mathrm{T}}\mathbf{W}_{\mathrm{m}}\right)\boldsymbol{n} - \mathbf{A}^{\mathrm{T}}\mathbf{W}_{\mathrm{d}}^{\mathrm{T}}\mathbf{W}_{\mathrm{d}}\boldsymbol{z} - \mathbf{W}_{\mathrm{m}}^{\mathrm{T}}\mathbf{W}_{\mathrm{m}}\boldsymbol{n}_0 = 0 \tag{A.5}$$

obtaining the pseudo solution for the data source $\widehat{\boldsymbol{n}}$

$$\widehat{\boldsymbol{n}} = \boldsymbol{n}_0 + \delta\widehat{\boldsymbol{n}} \tag{A.6}$$

where $\boldsymbol{n}_0$ is the initial guess and $\delta\widehat{\boldsymbol{n}}$, obtained by adding and subtracting the term $\mathbf{A}^{\mathrm{T}}\mathbf{W}_{\mathrm{d}}^{\mathrm{T}}\mathbf{W}_{\mathrm{d}}\mathbf{A}\boldsymbol{n_0}$ in Equation (A.6), is

$$\delta\hat{\boldsymbol{n}} = \left(\mathbf{A}^{\mathrm{T}}\mathbf{W}_{\mathrm{d}}^{\mathrm{T}}\mathbf{W}_{\mathrm{d}}\mathbf{A} + \varrho\,\mathbf{W}_{\mathrm{m}}^{\mathrm{T}}\mathbf{W}_{\mathrm{m}}\right)^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{W}_{\mathrm{d}}^{\mathrm{T}}(\boldsymbol{z} - \mathbf{A}\boldsymbol{n}_0) \tag{A.7}$$

i.e. the inversion applied to the difference between the measurement $\boldsymbol{z}$ and the predicted measurement $\mathbf{A}\boldsymbol{n}_0$.

By considering the matrix $\mathbf{W}_{\mathrm{d}}$ an identity matrix and the initial guess $\boldsymbol{n}_0$ a zero vector elements, the pseudo solution becomes

$$\hat{\boldsymbol{n}} = \left(\mathbf{A}^{\mathrm{T}}\mathbf{A} + \varrho\,\mathbf{W}_{\mathrm{m}}^{\mathrm{T}}\mathbf{W}_{\mathrm{m}}\right)^{-1}\mathbf{A}^{\mathrm{T}}\boldsymbol{z} \tag{A.8}$$

According to the choice of $\mathbf{W}_{\mathrm{m}}$, the inversion in Equation (A.7) could still be singular, giving an unstable solution. An even more robust stabilization can be obtained by considering the following functional cost

$$f^{\varrho,\alpha}(\boldsymbol{n}, \boldsymbol{b}) = \|\mathbf{A}\boldsymbol{n} - \boldsymbol{z}\|^2 + \varrho\|\mathbf{W}_{\mathrm{m}}\boldsymbol{n} - \mathbf{W}_{\mathrm{m}}\boldsymbol{n_0}\|^2 + \alpha\|\boldsymbol{n} - \boldsymbol{n_0}\|^2 \tag{A.9}$$

where a term$\|\boldsymbol{n} - \boldsymbol{n_0}\|^2$ has been added to force solution to be like the initial guess according to α. Equation (A.9) can be solved in same way by setting the derivative equal to zero obtaining the following solution

$$\hat{\boldsymbol{n}} = \boldsymbol{n}_0 + \delta\hat{\boldsymbol{n}} \tag{A.10}$$

where $\boldsymbol{n}_0$ is the initial guess and $\delta\hat{\boldsymbol{n}}$

$$\delta\hat{\boldsymbol{n}} = \left(\mathbf{A}^{\mathrm{T}}\mathbf{A} + \varrho\mathbf{W}_{\mathrm{m}}^{\mathrm{T}}\mathbf{W}_{\mathrm{m}} + \alpha\mathbf{I}\right)^{-1}\mathbf{A}^{\mathrm{T}}(\boldsymbol{z} - \mathbf{A}\boldsymbol{n}_0) \tag{A.11}$$

If the source data set is decomposed through some basis functions, i.e.

$$\boldsymbol{b} = \mathbf{A}\boldsymbol{n} = \mathbf{A}\mathbf{K}\boldsymbol{x} \tag{A.12}$$

where $\mathbf{K}$ is an orthonormal matrix, the RLS solution will try to minimize the following functional

$$f^{\varrho,\alpha}(\boldsymbol{x}, \boldsymbol{z}) = \|\mathbf{A}\mathbf{K}\boldsymbol{x} - \boldsymbol{z}\|^2 + \varrho\|\mathbf{W}_{\mathrm{m}}\mathbf{K}\boldsymbol{x} - \mathbf{W}_{\mathrm{m}}\mathbf{K}\boldsymbol{x_0}\|^2 + \alpha\|\mathbf{K}\boldsymbol{x} - \mathbf{K}\boldsymbol{x_0}\|^2 \tag{A.13}$$

By taking the derivative and set it equal to zero it is obtained

$$\hat{\boldsymbol{x}} = \boldsymbol{x}_0 + \delta\hat{\boldsymbol{x}} \tag{A.14}$$

where

$$\delta\hat{\boldsymbol{x}} = \left(\mathbf{K}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{A}\mathbf{K} + \varrho\mathbf{K}^{\mathrm{T}}\mathbf{W}_{\mathrm{m}}^{\mathrm{T}}\mathbf{W}_{\mathrm{m}}\mathbf{K} + \alpha\mathbf{K}^{\mathrm{T}}\mathbf{K}\right)^{-1}\mathbf{K}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}(\boldsymbol{z} - \mathbf{A}\mathbf{K}\boldsymbol{x}_0) \tag{A.15}$$

is the inversion of the difference between the prediction $\mathbf{AK}\boldsymbol{x}_0$ and the observed $\boldsymbol{b}$ measurements. This differential inversion adds to the model the variations which can't be modelled or predicted without taking into account the model itself.

According to this decomposition the model resolution matrix can be written as

$$(\hat{\boldsymbol{x}} - \boldsymbol{x}_0) = \mathbf{R}_{\mathrm{m}}(\boldsymbol{x} - \boldsymbol{x}_0) \tag{A.16}$$

If the resolution matrix $\mathbf{R}_{\mathrm{m}}$ is an identity means that the inversion is adding the true information that is missing from the model.

Appendix B

This appendix derives the calibration matrix that was described in Chapter 5.

B.1 The calibration matrix

The forward problem was defined in Chapter 5 as

$$\boldsymbol{z} = \mathbf{A}\mathbf{K}\boldsymbol{x} + \mathbf{B}\boldsymbol{b} \tag{B.1}$$

where **A** is the projection matrix, $\boldsymbol{z}$ is the observation vector, **K** is the matrix containing the vertical and horizontal basis functions, **B** the projection matrix of the offsets $\boldsymbol{b}$.

The inverse problem is defined through the functional $f(\boldsymbol{x}, \boldsymbol{b})$

$$f(\boldsymbol{x}, \boldsymbol{b}) = \|\boldsymbol{z} - \mathbf{H}\boldsymbol{x} - \mathbf{B}\boldsymbol{b}\|^2 + \alpha\, \mathcal{P}(\boldsymbol{x}, \boldsymbol{b}) \tag{B.2}$$

where $\mathcal{P}(\boldsymbol{x}, \boldsymbol{b})$ is the regularization term weighted by the parameter α, and $\mathbf{H} = \mathbf{A}\mathbf{K}$.

Equation (B.2) can be rewritten as

$$f(\boldsymbol{y}) = \boldsymbol{z}^T\boldsymbol{z} + \boldsymbol{y}^T\boldsymbol{C}^T\boldsymbol{C}\boldsymbol{y} - 2\boldsymbol{z}^T\boldsymbol{C}\boldsymbol{y} + \alpha\, \mathcal{P}(\boldsymbol{y}) \tag{B.3}$$

where $\boldsymbol{y} = \begin{bmatrix}\boldsymbol{x}\\ \boldsymbol{b}\end{bmatrix}$ and $\boldsymbol{C} = [\mathbf{H} \quad \mathbf{B}]$.

The pseudo-solution is then found by minimizing Equation (B.3), i.e. by taking the derivative and set it to zero. Thus,

$$\left[\frac{\partial f(\boldsymbol{y})}{\partial \boldsymbol{y}}\right]^T = 2\boldsymbol{C}^T\boldsymbol{C}\hat{\boldsymbol{y}} - 2\boldsymbol{C}^T\boldsymbol{z} + \alpha\frac{\partial \mathcal{P}(\hat{\boldsymbol{y}})}{\partial \hat{\boldsymbol{y}}} = 0 \tag{B.4}$$

Equation (B.4) can be rewritten as

$$\frac{\alpha}{2}\frac{\partial \mathcal{P}(\hat{\boldsymbol{y}})}{\partial \hat{\boldsymbol{y}}} + \begin{bmatrix}\mathbf{H}^\mathbf{T}\mathbf{H} & \mathbf{H}^\mathbf{T}\mathbf{B}\\ \mathbf{B}^\mathbf{T}\mathbf{H} & \mathbf{B}^\mathbf{T}\mathbf{B}\end{bmatrix}\begin{bmatrix}\hat{\boldsymbol{x}}\\ \hat{\boldsymbol{b}}\end{bmatrix} = \begin{bmatrix}\mathbf{H}^\mathbf{T}\boldsymbol{z}\\ \mathbf{B}^\mathbf{T}\boldsymbol{z}\end{bmatrix} \tag{B.5}$$

With the assumption that the regularization term does not involve $\hat{\boldsymbol{b}}$, Equation B.5 can be written as

$$\frac{\alpha}{2}\begin{bmatrix}\frac{\partial \mathcal{P}(\hat{\boldsymbol{x}})}{\partial \hat{\boldsymbol{x}}}\\ 0\end{bmatrix} + \begin{bmatrix}\mathbf{H}^\mathbf{T}\mathbf{H} & \mathbf{H}^\mathbf{T}\mathbf{B}\\ \mathbf{B}^\mathbf{T}\mathbf{H} & \mathbf{B}^\mathbf{T}\mathbf{B}\end{bmatrix}\begin{bmatrix}\hat{\boldsymbol{x}}\\ \hat{\boldsymbol{b}}\end{bmatrix} = \begin{bmatrix}\mathbf{H}^\mathbf{T}\boldsymbol{z}\\ \mathbf{B}^\mathbf{T}\boldsymbol{z}\end{bmatrix} \tag{B.6}$$

From the second "row" of Equation (B.6) it is possible to calculate the estimated offsets $\hat{\boldsymbol{b}}$ as

$$\widehat{\boldsymbol{b}} = \left(\mathbf{B}^{\mathbf{T}}\mathbf{B}\right)^{-\mathbf{1}}\mathbf{B}^{\mathbf{T}}(\boldsymbol{z} - \mathbf{H}\widehat{\boldsymbol{x}}) \tag{B.7}$$

By substituting Equation (B.7) into the first "row" of Equation (B.6) the following formulation is obtained

$$\mathbf{H}^{\mathbf{T}}\left(\mathbf{I} - \mathbf{B}\left(\mathbf{B}^{\mathbf{T}}\mathbf{B}\right)^{-\mathbf{1}}\mathbf{B}^{\mathbf{T}}\right)\mathbf{H}\widehat{\boldsymbol{x}} - \mathbf{H}^{\mathbf{T}}\left(\mathbf{I} - \mathbf{B}\left(\mathbf{B}^{\mathbf{T}}\mathbf{B}\right)^{-\mathbf{1}}\mathbf{B}^{\mathbf{T}}\right)\mathbf{H}\widehat{\boldsymbol{x}} + \frac{\alpha}{2}\frac{\partial\mathcal{P}(\widehat{\boldsymbol{x}})}{\partial\widehat{\boldsymbol{x}}} = 0 \tag{B.8}$$

It is possible to demonstrate that the same solution that would be obtained from Equation (B.8) can be retrieved with the following functional

$$f(\boldsymbol{x}) = \|\boldsymbol{z} - \mathbf{H}\boldsymbol{x}\|_{\mathbf{C}}^2 + \alpha\,\mathcal{P}(\boldsymbol{x}) \tag{B.9}$$

As for Equation (B.2), the pseudo-solution is calculated by minimizing the functional of Equation (B.9), i.e. by taking the derivative and set it equals to zero, which produces the following equation

$$\mathbf{H}^{\mathbf{T}}\mathbf{C}\mathbf{H}\widehat{\boldsymbol{x}} - \mathbf{A}^{\mathbf{T}}\mathbf{C}\widehat{\boldsymbol{z}} + \frac{\alpha}{2}\frac{\partial\mathcal{P}(\widehat{\boldsymbol{x}})}{\partial\widehat{\boldsymbol{x}}} = 0 \tag{B.10}$$

By comparing Equation (B.10) with Equation (B.8) it is possible to determine the calibration matrix $\mathbf{C}$, i.e.

$$\mathbf{C} = \mathbf{I} - \mathbf{B}\left(\mathbf{B}^{\mathbf{T}}\mathbf{B}\right)^{-\mathbf{1}}\mathbf{B}^{\mathbf{T}} \tag{B.11}$$

Appendix C

This appendix shows the reconstructions obtained with the methods described in Chapter 7 in terms of normalized TEC. The figures are compared with the normalized electron density as measured by CHAMP.

C.1 19 November 2009

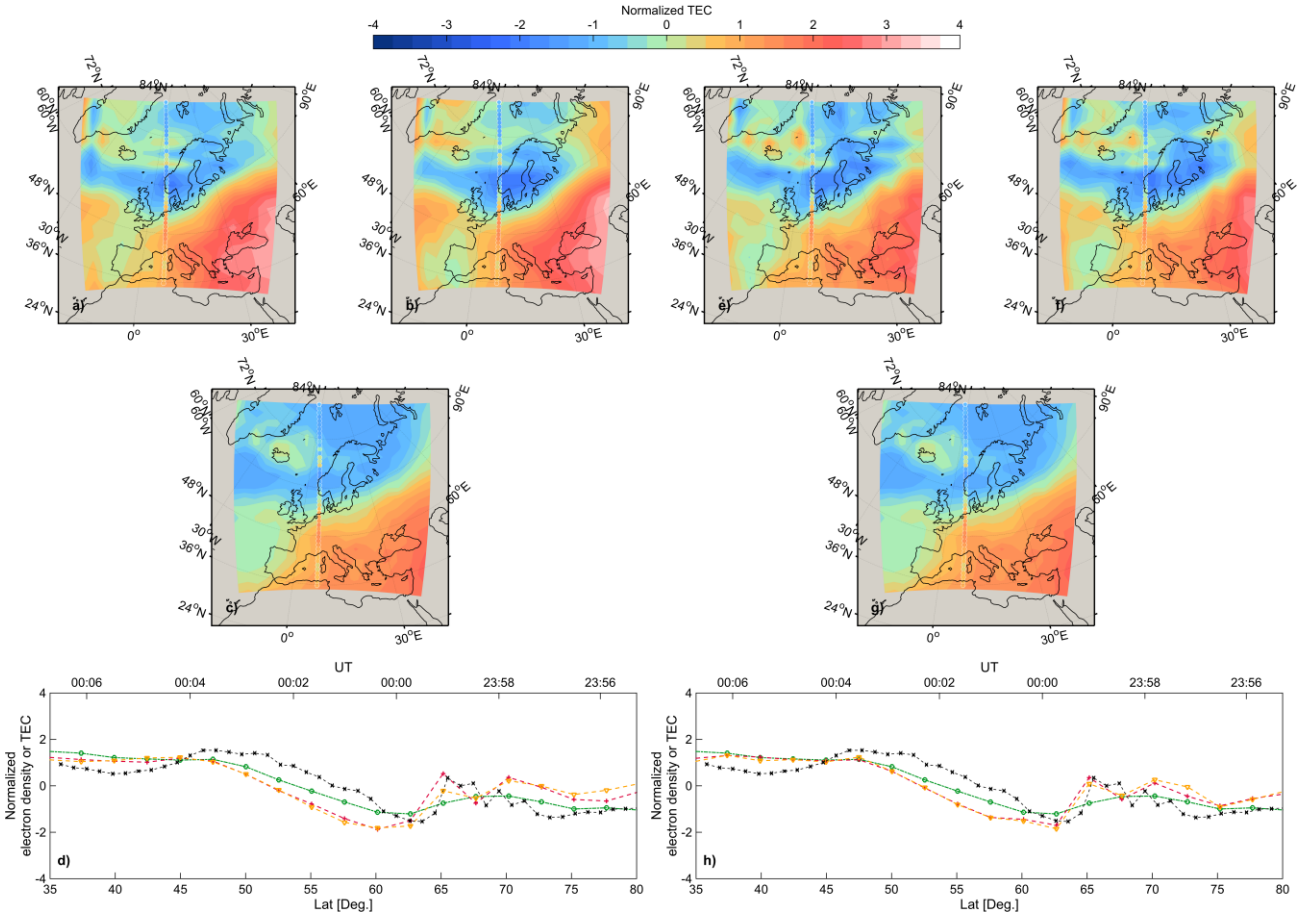

Figure C.1. Comparison between normalized CHAMP and normalized TEC maps by using a) Daubechies 4; b) discrete Meyer; c) MIDAS with the time-smooth algorithm, and d) Daubechies 4; e) discrete Meyer; f) MIDAS with the time-smooth algorithm. CHAMP is shown in terms of normalized electro density as coloured circles over the reconstruction maps. The colour is proportional to the normalized electron density. The bottom plot shows, instead, the normalized electron density (black) and normalized TEC for Daubechies 4 (red) and discrete Meyer (orange) and MIDAS (green) along the satellite CHAMP pass. The plot shows two axes: the latitude (degrees) and the CHAMP time when the measurement was taken. TEC maps are calculated at 00:00 UT of 19 November 2009.

C.2 05 December 2009

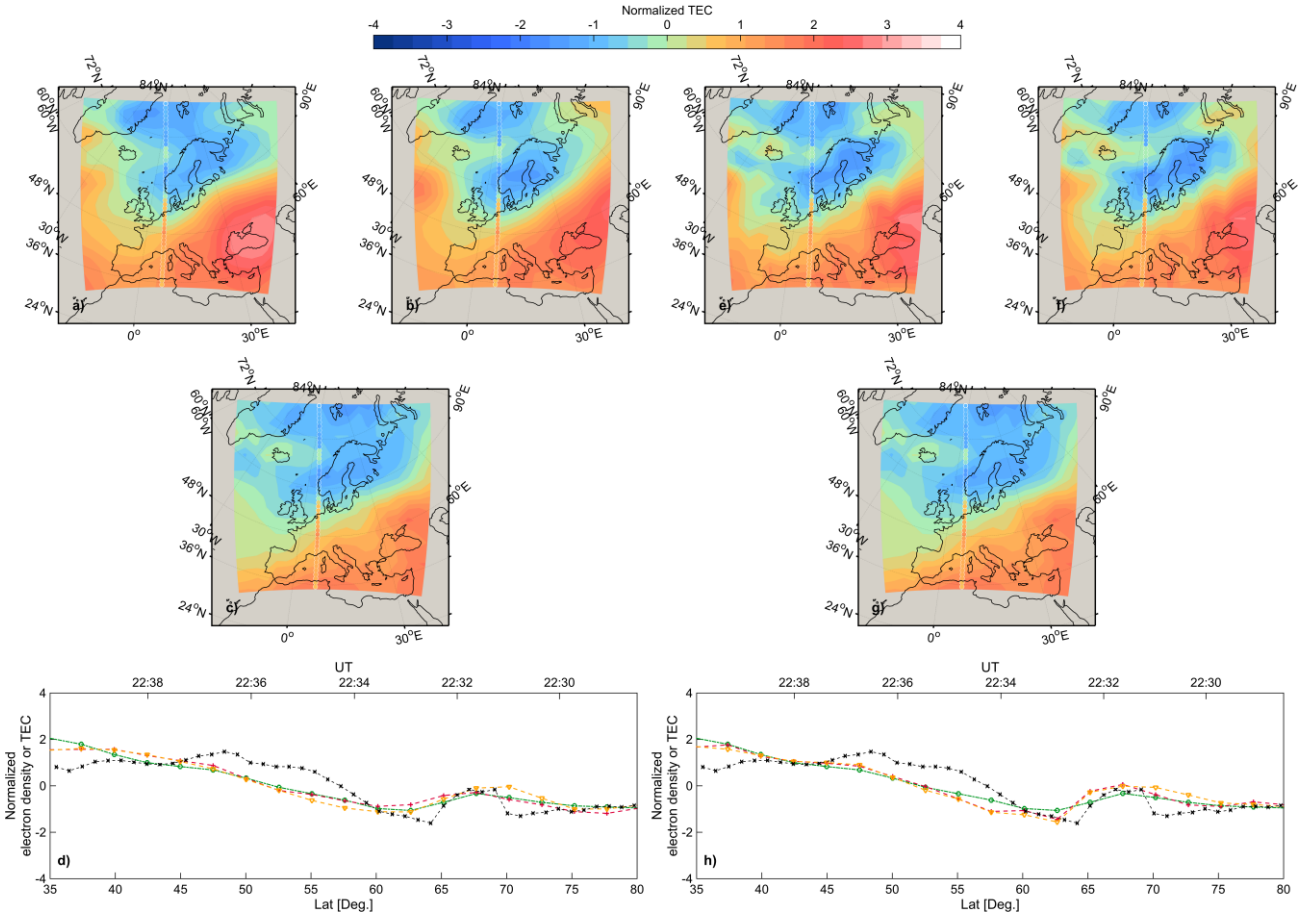

Figure C.2. Comparison between normalized CHAMP and normalized TEC maps by using a) Daubechies 4; b) discrete Meyer; c) MIDAS with the time-smooth algorithm, and d) Daubechies 4; e) discrete Meyer; f) MIDAS with the time-smooth algorithm. CHAMP is shown in terms of normalized electro density as coloured circles over the reconstruction maps. The colour is proportional to the normalized electron density. The bottom plot shows, instead, the normalized electron density (black) and normalized TEC for Daubechies 4 (red) and discrete Meyer (orange) and MIDAS (green) along the satellite CHAMP pass. The plot shows two axes: the latitude (degrees) and the CHAMP time when the measurement was taken. TEC maps are calculated at 22:30 UT of 05 December 2009.

C.3 14 July 2009

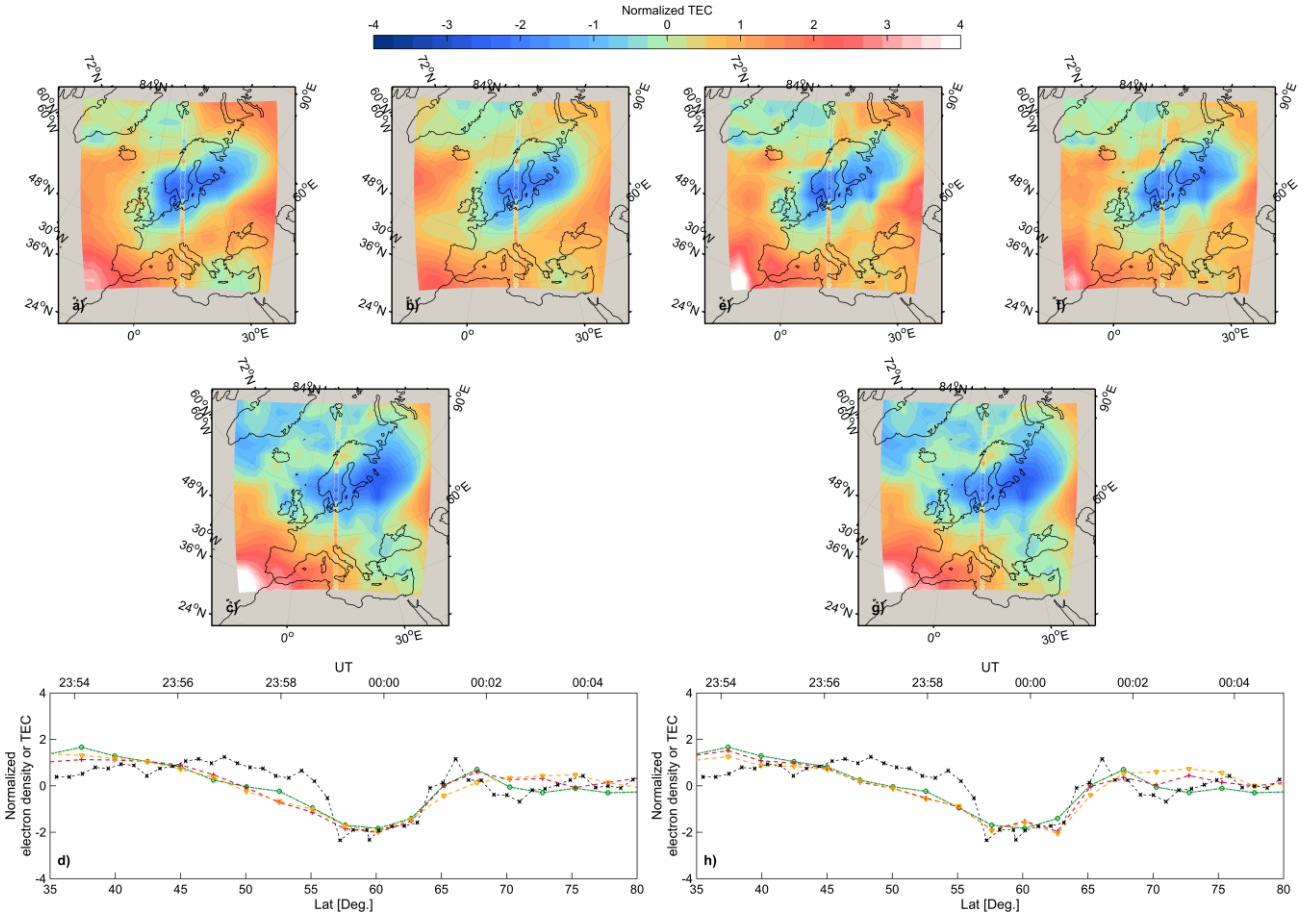

Figure C.3. Comparison between normalized CHAMP and normalized TEC maps by using a) Daubechies 4; b) discrete Meyer; c) MIDAS with the time-smooth algorithm, and d) Daubechies 4; e) discrete Meyer; f) MIDAS with the time-smooth algorithm. CHAMP is shown in terms of normalized electro density as coloured circles over the reconstruction maps. The colour is proportional to the normalized electron density. The bottom plot shows, instead, the normalized electron density (black) and normalized TEC for Daubechies 4 (red) and discrete Meyer (orange) and MIDAS (green) along the satellite CHAMP pass. The plot shows two axes: the latitude (degrees) and the CHAMP time when the measurement was taken. TEC maps are calculated at 00:00 UT of 15 July 2009.

C.4 23 July 2009

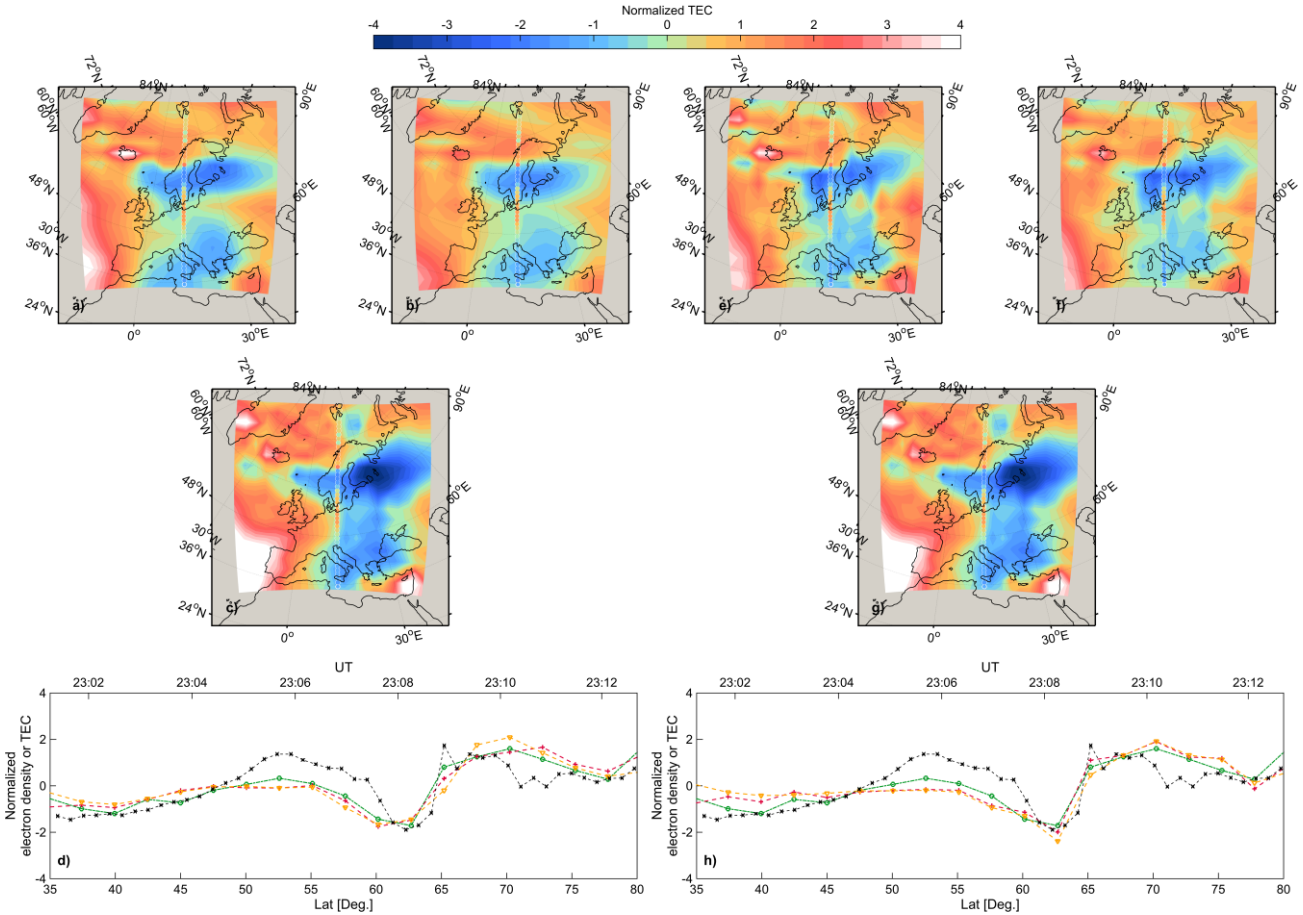

Figure C.4. Comparison between normalized CHAMP and normalized TEC maps by using a) Daubechies 4; b) discrete Meyer; c) MIDAS with the time-smooth algorithm, and d) Daubechies 4; e) discrete Meyer; f) MIDAS with the time-smooth algorithm. CHAMP is shown in terms of normalized electro density as coloured circles over the reconstruction maps. The colour is proportional to the normalized electron density. The bottom plot shows, instead, the normalized electron density (black) and normalized TEC for Daubechies 4 (red) and discrete Meyer (orange) and MIDAS (green) along the satellite CHAMP pass. The plot shows two axes: the latitude (degrees) and the CHAMP time when the measurement was taken. TEC maps are calculated at 23:10 UT of 23 July 2009.

Appendix D

This appendix shows the tables obtained for the case studies presented in Chapter 7.

D.1 Normalized TEC compared with CHAMP

The following table shows the RMS error, the C_{MAX} and MSSIS indices calculated from the reconstructed normalized TEC and CHAMP profiles. The percentage number of basis functions used for the reconstructions is also indicated.

Table D-1. The table shows different indices calculated from normalized VTEC maps obtained with MIDAS and wavelets using CHAMP as reference. The indices are the RMS error (values are in TECU), the maximum cross-correlation values (CMAX, normalized to 100) between the reconstructed values and CHAMP and the MSSIS index. The percentage of basis functions with non-zero coefficients is also shown. The time-smooth (TS) and time-sparse (TP) results are illustrated for each basis function and for different case studies.

		Type	Basis Functions	RMS error	C_{MAX}	MSSIS	Number of basis (%)
TEC	20-Nov-09		No	0.47	80.27	0.50	100.00
		TS	DB4	0.72	71.09	0.65	37.08
		TP		0.76	66.42	0.63	24.33
		TS	DM	0.88	62.21	0.45	31.08
		TP		0.82	63.71	0.60	12.66
	05-Dec-09		No	0.53	63.00	0.44	100.00
		TS	DB4	0.64	59.51	0.48	40.48
		TP		0.60	63.24	0.42	22.88
		TS	DM	0.66	58.62	0.45	32.22
		TP		0.68	60.50	0.34	11.93
	14-Jul-09		No	0.56	73.71	0.51	100.00
		TS	DB4	0.61	81.26	0.56	33.39
		TP		0.66	73.64	0.50	18.25
		TS	DM	0.69	74.76	0.47	27.72
		TP		0.77	71.74	0.31	10.89

		Type	Basis Functions	RMS error	C_{MAX}	MSSIS	Number of basis (%)
TEC	23-Jul-09		No	0.70	70.09	0.62	100.00
		TS	DB4	0.83	62.91	0.49	34.78
		TP		0.89	63.30	0.59	18.43
		TS	DM	0.90	59.37	0.44	28.52
		TP		0.91	62.90	0.51	11.47

D.2 Normalized TEC gradients compared with CHAMP

The following table summarizes the results for the normalized TEC and electron density gradients in terms of RMS error, correlation index (normalized to 100), MSSIS index and percentage of basis functions.

Table D-2. The table shows different indices calculated from normalized VTEC gradient maps obtained with MIDAS and wavelets using CHAMP as reference. The indices are the RMS error (values are in TECU), the maximum cross-correlation values (CMAX, normalized to 100) between the reconstructed values and CHAMP and the MSSIS index. The percentage of basis functions with non-zero coefficients is also shown. The time-smooth (TS) and time-sparse (TP) results are illustrated for each basis function and for different case studies.

		Type	Basis Functions	RMS error	C_{MAX}	MSSIS	Number of basis (%)
DTEC	20-Nov-09		No	0.36	2.64	0.37	100.00
		TS	DB4	0.39	5.26	0.39	37.08
		TP		0.40	5.63	0.41	24.33
		TS	DM	0.40	6.23	0.41	31.08
		TP		0.39	5.43	0.30	12.66
	05-Dec-09		No	0.24	1.90	0.54	100.00
		TS	DB4	0.22	3.56	0.44	40.48
		TP		0.26	2.80	0.26	22.88
		TS	DM	0.23	3.13	0.33	32.22
		TP		0.28	2.33	0.15	11.93

		Type	Basis Functions	RMS error	C_{MAX}	MSSIS	Number of basis (%)
	14-Jul-09		**No**	0.46	6.46	0.26	100.00
		TS	**DB4**	0.43	9.98	0.38	33.39
		TP		0.43	8.94	0.38	18.25
		TS	**DM**	0.45	8.36	0.36	27.72
		TP		0.46	7.49	0.38	10.89
	23-Jul-09		**No**	0.53	14.55	0.37	100.00
		TS	**DB4**	0.59	16.86	0.23	34.78
		TP		0.53	16.95	0.32	18.43
		TS	**DM**	0.57	12.69	0.21	28.52
		TP		0.52	15.15	0.31	11.47

References

Adcock, B., A. Hansen, C. Poon, and B. Roman (2013), Breaking the coherence barrier: asymptotic incoherence and asymptotic sparsity in compressed sensing, *Preprint*.

Adler, A., T. Dai, and W. R. B. Lionheart (2007), Temporal image reconstruction in electrical impedance tomography, *Physiological measurement*, *28*(7), S1.

Afraimovich, E., O. Pirog, and A. Terekhov (1992), Diagnostics of large-scale structures of the high-latitude ionosphere based on tomographic treatment of navigation-satellite signals and of data from ionospheric stations, *Journal of atmospheric and terrestrial physics*, *54*(10), 1265-1273.

Amerian, Y., M. H. Hossainali, B. Voosoghi, and M. R. Ghaffari (2010), Tomographic Reconstruction of the Ionospheric Electron Density in term of Wavelets, *Journal of Aerospace Science and Technology*, *7*(1), 19-29.

Andreeva, E., A. Galinov, V. Kunitsyn, Y. A. Mel'nichenko, E. Tereshchenko, M. Filimonov, and S. Chernyakov (1990), Radiotomographic reconstruction of ionization dip in the plasma near the Earth, *Letters to Jounal of Experimental and Theoretical Physics*, *52*(3), 145-148.

Austen, J. R., S. J. Franke, and C. H. Liu (1988), Ionospheric imaging using computerized tomography, *Radio Science*, *23*, 299-307.

Austen, J. R., S. J. Franke, C. H. Liu, and K. C. Yeh (1986), Application of Computerized Tomography Techniques to Ionospheric Research, paper presented at Proceedings of the Beacon Satellite Symposium 1986, University of Oulu, Finland.

Barclay, L. (2003), *Propagation of Radiowaves, 2nd Edition*, The Institution of Electrical Engineers.

Barrett, R., M. Berry, T. Chan, J. Demmel, J. Donato, J. Dongarra, V. Eijkhout, R. Pozo, C. Romine, and H. van der Vorst (1994), *Templates for the Solution of Linear Systems: Building Blocks for Iterative Methods*, 135 pp., Society for Industrial and Applied Mathematics.

Beck, A., and M. Teboulle (2009), A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems, *SIAM Journal on Imaging Sciences*, *2*(1), 183-202.

Beck, A., and M. Teboulle (2009), Fast Gradient-Based Algorithms for Constrained Total Variation Image Denoising and Deblurring Problems, *Image Processing, IEEE Transactions on*, *18*(11), 2419-2434.

Beynon, W. J. G., and P. J. S. Williams (1978), Incoherent scatter of radio waves from the ionosphere, *Reports on Progress in Physics*, *41*(6), 909-955.

Bilitza, D., D. Altadill, Y. Zhang, C. Mertens, V. Truhlik, P. Richards, L.-A. McKinnell, and B. Reinisch (2014), The International Reference Ionosphere 2012 – a model of international collaboration, *Journal of Space Weather and Space Climate*, *4*, A07.

Bowles, K. L. (1961), Incoherent scattering by free electrons as a technique for studying the ionosphere and exosphere: Some observations and theoretical considerations, *Journal of Research of National Bureau of Standards*, *65D*(1), 1-14.

Bracewell, R. N. (1956), Strip Integration in Radio Astronomy, *Australian Journal of Physics*, *9*(2), 198-217.

Bruckstein, A. M., D. L. Donoho, and M. Elad (2009), From Sparse Solutions of Systems of Equations to Sparse Modeling of Signals and Images, *SIAM Review*, *51*(1), 34-81.

Bust, G., D. Coco, and J. J. Makela (2000), Combined Ionospheric Campaign 1: Ionospheric tomography and GPS total electron count (TEC) depletions, *Geophysical research letters*, *27*(18), 2849-2852.

Bust, G., T. W. Garner, and T. L. Gaussiran II (2004), Ionospheric Data Assimilation Three-Dimensional (IDA3D): A global, multisensor, electron density specification algorithm, *Journal of Geophysical Research*, *109*(A11), A11312.

Bust, G., G. Crowley, N. Curtis, A. Reynolds, L. Paxton, C. Coker, and P. Bernhardt (2007a), IDA4D-a new ionospheric imaging algorithm using non-linear ground-based and spaced-based data sources, American Geophysical Union, Fall Meeting 2007, abstract #SA11B-06.

Bust, G., G. Crowley, T. W. Garner, T. L. Gaussiran II, R. W. Meggs, C. N. Mitchell, P. S. J. Spencer, P. Yin, and B. Zapfe (2007b), Four-dimensional GPS imaging of space weather storms, *Space Weather*, *5*(2), S02003.

Bust, G. S., and C. N. Mitchell (2008), History, current state, and future directions of ionospheric imaging, *Reviews of Geophysics*, *46*(1).

Censor, Y. (1983), Finite series-expansion reconstruction methods, *Proceedings of the IEEE*, *71*(3), 409-419.

Chambolle, A. (2004), An Algorithm for Total Variation Minimization and Applications, *Journal of Mathematical Imaging and Vision*, *20*(1-2), 89-97.

Chapman, S. (1931), The absorption and dissociative or ionizing effect of monochromatic radiation in an atmosphere on a rotating earth, *Proceedings of the Physical Society*, *43*(1), 26-45.

Chartier, A. T., C. N. Mitchell, and D. R. Jackson (2012), A 12 year comparison of MIDAS and IRI 2007 ionospheric Total Electron Content, *Advances in Space Research*, *49*(9), 1348-1355.

Choi, S.-C. T., and M. A. Saunders (2014), Algorithm 937: MINRES-QLP for symmetric and Hermitian linear equations and least-squares problems, *ACM Transactions on Mathematical Software*, *40*(2), 1-12.

Cooke, D. L., C. W. Turnbull, C. Roth, A. Morgan, and R. Redus (2003), Ion Drift-Meter Status and Calibration, in *First CHAMP Mission Results for Gravity, Magnetic and Atmospheric Studies*, edited by C. Reigber, H. Lühr and P. Schwintzer, pp. 212-219, Springer Berlin Heidelberg.

Da Dalt, F. (2015), Ionospheric modelling and data assimilation, Thesis (Ph.D.) - University of Bath, Bath.

Daley, R. (1991), *Atmospheric Data Analysis*, Cambridge University Press, Cambridge.

Daley, R., and E. Barker (1998), The NRL 3DVAR source book, *NRL internal report Naval Research Laboratory, Monterey, Ca.*, 146.

Danchik, R. J. (1998), An overview of transit development, *Johns Hopkins APL technical digest*, *19*(1), 19.

Daniell, R. E. (1991), Parameterized Real-Time Ionospheric Specification Model PRISM Version 1. 0, *Rep. PL-TR-91-2299*, Phillips Lab., Hanscom Air Force Base, Mass.

Daubechies, I. (1992), *Ten lectures on wavelets*, SIAM.

Daubechies, I., M. Defrise, and C. De Mol (2004), An iterative thresholding algorithm for linear inverse problems with a sparsity constraint, *Communications on Pure and Applied Mathematics*, *57*(11), 1413-1457.

Davies, K. (1990), *Ionospheric Radio*, The Institution of Engineering and Technology, London.

Dear, R. M., and C. N. Mitchell (2006), GPS interfrequency biases and total electron content errors in ionospheric imaging over Europe, *Radio Science*, *41*(6), RS6007.

Donoho, D. L. (1992), Wavelet shrinkage and WVD: A 10-minute tour, paper presented at the International Conference on Wavelets and Applications, Tolouse, France, Citeseer.

Donoho, D. L. (2006), For most large underdetermined systems of linear equations the minimal ℓ1-norm solution is also the sparsest solution, *Communications on Pure and Applied Mathematics*, *59*(6), 797-829.

Donoho, D. L., and J. M. Johnstone (1994a), Ideal spatial adaptation by wavelet shrinkage, *Biometrika*, *81*(3), 425-455.

Donoho, D. L., and J. M. Johnstone (1994b), Threshold selection for wavelet shrinkage of noisy data, paper presented at Engineering in Medicine and Biology Society, 1994. Engineering Advances: New Opportunities for Biomedical Engineers. Proceedings of the 16th Annual International Conference of the IEEE, IEEE, Baltimore, MD, 3-6 Nov 1994.

Donoho, D. L., M. Elad, and V. N. Temlyakov (2006), Stable recovery of sparse overcomplete representations in the presence of noise, *Information Theory, IEEE Transactions on*, *52*(1), 6-18.

Durmaz, M., and M. O. Karslioğlu (2011), Non-parametric regional VTEC modeling with Multivariate Adaptive Regression B-Splines, *Advances in Space Research*, *48*(9), 1523-1530.

Erturk, O., O. Arikan, and F. Arikan (2009), Tomographic reconstruction of the ionospheric electron density as a function of space and time, *Advances in Space Research*, *43*(11), 1702-1710.

Fodor, I. K., and C. Kamath (2003), Denoising through wavelet shrinkage: an empirical study, *Journal of Electronic Imaging*, *12*(1), 151-160.

Foster, J. C., et al. (1994), Russian-American tomography experiment, *International Journal of Imaging Systems and Technology*, *5*(2), 148-159.

Fougere, P. F. (1995), Ionospheric radio tomography using maximum entropy 1. Theory and simulation studies, *Radio Science*, *30*(2), 429-444.

Frank, J. (2006), *Electron tomography: methods for three-dimensional visualization of structures in the cell*, Springer.

Fremouw, E. J., J. A. Secan, and B. M. Howe (1992), Application of stochastic inverse theory to ionospheric tomography, *Radio Science*, *27*(5), 721-732.

Gorbunov, M., W. Schreiner, D. Feng, B. Herman, Y. Kuo, and X. Zou (1997), Analysis and validation of GPS/MET data in the neutral atmosphere, *Journal of Geophysical Research*, *102*(D25), 29,849-829,866.

Gordon, R., R. Bender, and G. T. Herman (1970), Algebraic Reconstruction Techniques (ART) for three-dimensional electron microscopy and X-ray photography, *Journal of Theoretical Biology*, *29*(3), 471-481.

Gordon, W. E. (1958), Incoherent Scattering of Radio Waves by Free Electrons with Applications to Space Exploration by Radar, *Proceedings of the IRE*, *46*(11), 1824-1829.

Guier, W. H., and G. C. Weiffenbach (1997), Genesis of satellite navigation, *Johns Hopkins APL technical digest*, *18*(2), 178-181.

Hajj, G., B. Wilson, C. Wang, X. Pi, and I. Rosen (2004), Data assimilation of ground GPS total electron content into a physics-based ionospheric model by use of the Kalman filter, *Radio Sci*, *39*(1).

Hajj, G. A., R. Ibañez-Meier, E. R. Kursinski, and L. J. Romans (1994), Imaging the ionosphere with the Global Positioning System, *International Journal of Imaging Systems and Technology*, *5*(2), 174-187.

Hansen, P. (1987), The truncated SVD as a method for regularization, *BIT Numerical Mathematics*, *27*(4), 534-553.

Hargreaves, J. K. (1995), *The Solar-Terrestrial Environment: An Introduction to Geospace - The Science of the Terrestrial Upper Atmosphere, Ionosphere, and Magnetosphere*, Cambridge University Press, New York.

Hounsfield, G. N. (1975), Method of and apparatus for examining a body by radiation such as x or gamma radiation, in *Patent File Date: 1968 Aug 23; Other Information: G01t1/00. Orig. Receipt Date: 30-JUN-76*, edited, p. Medium: X; Size: Pages: 14.

Howe, B. M., K. Runciman, and J. A. Secan (1998), Tomography of the ionosphere: Four-dimensional simulations, *Radio Science*, *33*(1), 109-128.

Huang, T. S., and P. Narendra (1975), Image restoration by singular value decomposition, *Applied Optics*, *14*(9), 2213-2216.

Jakowski, N. (1996), TEC monitoring by using satellite positioning systems, in *Modern Ionospheric Science – 50 Years of Ionospheric Research in Lindau*, edited by H. Kohl, R. Rüster and K. Schlegel, pp. 371-389, Produserv GmbH Verlagsservice, Germany.

Jakowski, N., S. Heise, A. Wehrenpfennig, and S. Schlüter (2001), TEC monitoring by GPS-a possible contribution to space weather monitoring, *Physics and Chemistry of the Earth, Part C: Solar, Terrestrial & Planetary Science*, *26*(8), 609-613.

Jones, D. G., I. K. Walker, and L. Kersley (1997), Structure of the poleward wall of the trough and the inclination of the geomagnetic field above the EISCAT radar, *Annales Geophysicae*, *15*(6), 740-746.

Kaplan, E. D., and C. J. Hegarty (2006), *Understanding GPS: principles and applications*, Artech House Publishers.

Keinert, F. (2003), *Wavelets and multiwavelets*, Chapman & Hall/CRC.

Kersley, L., J. A. T. Heaton, S. E. Pryse, and T. D. Raymund (1993), Experimental ionospheric tomography with ionosonde input and EISCAT verification, paper presented at Annales Geophysicae, Copernicus.

Kuklinski, W. S. (1997), Ionospheric tomography via iterative cross-entropy minimization, *Radio Science*, *32*(3), 1037-1049.

Kunitake, M., K. Ohtaka, T. Maruyama, M. Tokumaru, A. Morioka, and S. Watanabe (1995), Tomographic imaging of the ionosphere over Japan by the modified truncated SVD method, *Annales of Geophysics*, *13*, 1303-1310.

Kunitsyn, V. E., E. S. Andreeva, O. G. Razinkov, and E. D. Tereshchenko (1994a), Phase and phase-difference ionospheric radio tomography, *International Journal of Imaging Systems and Technology*, *5*(2), 128-140.

Kunitsyn, V. E., E. S. Andreeva, E. D. Tereshchenko, B. Z. Khudukon, and T. Nygrén (1994b), Investigations of the ionosphere by satellite radiotomography, *International Journal of Imaging Systems and Technology*, *5*(2), 112-127.

Leadbetter, M. R., G. Lindgren, and H. Rootzen (1983), *Extremes and Related Properties of Random Sequences and Processes*, Springer-Verlag, New York.

Leitinger, R. (1996), Tomography, *Modern Ionospheric Science – 50 Years of Ionospheric Research in Lindau*, 346-370.

Leitinger, R., G. Schmidt, and A. Tauriainen (1975), An evaluation method combining the differential Doppler measurements from two stations that enables the calculation of the electron content of the ionosphere, *Zeitschrift fur Geophysik*, *41*(2), 201-213.

Loris, I., F. Simons, I. Daubechies, G. Nolet, M. Fornasier, P. Vetter, S. Judd, S. Voronin, C. Vonesh, and J. Charléty (2010), Solving or resolving global tomographic models with spherical wavelets, and the scale and sparsity of seismic heterogeneity, paper presented at AGU Fall Meeting Abstracts.

Mallat, S. (2008), *A wavelet tour of signal processing: the sparse way*, Academic press.

Mandrake, L., B. Wilson, C. Wang, G. Hajj, A. Mannucci, and X. Pi (2005), A performance evaluation of the operational Jet Propulsion Laboratory/University of Southern California global assimilation ionospheric model (JPL/USC GAIM), *Journal of Geophysical Research*, *110*(A12), A12306.

Mannucci, A. J., B. A. Iijima, U. J. Lindqwister, X. Pi, L. Sparks, and B. D. Wilson (1999), *GPS and ionosphere*, W. Ross Stone (Editor) ed., 625-665 pp., Wiley-IEEE Press, New York.

Markkanen, M., M. Lehtinen, T. Nygren, J. PirttilaÈ, P. Henelius, E. Vilenius, E. D. Tereshchenko, and B. Z. Khudukon (1995), Bayesian approach to satellite radiotomography with applications in the Scandinavian sector, paper presented at Annales Geophysicae, Copernicus.

Meggs, R. W., C. N. Mitchell, and V. S. C. Howells (2005), Simultaneous observations of the main trough using GPS imaging and the EISCAT radar, *Annales Geophysicae*, *23*(3), 753-757.

Menke, W. (1989), *Geophysical data analysis: Discrete inverse theory*, Academic press, San Diego, California.

Mitchell, C. N., and P. S. J. Spencer (2003), A three-dimensional time-dependent algorithm for ionospheric imaging using GPS, *Annals of Geophysics*, *46*(4).

Mitchell, C. N., D. G. Jones, L. Kersley, S. E. Pryse, and I. K. Walker (1995), Imaging of field-aligned structures in the auroral ionosphere, paper presented at Annales geophysicae, Copernicus.

Mitchell, C. N., L. Kersley, S. E. Pryse, C. A. Willson, J. A. T. Heaton, P. S. Cannon, and N. C. Rogers (1999), Imaging and modelling of the main ionospheric trough using radio tomography, paper presented at Antennas and Propagation, 1999. IEE National Conference on., April 1 1999-March 31 1999.

Moffett, R. J., and S. Quegan (1983), The mid-latitude trough in the electron concentration of the ionospheric F-layer: a review of observations and modelling, *Journal of atmospheric and terrestrial physics*, *45*(5), 315-343.

Munk, W., and C. Wunsch (1979), Ocean acoustic tomography: A scheme for large scale monitoring, *Deep Sea Research Part A. Oceanographic Research Papers*, *26*(2), 123-161.

Na, H., and H. Lee (1991), Orthogonal decomposition technique for ionospheric tomography, *International Journal of Imaging Systems and Technology*, *3*(4), 354-365.

Na, H., and H. Lee (1992), Analysis of fundamental resolution limit of ionospheric tomography, paper presented at Acoustics, Speech, and Signal Processing, 1992. ICASSP-92., 1992 IEEE International Conference on, IEEE, San Francisco, CA, 23-26 Mar 1992.

Na, H., and C. Biswas (1994), A space-frequency algorithm for limited angle tomography, paper presented at Image Processing, 1994. Proceedings. ICIP-94., IEEE International Conference, IEEE.

Na, H. R. (1993), Tomographic imaging of ionospheric distributions, paper presented at Signals, Systems and Computers, 1993. 1993 Conference Record of The Twenty-Seventh Asilomar Conference on, 1-3 Nov 1993.

Natarajan, B. K. (1995), Sparse Approximate Solutions to Linear Systems, *SIAM journal on computing*, *24*(2), 227-234.

Natterer, F. (2001), *The mathematics of computerized tomography*, Society for Industrial Mathematics.

Nesterov, I., and V. Kunitsyn (2011), GNSS radio tomography of the ionosphere: The problem with essentially incomplete data, *Advances in Space Research*, *47*(10), 1789-1803.

Newbold, H. G. (1972), A method of and apparatus for examination of a body by radiation such as x or gamma radiation, Patent 1283915, U.S. Patent and Trademark Off., Washington D.C.

Paige, C. C., and M. A. Saunders (1982), LSQR: An Algorithm for Sparse Linear Equations and Sparse Least Squares, *ACM Transactions on Mathematical Software*, *8*(1), 43-71.

Panicciari, T., N. D. Smith, F. Da Dalt, and C. N. Mitchell (2014a), Robust ionospheric tomography using sparse regularization, paper presented at General Assembly and Scientific Symposium (URSI GASS), 2014 XXXIth URSI, 16-23 Aug. 2014.

Panicciari, T., N. D. Smith, F. Da Dalt, C. N. Mitchell, and G. S. Bust (2014b), Multiresolution Tomography of Ionospheric Electron Density, paper presented at Mitigation of Ionospheric Threats to GNSS: an Appraisal of the Scientific and Technological Outputs of the TRANSMIT Project, 2014-07-17.

Panicciari, T., N. D. Smith, C. N. Mitchell, F. Da Dalt, and P. S. J. Spencer (2015), Using sparse regularization for multiresolution tomography of the ionosphere, *Nonlinear Processes in Geophysics*, *22*(5), 613-624.

Petrie, E., M. Hernández-Pajares, P. Spalla, P. Moore, and M. King (2011), A Review of Higher Order Ionospheric Refraction Effects on Dual Frequency GPS, *Surveys in Geophysics*, *32*(3), 197-253.

Pi, X., C. Wang, G. A. Hajj, G. Rosen, B. D. Wilson, and G. J. Bailey (2003), Estimation of E× B drift using a global assimilative ionospheric model: an observation system simulation experiment, *Journal of Geophysical Research*, *108*(A2), 1075.

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery (2007), *Numerical recipes 3rd edition: The art of scientific computing*, Cambridge university press.

Pryse, S. E., and L. Kersley (1992), A preliminary experimental test of ionospheric tomography, *Journal of atmospheric and terrestrial physics*, *54*(7–8), 1007-1012.

Pryse, S. E., L. Kersley, M. J. Williams, and I. K. Walker (1998), The spatial structure of the dayside ionospheric trough, *Annales Geophysicae*, *16*(10), 1169-1179.

Pryse, S. E., L. Kersley, D. L. Rice, C. D. Russell, and I. K. Walker (1993), Tomographic imaging of the ionospheric mid-latitude trough, paper presented at Annales Geophysicae, Copernicus.

Radon, J. (1917), Über die Bestimmung von Funktionen durch ihre Integralwerte längs gewisser Mannigfaltigkeiten, *Berichte über die Verhandlungen der Sächsischen Akademie der Wissenschaften*, *69*, 262-267.

Raymund, T. D. (1994), Ionospheric tomography algorithms, *International Journal of Imaging Systems and Technology*, *5*(2), 75-85.

Raymund, T. D., S. J. Franke, and K. C. Yeh (1994), Ionospheric tomography: its limitations and reconstruction methods, *Journal of Atmospheric and Terrestrial Physics*, *56*(5), 637-657.

Raymund, T. D., J. Austen, S. Franke, C. Liu, J. Klobuchar, and J. Stalker (1990), Application of computerized tomography to the investigation of ionospheric structures, *Radio Science*, *25*(5), 771-789.

Reigber, C., H. Lühr, and P. Schwintzer (2002), CHAMP mission status, *Advances in Space Research*, *30*(2), 129-134.

Rodger, A. (2008), The Mid-Latitude Trough—Revisited, in *Midlatitude Ionospheric Dynamics and Disturbances*, edited, pp. 25-33, American Geophysical Union, Washington, D. C.

Rubinstein, R., A. M. Bruckstein, and M. Elad (2010), Dictionaries for Sparse Representation Modeling, *Proceedings of the IEEE*, *98*(6), 1045-1057.

Scherliess, L., R. W. Schunk, J. J. Sojka, D. C. Thompson, and L. Zhu (2006), Utah State University global assimilation of ionospheric measurements Gauss-Markov Kalman filter model of the ionosphere: model description and validation, *Journal of Geophysical Research*, *111*, A11315.

Schmidt, M. (2007), Wavelet modelling in support of IRI, *Advances in Space Research*, *39*(5), 932-940.

Schmidt, M., D. Bilitza, C. Shum, and C. Zeilhofer (2008), Regional 4-D modeling of the ionospheric electron density, *Advances in Space Research*, *42*(4), 782-790.

Schunk, R., L. Scherliess, J. J. Sojka, D. C. Thompson, and L. Zhu (2005), Ionospheric weather forecasting on the horizon, *Space Weather*, *3*(8), S08007.

Schunk, R. W., J. J. Sojka, and J. V. Eccles (1997), Expanded capabilities for the ionospheric forecast model, *Rep. AFRL-VS-HA-TR-98-0001*, 1-142 pp, Space Veh. Dir., Hanscom Air Force Base, Mass.

Schunk, R. W., L. Scherliess, J. J. Sojka, D. C. Thompson, and L. Zhu (2005), An Operational Data Assimilation Model of the Ionosphere, paper presented at Ionospheric Effects Symposium, Natl. Tech. Info. Serv., Springfield, Va.

Schunk, R. W., L. Scherliess, J. J. Sojka, D. C. Thompson, D. N. Anderson, M. Codrescu, C. Minter, T. J. Fuller-Rowell, R. A. Heelis, and M. Hairston (2004), Global assimilation of ionospheric measurements (GAIM), *Radio Science*, *39*(1), RS1S02.

Shi, J., J. Wielaard, R. T. Smith, and P. Sajda (2013), Perceptual Decision Making "Through the Eyes" of a Large-scale Neural Model of V1, *Frontiers in psychology*, *4*.

Simons, F. J., I. Loris, G. Nolet, I. C. Daubechies, S. Voronin, J. S. Judd, P. A. Vetter, J. Charléty, and C. Vonesch (2011), Solving or resolving global tomographic models with spherical wavelets, and the scale and sparsity of seismic heterogeneity, *Geophysical journal international*, *187*(2), 969-988.

Spencer, P. S. J., and C. N. Mitchell (2007), Imaging of fast moving electron-density structures in the polar cap, *Annals of Geophysics*, *50*(3), 427-434.

Spencer, P. S. J., and C. N. Mitchell (2011), Imaging of 3-D plasmaspheric electron density using GPS to LEO satellite differential phase observations, *Radio Science*, *46*(3), RS0D04.

Stollnitz, E. J., T. D. DeRose, and D. H. Salesin (1995), Wavelets for computer graphics: a primer. 1, *Computer Graphics and Applications, IEEE*, *15*(3), 76-84.

Sutton, E., and H. Na (1994), High resolution ionospheric tomography through orthogonal decomposition, paper presented at Image Processing, 1994. Proceedings. ICIP-94., IEEE International Conference, Austin, TX, 13-16 Nov 1994.

Thomson, J. J. (1906), Conductivity of electricity through gases, edited, Cambridge University Press.

Tikhonov, A. N., and V. Y. Arsenin (1977), *Solution of ill-posed problems*, 258 pp., Winston & Sons, Washington, D.C.

Tropp, J. A. (2004), Just relax: Convex programming methods for subset selection and sparse approximation, *ICES report*, *404*.

Tsaig, Y., and D. L. Donoho (2006), Breakdown of equivalence between the minimal ℓ1-norm solution and the sparsest solution, *Signal Processing*, *86*(3), 533-548.

Voiculescu, M., I. Virtanen, and T. Nygrén (2006), The F-region trough: seasonal morphology and relation to interplanetary magnetic field, *Annales Geophysicae*, *24*(1), 173-185.

Wang, C., G. Hajj, X. Pi, I. G. Rosen, and B. Wilson (2004), Development of the global assimilative ionospheric model, *Radio Science*, *39*(1), RS1S06.

Watermann, J., G. S. Bust, J. P. Thayer, T. Neubert, and C. Coker (2002), Mapping plasma structures in the high-latitude ionosphere using beacon satellite, incoherent scatter radar and ground-based magnetometer observations, *Annals of Geophysics*, *45*(1), 177-189.

Xiaoqing, P., W. Chunming, G. A. Hajl, and I. G. Rosen (2004), Assimilative modeling of low latitude ionosphere, paper presented at Position Location and Navigation Symposium, 2004. PLANS 2004, 26-29 April 2004.

Yeh, K. C., and T. D. Raymund (1991), Limitations of ionospheric imaging by tomography, *Radio Science*, *26*(6), 1361-1380.

Zeilhofer, C., M. Schmidt, D. Bilitza, and C. Shum (2009), Regional 4-D modeling of the ionospheric electron density from satellite data and IRI, *Advances in Space Research*, *43*(11), 1669-1675.

Zhdanov, M. (2002), *Geophysical Inverse Theory and Regularization Problems*, Elsevier, Eastbourne.

Zhou, W., A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli (2004), Image quality assessment: from error visibility to structural similarity, *Image Processing, IEEE Transactions on*, *13*(4), 600-612.