

# **GNSS and InSAR based water vapor tomography: A Compressive Sensing solution**

Zur Erlangung des akademischen Grades einer  
**Doktor-Ingenieurin**

von der Fakultät  
für Bauingenieur-, Geo- und Umweltwissenschaften  
des Karlsruher Instituts für Technologie

genehmigte  
**Dissertation**

von

M.Sc. Marion Elsa Amelie Heublein  
aus Nürnberg

Tag der mündlichen Prüfung: 22.2.2019

Hauptreferent: Prof. Dr.-Ing. habil. Stefan Hinz  
Institut für Photogrammetrie und Fernerkundung  
Karlsruher Institut für Technologie

Korreferenten: Dr.-Ing. Uwe Ehret  
Institut für Wasser und Gewässerentwicklung – Hydrologie  
Karlsruher Institut für Technologie

Prof. Dr.-Ing. habil. Xiaoxiang Zhu  
Signal Processing in Earth Observation  
Remote Sensing Technology Institute  
German Aerospace Center  
Technical University of Munich

**Karlsruhe 2019**



## Abstract

An accurate knowledge of the three-dimensional (3D) distribution of water vapor in the atmosphere is a key element for weather forecasting and climate research. In addition, a precise determination of water vapor is also required for accurate positioning and deformation monitoring using Global Navigation Satellite Systems (GNSS) and Interferometric Synthetic Aperture Radar (InSAR). Several approaches for 3D tomographic water vapor reconstruction from GNSS-based Slant Wet Delay (SWD) estimates exist. Yet, due to the usually sparsely distributed GNSS sites and due to the limited number of visible GNSS satellites, the tomographic system usually is ill-posed and needs to be regularized, e.g. by means of geometric constraints that risk to over-smooth the tomographic refractivity estimates.

Therefore, this work develops and analyzes a Compressive Sensing (CS) approach for neutrospheric water vapor tomographies benefiting of the sparsity of the refractivity estimates in an appropriate transform domain as a prior for regularization. The CS solution is developed because it does not include any geometric smoothing constraints as applied in common Least Squares (LSQ) approaches and because the sparse CS solution containing only a few non-zero coefficients may be determined, at a constant number of observations, based on less parameters than the corresponding LSQ solution. In addition to the developed CS solution, this work introduces SWDs obtained from both GNSS and InSAR into the tomographic system in order to dispose of a better spatial distribution of the observations. The novelties of this approach are 1) the use of both absolute GNSS and absolute InSAR SWDs for tomography and 2) the solution of the tomographic system by means of Compressive Sensing. In addition, 3) the quality of the CS reconstruction is compared with the quality of common LSQ approaches to water vapor tomography.

The tomographic reconstruction is performed, on the one hand, based on a real data set using GNSS and InSAR SWDs and, on the other hand, based on three different synthetic SWD data sets generated using wet refractivity information from the Weather Research and Forecasting (WRF) model. Thus, the validation of the achieved results focuses, on the one hand, on radiosonde profiles and, on the other hand, on a comparison of the refractivity estimates with the input WRF refractivities. The real data set resp. the first synthetic data set compares the reconstruction quality of the developed CS approach with LSQ approaches to water vapor tomography and investigates in how far the inclusion of InSAR resp. synthetic InSAR SWDs increases the accuracy and precision of the refractivity estimates. The second synthetic data set is designed in order to analyze the general effect of the observing geometry on the quality of the refractivity estimates. The third synthetic data set places a special focus on the sensibility of the tomographic reconstruction to different numbers of GNSS sites, varying voxel discretization, and different orbit constellations.

In case of the real data set, for both the GNSS only solution and a combined GNSS and InSAR solution, the refractivities estimated by means of the LSQ and CS methodologies show a consistent behavior, although the two solution strategies differ. The synthetic data sets show that CS can yield very precise and accurate results, if an appropriate tomographic setting is chosen. The reconstruction quality mainly depends on i) the accuracy of the functional model relating the SWD estimates to the refractivity parameters and to the distances passed by the rays within the voxels, ii) the number of available GNSS sites, iii) the voxel discretization, and iv) the variety of ray directions introduced into the tomographic system.

The sizes of the study areas associated to the real resp. to the synthetic data sets are about  $120 \times 120 \text{ km}^2$  and about  $100 \times 100 \text{ km}^2$ , respectively. In the real data set, a total of eight GNSS sites is available and SWD estimates of GPS and InSAR are introduced. In the synthetic data sets, different numbers of sites are defined and a variety of ray directions is tested.



## Zusammenfassung

Unvollständig oder ungenau erstellte Modelle atmosphärischer Effekte schränken die Qualität geodätischer Welt- raumverfahren wie GNSS (Globale Satelliten-Navigationssysteme) und InSAR (Interferometrisches Radar mit synthetischer Apertur) ein. Gleichzeitig enthalten Zustandsgrößen der Erdatmosphäre, allen voran die drei- dimensionale (3D) Wasserdampf-Verteilung, wertvolle Informationen für Klimaforschung und Wettervorhersa- ge, welche aus GNSS- oder InSAR-Beobachtungen abgeleitet werden können. Es gibt etliche Verfahren zur 3D- Wasserdampf-Rekonstruktion aus GNSS-basierten feuchten Laufzeitverzögerungen. Aufgrund der meist spärlich verteilten GNSS-Stationen und durch die begrenzte Anzahl sichtbarer GNSS-Satelliten, treten in tomographischen Anwendungen in der Regel jedoch schlecht gestellte Probleme auf, die z.B. über geometrische Zusatzbedingungen regularisiert werden, welche oft glättend auf die Wasserdampf-Schätzungen wirken.

Diese Arbeit entwickelt und analysiert daher einen Ansatz, der auf einer Compressive Sensing (CS) Lösung des to- mographischen Modells beruht. Dieser Ansatz nutzt die Spärlichkeit der Wasserdampf-Verteilung in einem geeig- neten Transformationsbereich zur Regularisierung des schlecht gestellten tomographischen Problems und kommt somit ohne glättende geometrische Zusatzbedingungen aus. Eine weitere Motivation für die Nutzung einer spär- lichen Compressive Sensing Lösung besteht darin, dass die Anzahl an zu bestimmenden von Null verschiede- nen Koeffizienten bei gleichbleibender Anzahl an Beobachtungen in Compressive Sensing geringer sein kann als die Anzahl an zu schätzenden Parametern in üblichen Kleinste Quadrate (LSQ) Ansätzen. Zur Erhöhung der räumlichen Auflösung der Beobachtungen führt diese Arbeit zudem sowohl feuchte Laufzeitverzögerungen aus GNSS als auch aus InSAR in das tomographische Gleichungssystem ein. Die Neuheiten des vorgestellten Ansatz- es sind 1) die Nutzung von sowohl GNSS als auch absoluten InSAR Laufzeitverzögerungen für die tomographi- sche Wasserdampf-Rekonstruktion und 2) die Lösung des tomographischen Systems mittels Compressive Sensing. Zudem wird 3) die Qualität der CS-Rekonstruktion mit der Qualität üblicher LSQ-Schätzungen verglichen.

Die tomographische Rekonstruktion der durch feuchte Refraktivitäten beschriebenen atmosphärischen Wasserdampf-Verteilung beruht auf der einen Seite auf realen feuchten Laufzeitverzögerungen aus GNSS und InSAR und auf der anderen Seite auf drei verschiedenen synthetischen Datensätzen feuchter Laufzeitverzöge- rungen, die aus Wasserdampf-Simulationen des Weather Research and Forecasting (WRF) Modells abgeleitet wurden. Die Validierung der geschätzten Wasserdampf-Verteilung stützt sich somit zum einen auf Radiosonden- Profile und zum anderen auf einen Vergleich der geschätzten Refraktivitäten mit den WRF Refraktivitäten, die zugleich als Eingangsdaten zur Generierung der synthetischen Laufzeitverzögerungen genutzt werden. Der reale bzw. der erste synthetische Datensatz vergleicht die Rekonstruktionsqualität des entwickelten CS-Ansatzes mit üblichen Kleinste Quadrate Wasserdampf-Schätzungen und untersucht, inwieweit die Nutzung von InSAR Lauf- zeitverzögerungen bzw. von synthetischen InSAR Laufzeitverzögerungen die Genauigkeit und die Präzision der Wasserdampf-Rekonstruktion erhöht. Der zweite synthetische Datensatz wurde dafür ausgelegt, den allgemeinen Einfluss der Beobachtungsgeometrie auf die Refraktivitätsschätzungen zu analysieren. Der dritte synthetische Da- tensatz untersucht insbesondere die Empfindlichkeit der tomographischen Rekonstruktion gegenüber variierenden GNSS-Stationenzahlen, unterschiedlichen Voxel-Diskretisierungen und verschiedenen Orbit-Konstellationen.

Im realen Datensatz verhalten sich die Kleinste Quadrate Schätzung und der Compressive Sensing Ansatz sowohl für die reine GNSS-Lösung als auch für die kombinierte GNSS- und InSAR-Lösung konsistent. Die synthetischen Datensätze zeigen, dass Compressive Sensing in geeigneten Szenarien sehr genaue und präzise Ergebnisse liefern kann. Die Qualität der Wasserdampf-Schätzungen hängt in erster Linie ab i) von der Genauigkeit des funktionalen Modells, das die feuchten Laufzeitverzögerungen, die zu schätzenden Refraktivitäten und die von den Strahlen in den Voxeln zurückgelegten Distanzen in Beziehung zueinander setzt, ii) von der Anzahl verfügbarer GNSS- Stationen, iii) von der Voxel-Diskretisierung, und iv) von der Vielseitigkeit der in das tomographische System eingebauten Strahlrichtungen.

Die mittels des realen Datensatzes bzw. mittels der synthetischen Datensätze untersuchten Regionen sind etwa  $120 \times 120 \text{ km}^2$  bzw.  $100 \times 100 \text{ km}^2$  groß. Im realen Datensatz stehen acht GNSS-Stationen zur Verfügung und es werden feuchte Laufzeitverzögerungen von GPS InSAR genutzt. In den synthetischen Datensätzen werden unterschiedliche Stationsanzahlen definiert und vielseitige Strahlrichtungen getestet.



# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Relevance of determining the 3D atmospheric water vapor distribution . . . . .	1
1.2	Relevance of the innovations in the proposed approach . . . . .	2
1.3	Objectives . . . . .	3
1.4	Contributions . . . . .	4
1.5	Outline . . . . .	4
<b>2</b>	<b>The Earth’s atmosphere: physical foundations and measurement techniques</b>	<b>7</b>
2.1	The Earth’s atmosphere . . . . .	8
2.1.1	Ionosphere . . . . .	8
2.1.2	Neutrosphere . . . . .	9
2.2	Meteorological quantities describing the neutrospheric water vapor distribution . . . . .	9
2.3	Numerical weather models . . . . .	12
2.4	Measurement techniques quantifying the neutrospheric water vapor content . . . . .	13
2.5	Interactions of radio waves and the neutrosphere . . . . .	16
2.6	Path delay modeling in GNSS and InSAR . . . . .	17
2.6.1	Modeling the hydrostatic component of the delay using surface meteorology . . . . .	17
2.6.2	Determining the wet component of the delay content using GNSS PPP . . . . .	19
2.6.3	Determining the wet component of the delay using PS-InSAR . . . . .	20
2.7	Summary . . . . .	22
<b>3</b>	<b>State of the art in tomographic water vapor reconstruction</b>	<b>23</b>
<b>4</b>	<b>Tomographic model</b>	<b>29</b>
4.1	Tomographic equation within a discretized refractivity field . . . . .	30
4.2	Side rays for better vertical resolution . . . . .	31
4.3	Effects of the voxel discretization . . . . .	32
4.4	Raytracing on the ellipsoid . . . . .	35
<b>5</b>	<b>Methodology for 3D tomographic reconstruction</b>	<b>41</b>
5.1	Mathematical basics for the solution of inverse problems . . . . .	41
5.2	$L_2$ solution vs. $L_1$ solution to an inverse problem . . . . .	44
5.3	Constrained Least Squares solution . . . . .	46
5.4	Sparse Compressive Sensing solution . . . . .	50
5.4.1	Sparse dictionaries using Kronecker bases . . . . .	51
5.4.2	Application of CS to the tomographic reconstruction of atmospheric water vapor . . . . .	53
5.5	Summary . . . . .	57
<b>6</b>	<b>Study areas and data sets</b>	<b>59</b>
6.1	Study areas . . . . .	60
6.2	Real data set . . . . .	61
6.2.1	GNSS data availability . . . . .	61
6.2.2	InSAR: data availability, PSI processing, and absolute InSAR ZWD generation . . . . .	62
6.2.3	Surface meteorology . . . . .	63

6.2.4	Radiosonde data . . . . .	63
6.2.5	Real SWD data set based on GNSS and InSAR . . . . .	65
6.3	Synthetic data set . . . . .	68
6.3.1	WRF data set . . . . .	68
6.3.2	Synthetic SWD data set based on WRF . . . . .	69
6.4	Geodetic and meteorological height systems . . . . .	70
6.5	Summary . . . . .	73
<b>7</b>	<b>Tomography results</b>	<b>75</b>
7.1	Real data set . . . . .	77
7.1.1	Tomographic settings in the real data set . . . . .	77
7.1.2	Consistency of GNSS, InSAR, and radiosondes . . . . .	79
7.1.3	Validation of GNSS and InSAR based wet refractivities from CS and LSQ using radiosonde profiles . . . . .	80
7.2	Synthetic data set . . . . .	82
7.2.1	Synthetic data set comparable to real data set . . . . .	82
7.2.2	Effect of the observing geometry on the tomographic results within a general study area . . . . .	86
7.2.3	Effect of the orbits and of the voxel discretization in a specific study area . . . . .	99
7.2.4	Summary . . . . .	107
<b>8</b>	<b>Discussion and Outlook</b>	<b>111</b>
<b>A</b>	<b>Appendices</b>	<b>117</b>
A.1	Niell mapping function . . . . .	117
A.2	PPP processing in Bernese GPS Software 5.2 . . . . .	117



# 1 Introduction

An accurate knowledge of the three dimensional (3D) distribution of water vapor in the atmosphere is a key element for weather forecasting and climate research. Moreover, atmospheric water vapor causes a delay in the microwave signal propagation. Thus, a precise determination of water vapor is required for accurate positioning and deformation monitoring using Global Navigation Satellite Systems (GNSS) and Interferometric Synthetic Aperture Radar (SAR) (InSAR). Yet, due to its high variability in time and space, the atmospheric water vapor distribution is difficult to model. This work therefore meets the challenge of tomographically reconstructing the 3D water vapor field by means of developing an innovative Compressive Sensing (CS) solution that includes both GNSS and InSAR observations. The CS solution is motivated by the fact that it does not include any geometric smoothing constraints as applied in common Least Squares (LSQ) approaches and by the fact that the sparse CS solution containing only a few non-zero coefficients may be determined, at a constant number of observations, based on fewer parameters than the corresponding LSQ solution. Within this work, the developed CS approach is compared to common LSQ solution strategies to water vapor tomography. The following details the relevance of an accurate determination of the 3D water vapor field for meteorology, climatology, and space geodesy. Thereafter, a motivation is given for the proposed CS approach using both GNSS and InSAR and for the analysis of the observing geometry's effect on the quality of the tomographic reconstruction. Finally, the goals and the contributions of this work are summarized and an outline of the thesis is given.

## 1.1 Relevance of determining the 3D atmospheric water vapor distribution

As water vapor is a necessary precondition to rainfall, it is important to accurately determine the 3D water vapor distribution in the atmosphere. According to Tuller [1973], the air has to be saturated with water vapor in order to form precipitation. Saturation is reached if the relative humidity RH attains 100%. If  $RH = 100\%$ , an air parcel is saturated with water vapor. Given a constant temperature, it is not possible to add any further water vapor to such a saturated air parcel. E.g. at a temperature of  $30^{\circ}\text{C}$  and a pressure of 1 bar, 1 kg of air can hold about 26 g of water vapor. Following the Clausius-Clapeyron equation describing the relationship between the saturation vapor pressure and the air temperature, the relative humidity increases if the water vapor density increases at a constant temperature or if the temperature decreases. Therefore, one key mechanism for attaining saturated air is a decrease in temperature, which may e.g. be reached by means of lifting the air parcel to higher altitudes as at the foot of mountains. Consequently, the windward side of a mountain range usually disposes of more precipitation than the lee side of the same mountain range. At the altitude at which  $RH = 100\%$ , clouds are generated.

However, water vapor only represents a necessary condition to rainfall. Water vapor does not correspond to a sufficient condition for the formation of rainfall. Besides the saturation of air with water vapor, condensation nuclei are necessary in order to form precipitation and the saturated air needs to condense on these particles in order to let grow the water droplets that, if the prevalent buoyant force and updraft is overcome by the weight of the droplets, fall down as precipitation. Condensation nuclei may e.g. originate from products of combustion, oxides of nitrogen, aerosols, or salt particles. According to Park [1999] and Xin and Reuter [1996], the low level water vapor content regulates the timing and the persistence of clouds and moisture convergence below the clouds and the timing and the quantity of rainfall. Thus, variations in the atmospheric water vapor content can cause significant changes in convective rainfall. A 1% moisture variation within a storm cell significantly affects

the storm intensity. Park and Drogemeier [2000] showed the effect of in-cloud moisture on the generation of secondary storm cells and the role of environmental moisture in strengthening a main storm cell. In addition, Stull [2016] states that storms get much of their energy from water vapor and that the amount of precipitation within a storm depends on the amount of moisture in the storm.

Besides its importance as a precondition to rainfall, water vapor corresponds to an important factor within the hydrological cycle. As stated in Mockler [1995], water vapor moves quickly through the atmosphere and redistributes energy associated with its evaporation and recondensation. In the case of evaporation, energy is taken up from the surroundings and the environment is cooled, analogously to the cooling effect of a sweating human body. When water vapor condenses, it releases energy and warms the environment. These heat exchanges influence the climate, and thus, good observations of the atmospheric water vapor are essential for understanding climatological processes.

In addition, an accurate knowledge of the 3D water vapor field is crucial for climatology, because water vapor is the most important natural greenhouse gas. According to Bowman [1990], without the natural greenhouse effect, the Earth's average temperature would be around 30°C lower than it is now. However, in contrast to the non-condensable or long-living greenhouse gases like CO<sub>2</sub>, methane, N<sub>2</sub>O, and halocarbons, atmospheric water vapor and clouds do not represent drivers of the greenhouse effect but act as fast feedbacks. I.e. water vapor reacts rapidly on changes in temperature, e.g. by evaporation, condensation, or precipitation. In case of anthropogenic emissions of CO<sub>2</sub>, methane, and other gases causing a rise in temperature, the evaporation rates and thus the atmospheric water vapor density is increased. The increased amount of water vapor, in turn, acts again as a greenhouse gas, absorbs energy that would otherwise escape to space, and so causes additional warming.

Finally, water vapor is an important error source in spaceborne GNSS and InSAR processing. In both techniques, radio wave signals are emitted by the satellite. On their way to the receiver on the ground resp. to the backscattering surface of the Earth (and back to the satellite, in the case of InSAR), these radio wave signals travel through the atmosphere and are delayed by the neutrospheric water vapor. As the neutrospheric water vapor distribution is highly variable both in time and space, the delays caused by water vapor are difficult to correct, when aiming at e.g. geodetic positioning or deformation modeling resp. surface movement applications in GNSS, or at e.g. topographic or deformation applications in InSAR. When compared to the dispersive ionospheric delay accelerating the signal propagation, the wet delay due to water vapor is much smaller but much more difficult to handle, as it cannot be corrected or modeled by multi-frequency measurements. According to Rothacher [2002], the neutrospheric delay in GNSS is highly correlated with the site height and with the receiver clock correction. Therefore it significantly affects the vertical component of the positioning and its effect on the site height needs to be reduced when aiming at precise positioning applications. Concerning InSAR, Zebker et al. [1997] state that the effects of water vapor are larger in magnitude and less evenly distributed throughout the interferogram than neutrospheric variations caused by pressure and temperature changes. Pressure variations typically cause 1.0 to 1.5 mbar root mean square pressure variabilities in temperate regions, where a 1 mbar pressure change leads to a delay change of about 2.3 mm. In contrast, Zebker et al. [1997] report phase changes due to neutrospheric water vapor corresponding to a delay of up to 30 cm. When translating these humidity variations into final deformation or topography products, according to Zebker et al. [1997], changes of 20 % in relative humidity lead to 10 cm errors in deformation products and perhaps 100 m of error in topographic maps for pass pairs with unfavorable baseline geometries.

## **1.2 Relevance of the innovations in the proposed approach**

In contrast to existing water vapor tomography approaches usually based on GNSS Slant Wet Delay (SWD) estimates only, this work also includes absolute wet delays obtained from InSAR. The inclusion of InSAR SWDs is relevant, because the profile-wise GNSS observations are, in general, sparsely distributed over the study area. This causes, especially in low atmospheric layers, a very low spatial resolution of the observations, which corresponds approximately to the inter-site distance of the GNSS sites of some tens of kilometers. Due to this poor spatial resolution of the GNSS rays in the lowest tomographic layers, the tomographic system of equations is ill-posed

and needs to be regularized. Including area wide InSAR SWDs may help to improve the observing geometry such that the observations are more evenly distributed.

Moreover, thanks to the launch of modern SAR missions such as Environmental Satellite (Envisat), TerraSAR, CosmoSkymed, or Sentinel-1, activities of Persistent Scatterer Interferometry (PSI) increased a lot. As stated in Hansen and Yu [2001], Parker [2017], or Tang et al. [2016], during PSI processing, the atmospheric phase screen can be estimated over wide areas at a temporal resolution of up to six days. Consequently, InSAR can be considered as a valuable resource for water vapor research and should be included within tomographic water vapor reconstructions.

When compared to previous research in the field of water vapor tomography, which is often based on LSQ, a CS solution approach is developed within this work. As in the case of the inclusion of InSAR data into the tomographic system, the use of CS is also motivated by the fact that the tomographic system of equations is ill-posed and needs to be regularized. The regularization can be performed e.g. by including additional observations (as those from InSAR or from surface meteorology) or by imposing constraints on the solution. Yet, the geometric smoothing constraints typically introduced within a LSQ solution to the tomographic system of equations showed to impose unnatural behavior to the water vapor distribution. Therefore, research on alternative regularization schemes is required. Based on the assumption that the signal can be sparsely represented in some appropriate transform domain, CS exploits the sparsity of the transformed water vapor signal for regularizing the tomographic system.

Finally, a reasonable discretization of the analyzed atmospheric volume into volumetric pixels (voxels) of constant water vapor content is essential for most water vapor tomography approaches. Independently of the choice of a LSQ or a CS reconstruction algorithm, the observing geometry, composed e.g. of the satellite positions, the GNSS site distribution, and the voxel discretization should be carefully chosen in order to enable a meaningful solution to the tomographic system. If too few observations are available in order to reconstruct the water vapor distribution at the chosen spatial resolution, or if these observations are too unevenly distributed, the applied constraints may falsify the solution and feign a much higher resolved, but unrealistic solution. This research emphasizes the importance of appropriate voxel sizes for accurately reconstructing the 3D water vapor field.

### 1.3 Objectives

This thesis aims at developing and analyzing a CS solution to tomographic water vapor reconstructions based on GNSS SWD estimates. In addition, an approach for including InSAR SWD estimates into the tomographic system is proposed. When comparing the reconstruction qualities of the LSQ and the CS solution strategy to the tomographic model, the thesis specially focuses on the questions i) which solution approach is more accurate and more precise, ii) in how far one of the strategies is more flexible, i.e. less constraint-driven, and iii) if the CS solution can do with fewer observations than LSQ. Alternatively, question iii) investigates in how far CS is able to estimate the neutrospheric water vapor field at a higher spatial resolution than LSQ, or if CS can estimate the water vapor field more accurately and more precisely than LSQ, given a certain number of observations and a certain spatial resolution. The influence of the observing geometry on the tomographic results is investigated by means of first analyzing the effect of the number of GNSS sites and of their horizontal distribution within the considered study area on the accuracy and on the precision of the tomographic results. This includes the questions i) if the site distributions should differ in different latitudes and ii) if the sites should be randomly distributed within the analyzed study area or rather situated along a regular grid. In addition, the effect of the ray geometry and of the voxel discretization on the tomographic results is investigated. This is done by means of focusing on the question in how far the inclusion of rays of more satellites than given by the Global Positioning System (GPS) orbits improves the repeatability of the results and by means of investigating how large the tomographic voxels should be in order to yield repeatable results within a changing orbit geometry.

## 1.4 Contributions

This thesis presents CS as a powerful tool for accurately reconstructing the 3D water vapor field. The sparsity of the water vapor field in an inverse Discrete Cosine Transform (iDCT) Euler Dirac transform domain is shown to be able to serve as a prior for regularizing the ill-posed system of equations encountered in the tomographic reconstruction of water vapor. Moreover, the capacity of InSAR SWD estimates to regularize the tomographic system resp. to improve the accuracy of the tomographic water vapor reconstruction is tested. Finally, the differences between the proposed CS approach and common LSQ approaches resp. between GNSS only resp. GNSS and InSAR solutions to water vapor tomography are analyzed. The thesis shows that, in the case of a CS solution, the observing geometry is of particular importance for the accurate determination of the 3D water vapor field based on profile wise SWD estimates. Moreover, the thesis points out the relevance of carefully selecting the size of the tomographic voxels and investigates the effect of the exact horizontal position of the GNSS sites on the accuracy and on the precision of the tomographic results. The developed tomography approach is based on geometrical and physical models that allow to combining and evaluating very different observation types and measurement techniques at a time. Therefore, the tomography approach represents a very flexible tool, especially under the light of new and heterogeneous satellite missions.

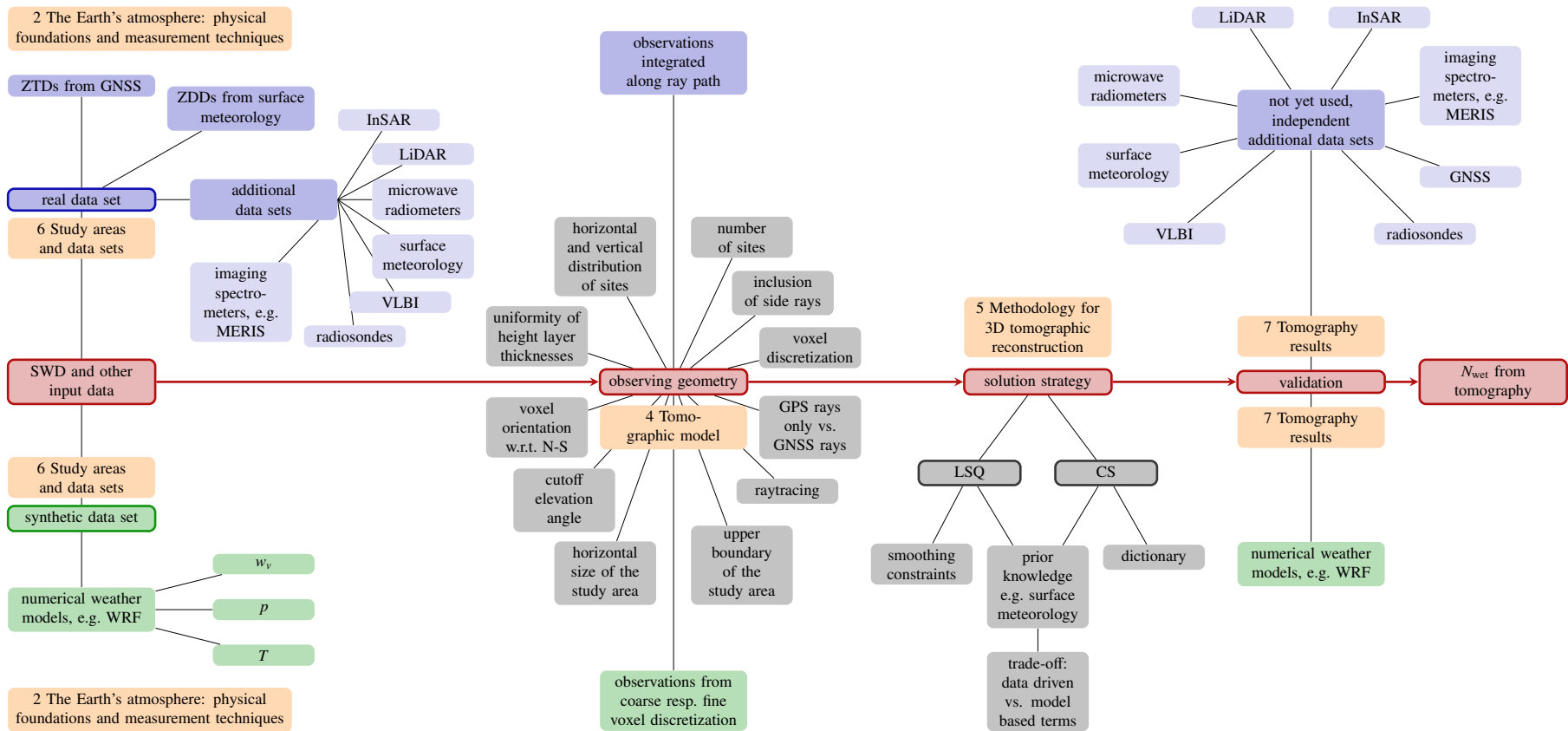
## 1.5 Outline

Many aspects need to be taken into account within a tomographic reconstruction of the 3D water vapor field. Therefore, Figure 1.1 shows an advance organizer illustrating the steps yielding 3D water vapor estimates based on e.g. slant wet delay input data. In the case of real data sets, common approaches to water vapor tomography mainly refer on input observations from GNSS and from surface meteorology. If available, e.g. InSAR SWD estimates and observations of Light Detection And Ranging (LiDAR) sensors, microwave radiometers, Very Long Baseline Interferometry (VLBI), radiosondes, or imaging spectrometers may be added. In contrast, in the case of synthetic data sets, the input SWDs are commonly deduced from meteorological quantities like temperature, pressure, and water vapor mixing ratio simulated by Numerical Weather Models (NWM).

In general, the available input data automatically define the number and the spatial distribution of the GNSS sites introduced within the tomographic system. Yet, many other components of the observing geometry need to be defined by the user, e.g. the horizontal size of the study area as well as its upper boundary, the horizontal voxel discretization as well as the thickness of the voxels and their orientation w.r.t. the North-South direction, the cutoff elevation angle for the introduced GNSS rays, the decision for GPS observations only or for GNSS observations, and the inclusion of rays entering the study area only on its top resp. also at the side.

Once the observing geometry settings are defined, a solution strategy can be selected. In this work, LSQ and CS solution approaches are distinguished for the determination of the 3D water vapor distribution. The validation of water vapor distributions estimated based on real data is carried out based on independent data sets that are not yet introduced within the tomographic system. In the case of synthetic data sets, a 3D validation of the estimated water vapor field is possible based on the simulations of e.g. temperature  $T$ , pressure  $p$ , and water vapor mixing ratio  $w_v$  resulting from NWM.

The main steps from the SWD input data to the interpretation of the tomography solution include the definition of the tomographic model according to the available data sets and the associated observing geometry, choosing a solution strategy, and validating the estimated 3D water vapor field. For each of these main steps, the advance organizer in Figure 1.1 indicates which section of this thesis describes resp. analyzes the respective step. In order to better guide the reader through the thesis, selected parts of the advance organizer will reappear at the beginning of the sections corresponding to the main steps of the tomographic reconstruction of the 3D water vapor field. The aspects relevant within the respective sections will then be highlighted within the advance organizer.

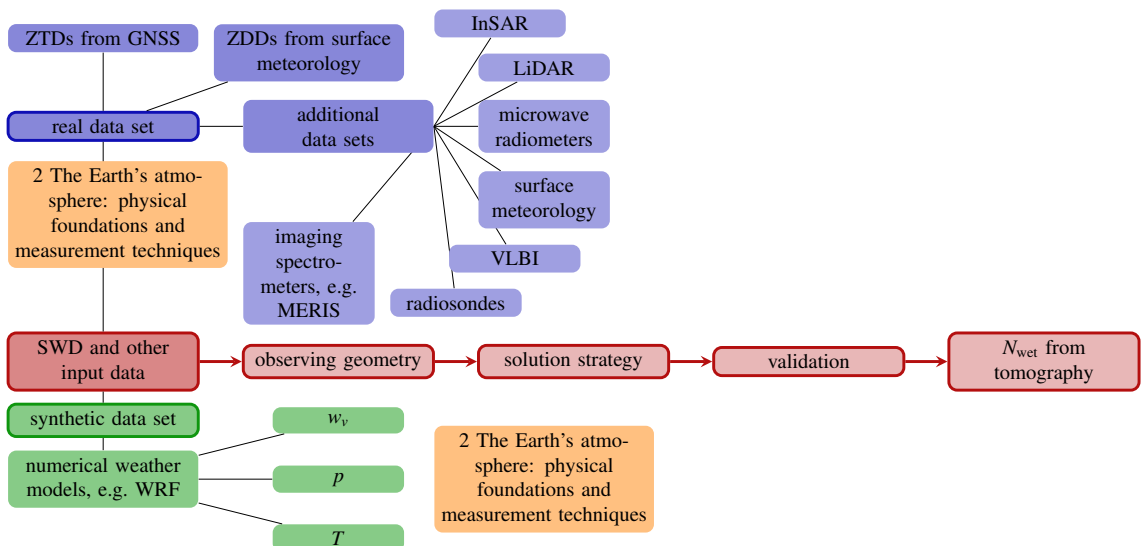


**Figure 1.1:** Advance organizer illustrating the steps yielding 3D water vapor estimates  $N_{wet}$  based on the input slant wet delay observations. The figure shall give an overview over the main steps of a water vapor tomography illustrated in red. In this way, the advance organizer shall prepare the reader to more efficiently handle and to better understand the thesis. At any one time that one of the steps is explained or investigated within the thesis, the advance organizer will appear and the respective step will be highlighted. For each of the main steps shown in red, relevant aspects effecting the respective step of the tomographic reconstruction are summarized within the advance organizer. Blue colors refer to the case of a tomographic reconstruction of water vapor using real data, whereas green colors stand for synthetic data. Gray colors are relevant for both real and synthetic data sets. The orange boxes indicate the sections in which the respective aspects are described resp. analyzed.



## 2 The Earth's atmosphere: physical foundations and measurement techniques

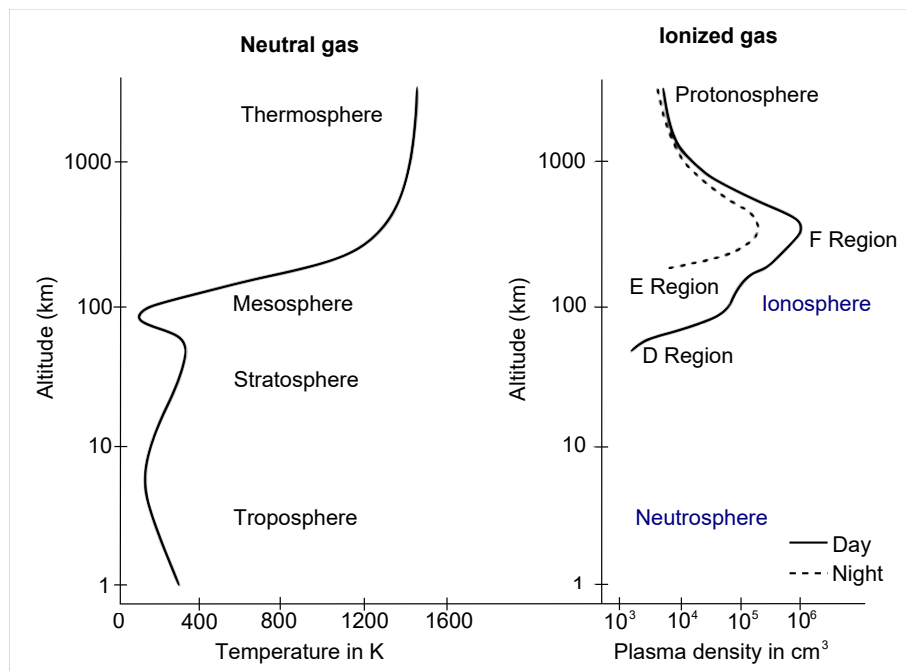
Tomographic water vapor modeling with space geodetic techniques like GNSS and InSAR necessitates a good understanding of the atmosphere's structure, of measurement techniques quantifying the water vapor distribution, and of interactions of electromagnetic waves with the atmosphere. Therefore, this chapter first gives an overview over the Earth's atmosphere and its common subdivision w.r.t. the temperature profile resp. w.r.t. the plasma density in Section 2.1. When subdividing the atmosphere according to the prevailing plasma density, the two main horizontal layers *neutrosphere* and *ionosphere* are distinguished. In this work, the focus is set on the electrically neutral lower  $\approx 10\text{km}$  of the atmosphere, the neutrosphere, in which most of the atmospheric water vapor resides. Therefore, Section 2.2 introduces meteorological quantities describing the neutrospheric water vapor distribution. The meteorological quantities are crucial for understanding the relation between the profile-wise SWD input data to water vapor tomography and the 3D water vapor field that shall be tomographically reconstructed. As illustrated in Figure 2.1, this work relies on both a real data set and synthetic data sets. The synthetic data sets are based on simulations of the Weather Research and Forecasting Model (WRF) model. Thus, Section 2.3 generally explains the modeling of the atmospheric state in NWMs for the synthetic data sets, and Section 2.4 presents, for the real data set, both meteorological and geodetic measurement techniques quantifying the neutrospheric water vapor content. Since radio waves and the neutrosphere interact, GNSS and InSAR radio wave observations can be used for water vapor tomography (Section 2.5). The GNSS and InSAR path delay modeling is explained in Section 2.6. In the case of InSAR, a special focus is set on illustrating how absolute wet delays per Persistent Scatterer (PS) can be obtained based on the temporally differentiated, short-wavelength water vapor signal observed as the InSAR atmospheric phase.



**Figure 2.1:** Water vapor tomographies aim at reconstructing the 3D water vapor field  $N_{\text{wet}}$  based on e.g. slant wet delay input data. In the case of real data, such SWD input data can e.g. be derived from GNSS, InSAR, and surface meteorological information. Alternatively, in synthetic data sets deduced from numerical weather models, the SWD input information to tomography is commonly deduced from meteorological quantities like pressure  $p$ , temperature  $T$ , and water vapor mixing ratios  $w_v$ .

## 2.1 The Earth's atmosphere

The Earth's atmosphere can be structured into different layers by means of, e.g. considering the prevailing temperature profile or the plasma density. Figure 2.2 shows these two subdivisions of the atmosphere. When considering the subdivision of the atmosphere with respect to the prevailing temperature profile, the lowest layer, extending from the Earth's surface up to the first temperature minimum, is called *troposphere*. In contrast, from the perspective of the ionization of the atmosphere, the lowest layer is called *neutrosphere*, followed, in higher altitudes, by the *ionosphere*. As stated in Kelley [2009], due to the pervasive influence of gravity, the neutrosphere and the ionosphere are to first order horizontally stratified. In this work, the focus is set on the non-ionized atmospheric regions, especially on the lowest 10km of the neutrosphere, containing a large amount of the atmospheric water vapor. Therefore, the term *neutrosphere* is preferred to the term *troposphere* in this thesis. The following two subsections give a short overview over the ionosphere and the neutrosphere.

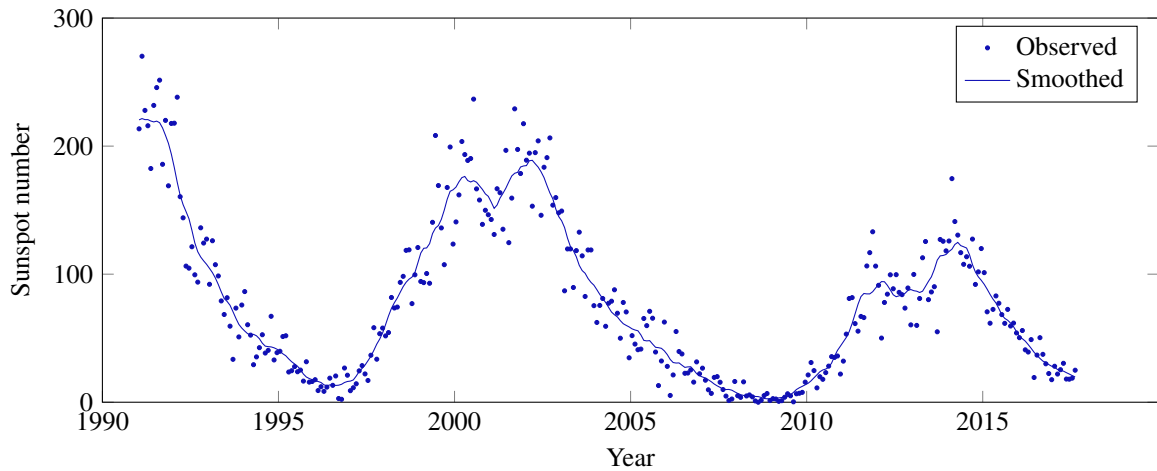


**Figure 2.2:** Typical profiles of atmospheric temperature and ionospheric plasma density according to Kelley [2009]. On the left, the various layers are distinguished by means of their temperature. On the right, the different layers are characterized by their ionization. The terms *ionosphere* and *neutrosphere* in blue were added to the figure from Kelley [2009].

### 2.1.1 Ionosphere

League [1997] describes the ionosphere, starting at a height of about 60km above the Earth's surface, as a region in which the air pressure is so low that free electrons and ions can move for a while without getting close enough to recombine into neutral atoms. The ionization is mainly caused by the ultraviolet solar radiation in the outer regions of the atmosphere. However, the ionization does not uniformly increase with the distance from the Earth's surface. Instead, there are different layers of high ionization at well-defined heights between about 40km to 300km above the Earth's surface. As the effect of the ionosphere on radio wave signals is related to the time of day, the season of the year, and variations in the solar activity characterized by the sunspot number, the intensity of the ionization within each layer and the exact layer heights vary over time. For the last two decades, Figure 2.3 exemplary shows that maximum sunspot numbers are observed at a repeat cycle of about eleven years. This repeat cycle is superposed by an 80 to 100 years super cycle.





**Figure 2.3:** Sunspot numbers of the last decades based on data from the American National Oceanic and Atmospheric Administration (NOAA) Space Weather Prediction Center <ftp://ftp.swpc.noaa.gov/pub/weekly/RecentIndices.txt> (09.07.2018).

In addition, as the ionosphere is a dispersive medium, the ionospheric delay on radio wave signals depends on the signal's frequency. Thus, as described in Hofmann-Wellenhof et al. [2008], multi frequency measurements enable a significant reduction of the ionospheric effect on the GNSS signal propagation time. Böhm and Schuh [2013] relate various GPS first-order measured parameters and the Total Electron Content (TEC) of the Earth's ionosphere. E.g. they state that a 1.000m range error, i.e. a delay of 3 ns, is caused by an electron content of  $6.15 \times 10^{16} \text{ el/m}^2$  resp. of  $3.73 \times 10^{16} \text{ el/m}^2$  at the GPS frequencies  $f_1$  resp. at  $f_2$ . In the case of InSAR, the ionospheric delay, however, is not corrected by dual frequency measurements. Thanks to the small wavelength of about 6 cm in the case of C-band SAR observations and because of the low solar activity around the data acquisitions in 2005, the ionospheric delay on InSAR as well as the Faraday rotation significant at L-band or lower frequencies (Jehle et al. [2005]) are neglected in this work.

### 2.1.2 Neutrosphere

According to Caballero [2014], the neutrosphere can be considered, to an excellent approximation, as a two component gas consisting of variable proportions of dry air and water vapor. Curry and Webster [1999] state that "water vapor is the most important gas in the atmosphere from a thermodynamic point of view because of its radiative properties as well as its ability to condense under atmospheric conditions". Water is the only atmospheric constituent that attains all three phases – gaseous, liquid, and solid – at the typical pressures and temperatures experienced in the Earth's atmosphere. According to Seidel [2002], nearly half of the atmospheric water vapor resides in the lowest 1.5 km of the atmosphere, and less than 5% are contained above 5 km. Radiosonde observations have shown that the water vapor content in heights above about 10 km is negligible in the study areas considered within this research.

## 2.2 Meteorological quantities describing the neutrospheric water vapor distribution

In order to get a better understanding of the both temporally and spatially highly variable neutrospheric water vapor distribution and the relation between water vapor, temperature, and humidity, this section introduces several meteorological quantities describing the neutrospheric water vapor field. The introduced quantities are essential

for relating the tomography input SWDs to the 3D water vapor field. According to Stull [2016], the 3D water vapor distribution can e.g. be expressed by the water vapor mixing ratio

$$w_v = \frac{m_{\text{water vapor}}}{m_{\text{dry air}}}, \quad (2.1)$$

which is commonly given in g/kg because the mass of water vapor  $m_{\text{water vapor}}$  within the air is much smaller than the mass of dry air  $m_{\text{dry air}}$ . Alternatively, the neutrospheric water vapor distribution can be described by the specific humidity  $q_v$ , also commonly given in g/kg,

$$q_v = 1000 \cdot \frac{w_v \text{ in kg/kg}}{1 + w_v \text{ in kg/kg}} = \frac{m_{\text{water vapor in g}}}{m_{\text{total air in kg}}}, \quad (2.2)$$

which can be related to the 3D water vapor field by solving

$$q_v = \frac{\epsilon' \cdot e}{p - e(1 - \epsilon')} \quad (2.3)$$

from Stull [2016] for the partial pressure  $e$  in hPa of water vapor:

$$e = \frac{q_v \cdot p}{\epsilon' + q_v \cdot (1 - \epsilon')} \quad (2.4)$$

In the Equations 2.3 and 2.4,  $p$  in hPa is the atmospheric pressure and the ratio between the gas constant of dry air and the gas constant of pure water vapor  $\epsilon' = 0.622$  is used. The factor  $0.622 = 18/29$  corresponds to the ratio of the molecular masses of water and air.

Finally, the 3D water vapor distribution can also be described by the wet refractivity field  $N_{\text{wet}}$ . The term *refractivity* will be explained in more detail in Section 2.5. According to Bevis et al. [1992],  $N_{\text{wet}}$  in ppm, with ppm standing for mm/km, is related to the partial pressure of water vapor  $e$  in hPa and to the temperature  $T$  in K as follows:

$$N_{\text{wet in ppm}} = k'_2 \cdot \frac{e}{T} + k_3 \cdot \frac{e}{T^2} \quad (2.5)$$

with

$$k'_2 = k_2 - k_1 \cdot \frac{M_{\text{water vapor}}}{M_{\text{dry air}}}, \quad (2.6)$$

from Davis et al. [1985], where the constants  $M_{\text{water vapor}} = 18.0153 \text{ g/mol}$  and  $M_{\text{dry air}} = 28.9647 \text{ g/mol}$  in Equation 2.6 are the molar masses of water vapor and of dry air. The constant factors  $k_1$ ,  $k_2$ , and  $k_3$  are given in Smith and Weintraub [1953] as:

$$\begin{aligned} k_1 &= 77.6 \text{ K/hPa} \\ k_2 &= 72 \text{ K/hPa} \\ k_3 &= 3.75 \cdot 10^5 \text{ K}^2/\text{hPa} \end{aligned} \quad (2.7)$$

Besides Smith and Weintraub [1953], many research on the refractive indices was carried out. Therefore, Rüeger [2002] summarizes the developments in refractive index equations for radio wave and millimeter waves over the years 1970 to 2000. In Table 2.1, typical values for  $q_v$  and  $e$  from Stull [2016] as well as the corresponding wet refractivities computed using Equation 2.5 are given for a pressure of  $p = 1013.25 \text{ hPa}$ .

**Table 2.1:** Typical values of  $q_v$  and  $e$  at a pressure of  $p = 1013.25 \text{ hPa}$ .

$T$ in $^\circ\text{C}$	$q_v$ in g/kg	$e$ in hPa	$N_{\text{wet}}$ in ppm
-10	1.7666	2.875	15.8284
-5	2.5956	4.222	22.3923
0	3.7611	6.113	31.2556
5	5.3795	8.735	43.0839
10	7.6005	12.320	58.6574

If the dew point temperature  $T_d$  and the temperature  $T$  are given instead of the partial pressure of water vapor, the value of  $e$  can be derived from the relative humidity RH and the temperature, divided by Kelvin, using

$$e = \frac{\text{RH}}{100} \cdot \exp(-37.2465 + 0.2131665 \cdot T - 0.000256908 \cdot T^2) \quad (2.8)$$

from Xu and Xu [2007]. Lawrence [2005] state that the relative humidity can be computed based on the temperature and the dew point temperature, both divided by Kelvin, by means of

$$\text{RH} = 100 - 5 \cdot (T - T_d). \quad (2.9)$$

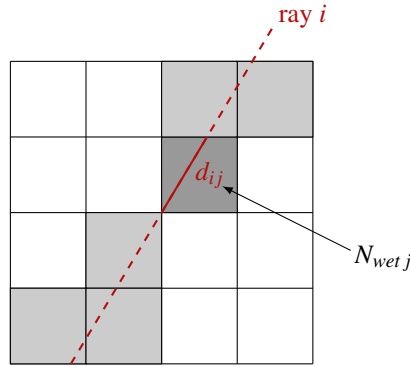
Integrating  $N_{\text{wet}}$  along the  $i$ th slant ray path  $\text{sp}_i$  with differential  $dl$  yields the observation equation for the  $\text{SWD}_{i,\text{cont}}$  along that path

$$\text{SWD}_{i,\text{cont}} \text{ (in m)} = 10^{-6} \cdot \int_{\text{sp}_i} N_{\text{wet}} \text{ (in ppm)} dl \text{ (in m)}, \quad (2.10)$$

or the observation equation for  $\text{SWD}_{i,\text{disc}}$  along discretized segments  $d_{ij}$  of the slant ray path in mm instead of m

$$\text{SWD}_{i,\text{disc}} \text{ (in mm)} = \sum_{j=1}^L N_{\text{wet},j} \text{ (in ppm)} \cdot d_{ij} \text{ (in km)}. \quad (2.11)$$

As illustrated in Figure 2.4, the variable  $d_{ij}$  corresponds to the distance passed by the slant ray path  $i$  within voxel  $j$ , and  $L$  is the total number of voxels within some tomographic grid.



**Figure 2.4:** Ray crossing the tomographic voxel grid.

In addition, the 3D wet refractivity field can be related to further integrated quantities like the Precipitable Water (PW), the Integrated Water Vapor (IWV), or the Zenith Wet Delay (ZWD) as indicated in Figure 2.5. The discretized formula for obtaining PW is given in the following equation:

$$\text{PW (in mm)} = \sum_{j=1}^L q_{v_j} \text{ (in g/kg)} \cdot \rho \cdot d_{\text{zenith},j} \text{ (in km)} \quad (2.12)$$

where  $d_{\text{zenith},j}$  represents the distance passed by the zenith ray path crossing the voxel  $j$ , and  $\rho = 1 \text{ g/cm}^3$  is the density of water.

According to Bevis et al. [1992], the precipitable water is related to the integrated water vapor in the zenith direction ( $\text{IWV}^{\text{zenith}}$ ) and to the ZWD as follows

$$\text{PW} = \frac{\text{IWV}^{\text{zenith}}}{\rho} = \Pi \cdot \text{ZWD} \quad (2.13)$$

and the conversion factors  $Q$  and  $\Pi$

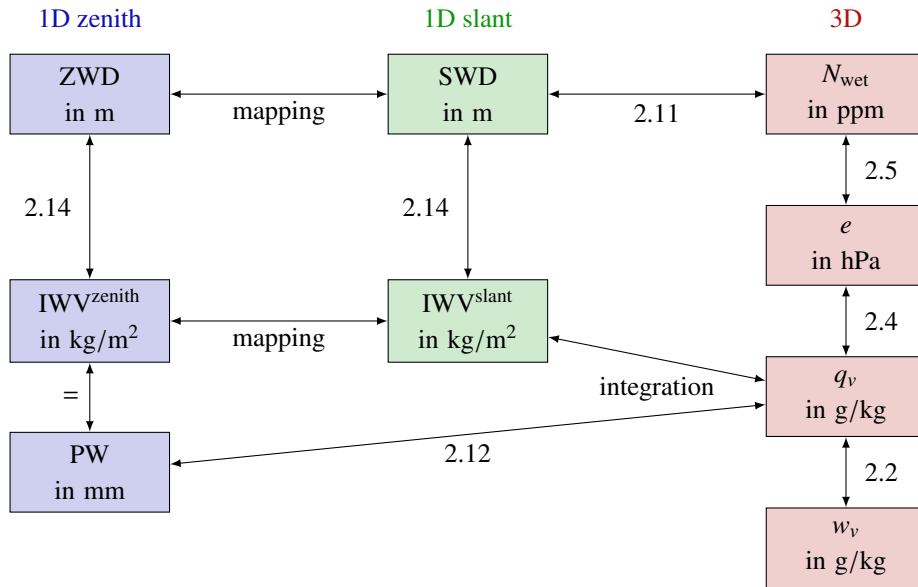
$$Q = 0.1022 + \frac{1708.08}{T_m \text{ in K}} = \frac{1}{\Pi} \quad (2.14)$$

between delays and IWV and its inverse  $\Pi$  can be approximated using

$$T_m \approx 70.2 + 0.72 \cdot T_0 \quad (2.15)$$

for the computation of the neutrospheric mean temperature  $T_m$  based on the surface temperature  $T_0$  in K. Alshawaf [2013] analyzed the sensitivity of  $Q$  w.r.t. to the surface temperature.

Figure 2.5 summarizes the relation between GNSS or InSAR integrated wet delays or precipitable water and the 3D water vapor mixing ratios simulated by the WRF model. The figure shows the relation between one dimensional (1D) wet delay input data to tomography and the 3D wet refractivity field  $N_{\text{wet}}$  to be estimated. In addition, related 1D quantities like the precipitable water or the integrated water vapor as well as related 3D quantities like the partial pressure of water vapor, the specific humidity, or water vapor mixing ratios are visualized. Understanding the relations between the respective quantities is important in order to be able to compare the water vapor tomography approaches resp. the tomography results proposed by different research groups. Moreover, in this work, the relations between the different humidity variables are essential when deducing synthetic SWD data sets from NWM.



**Figure 2.5:** Relations of the meteorological quantities describing the 3D water vapor distribution in the neutrosphere resp. 1D integrals of the neutrospheric water vapor field. The numbers on the arrows indicate the equation numbers corresponding to the respective conversions. A mapping projects the slant integrated quantities in the middle column into the quantities in the left column referring to the zenith direction.

## 2.3 Numerical weather models

Numerical weather models serve for describing the atmospheric behavior at a certain time and a certain location and are, in this work, used in order to generate synthetic SWD data sets for water vapor tomography. Alternatively, they could be used as prior knowledge to water vapor tomography or could be used for validation. Numerical weather models are mainly based on the equations of motion solving Eulerian equations for the three wind components, the first law of thermodynamics describing the air temperature, the water vapor budget equation modeling the total-water mixing ratio, the continuity equation describing the air density, and the general gas equation containing information on the air pressure. According to Stull [2016], all mentioned equations have a tendency term, i.e. all of them contain a description of the temporal changing rate of the respective variable. In addition, advection is included in all equations. Huschke et al. [1959] define advection as transport of an atmospheric property by the mass motion of the air. Besides, the forecast equations describing the behavior of the wind components,

the temperature, and the water vapor mixing ratio include a turbulence flux divergence term. As the equations describing the atmospheric behavior are non-linear and coupled, i.e. as each equation contains variables that need to be computed based on some of the other equations, all equations need to be solved together. Stull [2016] states that no analytical solution for the whole set of equations has been found so far, i.e. it is not possible to establish an algebraic equation that can be applied to every single location in the atmosphere.

In principle, three alternatives to such an analytical solution exist. Firstly, Stull [2016] proposes that an exact analytical solution could be established to a simplified version of the governing equations. A second alternative is to conceive a simplified physical model for which exact equations can be solved. Finally, an approximate numerical solution to the full governing equations can be computed. This latter method is pursued in the case of modern numerical weather prediction, in which the solutions of numerical approximations of the mentioned equations are determined at discrete, regularly spaced grid points only. These approximate solutions include both temporal and spatial discretization. That is, the continuum of space is subdivided into discrete grid cells, and instead of considering a smooth progression of time, the computations are performed at discrete time steps only.

In addition to the variables described by the equations explained above, there may be further atmospheric processes that are not well understood although their effects can be measured. Alternatively, there may be processes that may involve motions that are too small to be resolved or that might be too complicated to compute in finite time. As such physical effects cannot be neglected, they are physically or statistically parameterized by means of one or more known terms within numerical weather models. Besides others, the physics parameterizations in Numerical Weather Prediction (NWP) include cloud coverage, deep convection, radiation, and turbulence.

Typical grid cell sizes in NWP extend from one to hundreds of kilometers horizontal length and from one to hundreds of meters in the vertical thickness. The smaller the grid cells, the higher the computational cost of the model solution. Therefore, in the horizontal, a fast-running coarse grid can be used over a large domain, in order to then embed a smaller-domain nested grid inside it. The coarse grid serves for modeling synoptic weather systems, whereas the finer grid can capture mesoscale details. According to Stull [2016], the nesting can be continued within the finer grid with successively finer nests. Due to important small-scale motions and strong vertical gradients, the highest resolution in the vertical direction is necessary close to the Earth's surface.

In contrast, the spatial resolution of the GNSS SWD observations commonly used as input for the tomographic reconstruction of the water vapor field, is particularly low in the lowest atmospheric layers. Therefore, many approaches to water vapor tomography include additional measurement techniques quantifying the atmospheric water vapor distribution.

## **2.4 Measurement techniques quantifying the neutrospheric water vapor content**

Both meteorological and geodetic measurement techniques can be used in order to sense the atmospheric water vapor. As shown in Table 2.2, meteorological and geodetic sensors contributing to the determination of the atmospheric water vapor content can be classified depending on their position during the measurement. On the one hand, ground-based observations are performed in order to quantify the atmospheric water vapor content, on the other hand, balloon-, air- or spaceborne sensors are used for the measurement of the neutrospheric water vapor content or of related quantities like the specific humidity, the partial pressure of water vapor, or the wet refractivity introduced in Section 2.2. Ground-based meteorological measurement techniques include surface meteorological sites, radiometers, and LiDAR.

Surface meteorological sites mainly collect meteorological observations of pressure, temperature, and humidity. However, as surface meteorological measurements are strongly related to land-air interactions, Braun [2004] warns that the observations of surface meteorological sites in general do not accurately represent the entire boundary layer.

According to Stull [2016], microwave radiometers are passive sensors measuring the intensity of upwelling electromagnetic radiation at millimeter-to-centimeter wavelengths (microwaves) emitted from the Earth and from the atmosphere. As stated in Rocken et al. [1993], this is done by means of measuring brightness temperatures. In the case of ground-based Water Vapor Radiometers (WVR), the IWV along the radiometer's line of sight is deduced from measurements of water vapor brightness temperatures around two different frequencies, one of them sensitive to water vapor, and the other one sensitive to liquid water. In this way, the portion of liquid water can be separated from that of water vapor. Using the optimal estimation technique described in Rodgers [2000], vertical water vapor profiles can be deduced from the measured water vapor brightness temperature spectrum. The temporal resolution of water vapor radiometer observations is generally very high. However, the spatial resolution of ground-based water vapor sensing techniques is usually poor. In contrast to this, spaceborne down-looking WVRs provide observations at high spatial and poor temporal resolution.

Similar to WVRs, that can be installed on the ground, on aircrafts, and on satellites, LiDAR sensors can also operate satellite-based, airborne, and ground-based. According to Stull [2016], the basic principle of atmospheric LiDAR measurements consists in transmitting electromagnetic radiation at two neighboring frequencies. One of these frequencies is required to be affected by water vapor and the other one should not be affected by water vapor. Whiteman et al. [1992] state that the water vapor mixing ratio can be determined from the LiDAR data by evaluating the Raman shifted signals from water vapor and from nitrogen. Nitrogen is used, because it is known to be in constant proportion to dry air.

Besides WVRs and LiDAR, satellites might carry imaging spectrometers for water vapor sounding like the MEdium Resolution Imaging Spectrometer (MERIS) resp. the MODerate Resolution Imaging Spectrometer (MODIS). Under clear skies, these infrared sensors are capable to measure the IWV with horizontal resolutions of 1 km resp. of 300m in the case of MODIS resp. in the case of MERIS. The IWV is deduced from the attenuation of near-infrared (IR) solar radiation by water vapor. Fischer and Bennartz [1997] explain that ratios of two close channels, one within and one outside the absorbing band of water vapor, are computed in order to get information on the IWV. In the case of clouds, the measured IWV corresponds to the water vapor content from the sensor to the top of the cloud. Therefore, in order to avoid misinterpretation when dealing with IWV information, no water vapor values are given in cloudy areas of MERIS images.

Imaging spectrometers can be subdivided into visible, IR, and water vapor imager. Stull [2016] states that IR imaging spectrometers use long wavelengths in an IR transmittance window. Consequently, they are able to clearly see through the atmosphere to the Earth's surface resp. to the top of the highest cloud. IR imaging spectrometers dispose of a day and night imaging capability because the Earth never cools down to absolute zero at night and thus always emits IR radiation. In contrast to these IR imaging spectrometers, water vapor imaging spectrometers obtain images by picking a wavelength that is situated outside of the transmittance windows of the atmosphere. The IR radiation emitted by the Earth can only reach the satellite in case of a small amount of water vapor present in the atmosphere. Satellite images from the visible wavelength range capture all those features that the human eye would also capture. As a consequence, water vapor cannot be seen in such satellite images. Examples for imaging spectrometers sensing the integrated water vapor are e.g. MODIS and MERIS, parts of the payloads of the Terra, Aqua, and Envisat satellites.

Balloon-borne radiosonde observations complete the ground-based, air-, and spaceborne techniques measuring the atmospheric water vapor. According to Dabberdt et al. [2002], a radiosonde is a device measuring the vertical profile of meteorological variables and transmitting the measured data to a ground-based receiving and processing station. From the surface up to heights of about 30 km, the vertical variation of temperature, pressure, and humidity along the balloon ascent are routinely measured.

When considering geodetic sensors, VLBI, GNSS, and InSAR can be applied in order to deduce information on the atmospheric water vapor content. VLBI is based on the observation of extragalactic radio sources (e.g. quasars) simultaneously by at least two ground-based radio antennas. Whereas a local radio interferometry uses a direct cable to connect the two antennas with a correlator evaluating the observed signals, each VLBI site needs an accurate atomic clock in order to combine the signals of very distant antennas and to thus enable an analysis

**Table 2.2:** Meteorological and geodetic sensors resp. techniques measuring the atmospheric water vapor content, humidity, or similar. The first five lines contain meteorological humidity measurement techniques, whereas the last three lines are rather considered as geodetic techniques that can be applied for water vapor sensing. The letter ‘M’ in the first column indicates meteorological sensors resp. measurement techniques, the letter ‘G’ stands for geodetic sensors and techniques. In case of an entry ‘none’ in the sensing requirements, the sensors or techniques are able to measure day and night and (nearly) independently of the weather. The sensors resp. techniques available within the real data set in this work are highlighted in green.

	Sensor/Technique	Observing geometry	Humidity parameters	Spatial resolution	Temporal resolution	Sensing requirements
M	Surface meteorology	ground-based	$p$ , $T$ , RH, and others	tens of km (low)	30 min ... 2 hourly	none
M	Microwave radiometers	ground-based, air-, or spaceborne	IWV using brightness temperature; water vapor profiles from inversion	very low (few operational instruments)	seconds ... minutes	no rain on instrument
M	LiDAR	ground-based, air-, or spaceborne	water vapor profiles	low ... high depending on observing geometry	low ... high depending on observing geometry	no clouds or observation only from surface to cloud
M	Radiosondes	balloon-borne	$p$ , $T$ , RH, IWV, and others	very low (2003 in average 315 km in the US)	typically 12h	none
M	Imaging spectrometers, e.g. MERIS or MODIS	spaceborne	IWV using the attenuation of near IR solar radiation by water vapor	high	low (depending on orbit repeat time)	cloud-free sky
G	VLBI	ground-based	IWV	very low (few sites)	days ... minutes (depending on schedule)	none
G	GNSS	spaceborne	IWV	tens of km (low)	minutes	none
G	InSAR	spaceborne	partial IWV differences	e.g. $5 \times 20 \text{ m}^2$ in C-band PS InSAR (high)	at least 6 days	none

of the observed radio wave signals. The main application of VLBI is an accurate determination of the Earth Orientation Parameters (EOP) and of the coordinates of the VLBI sites resp. the observed radio sources. Yet, as these parameters are correlated with the neutrospheric delay and thus with the atmospheric water vapor content, Böhm [2004] states that a determination of the EOP or of the VLBI site coordinates requires an exact modeling of the water vapor content. Therefore, approaches estimating a site's position, nutation parameters, and the prevailing wet delay at a time are possible as described in Niell et al. [2001].

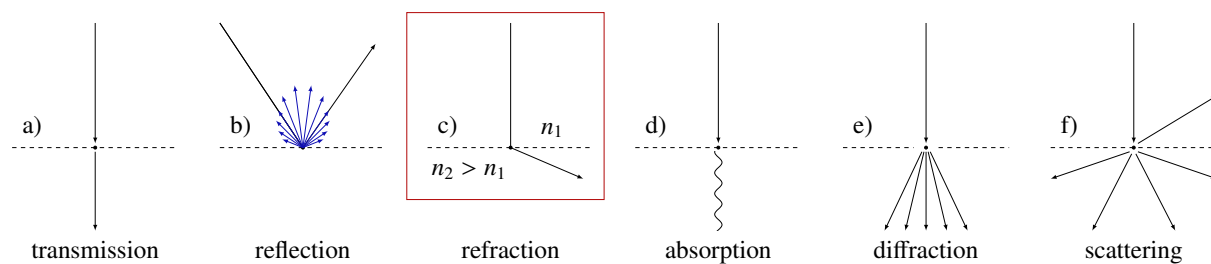
As VLBI, GNSS was not designed to measure the atmospheric water vapor distribution, but rather with the intention of positioning or timing applications. Though, as shown by Bevis et al. [1992], GNSS is a powerful tool for determining the atmospheric water vapor content, which is, similarly to the case of VLBI, correlated with the GNSS site height and with the receiver clock parameters. Analogously to GNSS, although originally designed e.g. for capturing surface displacements or generating Digital Elevation Models (DEM), InSAR is able to determine partial IWV differences using a certain combination of spatio-temporal filtering routines. Further information on the neutrospheric effects on GNSS and InSAR and on the determination of water vapor information using these measurement techniques are given in Section 2.5.

## 2.5 Interactions of radio waves and the neutrosphere

As radio waves and neutrosphere interact, the neutrosphere effects both the GNSS and InSAR measurement techniques, and thus a tomographic reconstruction of the atmospheric water vapor field by means of GNSS and InSAR becomes possible. This section explains different kinds of interaction of radio waves and the neutrosphere. This helps to accentuate the interaction between the neutrosphere and radio wave signals on which approaches to water vapor tomography rely.

According to Barclay [2003], the atmospheric effect on radio wave signals can be subdivided into two main behaviors, depending on the state of matter of the atmospheric constituents interacting with the radio waves. In general, influences of atmospheric gases like reflection, absorption, or effects on the refractivity index (causing delays and bending) are distinguished from interactions with solid or liquid constituents of the atmosphere (as clouds or aerosols) like scattering, absorption, and scintillation.

Figure 2.6 shows some of the possible interactions of a wave with the medium through which the wave is traveling. Besides others, the wave can be transmitted, reflected, refracted, absorbed, diffracted, or scattered.



**Figure 2.6:** Wave behaviors: a) transmission; b) black: specular reflection, blue: diffuse reflection; c) refraction; d) absorption; e) diffraction; f) scattering. The red frame highlights the refraction causing the neutrospheric delays used in this work in order to reconstruct the 3D refractivity field.

If a wave is transmitted through the object, it can pass through the object without interacting with it. In the case of a specular reflection, the wave hits a smooth surface and bounces off. In this context, the signal's wavelength must be large when compared with the roughness of the smooth surface. A wave (or a part of a wave's energy) is absorbed if the wave causes a vibration of the atoms and molecules within the hit object. The vibration heats up the object and the heat is emitted as thermal energy. Diffraction corresponds to the bending of a wave when it encounters an obstacle. Scattering occurs when a wave encountering an object bounces off in many different directions.



Refraction occurs if a wave travels from a medium with some refractive index  $n_1$  to a medium with some other refractive index  $n_2$  resulting in a bending of the ray path as also observed in electronic distance measurements (EDM).

More information on refraction is given in Section 2.6, focusing particularly on ionospheric and neutrospheric refraction. The neutrospheric refraction is subdivided into refraction caused by the dry part of the neutrosphere and refractivity causing wet delays on radio wave signals. Thereafter, Section 2.6.1 explains how the dry delay can be modeled based on surface meteorological measurements, and the Sections 2.6.2 resp. 2.6.3 describe the effects of the wet refractivity on GNSS resp. InSAR observations and how information on the atmospheric water vapor can be deduced from these measurement techniques.

## 2.6 Path delay modeling in GNSS and InSAR

Both GNSS and InSAR satellites emit electromagnetic waves propagating, in vacuum, with the speed of light  $c$ . On their way from the satellites to the ground (and back to the satellite, when thinking of InSAR), the signals pass the Earth's atmosphere. Equation 2.16 describes the relation between the propagation velocity  $v$  in a medium with refractivity  $n$ :

$$n = \frac{c}{v} \quad (2.16)$$

Fermat's principle cited in Hofmann-Wellenhof et al. [2008] states that an electromagnetic signal follows the path between source and observer which takes the least amount of time. Due to the variation of the air masses or densities along the ray path of a radio wave, the refractive index  $n$  along the signal transmitting path differs from that in vacuum and takes a value slightly greater than unity, varying along the signal's path through the neutrosphere. As a consequence, according to Forssell [2008], the ray path deviates from a straight line and the signal propagation is delayed. The refractive index in Equation 2.16 is related to the total refractivity  $N_{\text{total}}$  by means of

$$N_{\text{total}} = 10^6 \cdot (n - 1). \quad (2.17)$$

The ionospheric refraction is dispersive, i.e. depends on the signal's frequency, whereas the neutrospheric refraction is related to the refractivity of gases, hydrometeors, and other particulates. The total refractivity and the total delay on radio wave signals caused by refractivity are commonly subdivided into two parts, e.g. into a *dry* and a *wet* component or into a *hydrostatic* and a *non-hydrostatic* part. The following sections explain how to model the hydrostatic component of the delay by means of surface meteorology, and describe how to determine the wet component of the delay using GNSS Precise Point Positioning (PPP) resp. PS-InSAR.

### 2.6.1 Modeling the hydrostatic component of the delay using surface meteorology

As stated in Section 2.5, the total refractivity and the total delay on radio wave signals caused by refractivity are commonly subdivided into two parts, e.g. into a *dry* and a *wet* component or into a *hydrostatic* and a *non-hydrostatic* part. According to Saastamoinen [1972], the hydrostatic component ZHD of the neutrospheric delay only contains the delay caused by dry gases and can be computed based on the surface pressure  $p_0$ , the zenith angle  $z$ , and correction terms  $D$  and  $B$  using

$$\text{ZHD} = 0.002277 \cdot D \cdot [p_0 - B \cdot \tan^2 z]. \quad (2.18)$$

In contrast, when applying a subdivision into dry and wet components, the dry component

$$\text{ZDD} = 0.002277 \cdot D \cdot [p_0 - 0.155471 \cdot e_0 - B \cdot \tan^2 z] \quad (2.19)$$

also contains contributions of the partial pressure of water vapor at the surface  $e_0$ . In the Equations 2.18 and 2.19, the components  $D$  and  $B$  are correction terms, where  $D$  is calculated from the ellipsoidal site height  $h$  and the latitude  $\varphi$  using the formula

$$D = 1 + 0.0026 \cdot \cos 2 \cdot \varphi + 0.00028 \cdot h \text{ in km}, \quad (2.20)$$

and  $B$  is taken from lookup tables depending on the site height  $h$ .

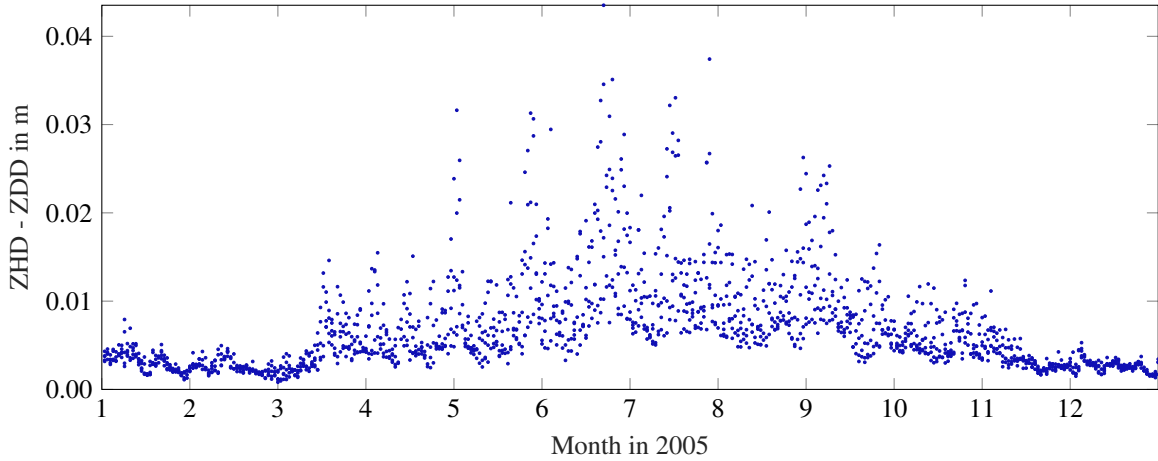
If a hydrostatic equilibrium can be assumed, the hydrostatic component can be accurately computed based on surface pressure. Therefore, in this work, the total refractivity or delay is subdivided into a hydrostatic and a non-hydrostatic part. However, for reasons of readability, and consistently with the International Earth Rotation and Reference Systems Service (IERS) conventions of Petit and Luzum [2010], the terms *wet* refractivity resp. *wet* delay are used in the following for the non-hydrostatic component  $N_{\text{wet}}$  of the refractivity resp. of the delay, yielding, with a hydrostatic refractivity  $N_{\text{hydrostatic}}$ ,

$$N_{\text{total}} = N_{\text{hydrostatic}} + N_{\text{wet}} \quad (2.21)$$

resp.

$$\text{ZTD} = \text{ZHD} + \text{ZWD}. \quad (2.22)$$

Figure 2.7 shows differences between dry and hydrostatic delays, which are based on surface meteorological observations in Karlsruhe. In humid summer months, the difference between Zenith Hydrostatic Delays (ZHD) and Zenith Dry Delays (ZDD) composing the Zenith Total Delays (ZTD) attains values of up to 4 cm.



**Figure 2.7:** Differences in m between dry and hydrostatic delays from surface meteorological measurements at Karlsruhe (synoptic site 4444).

As described in Section 2.6.3, a computation of the zenith hydrostatic delay is necessary both for deducing GNSS ZWDs and for deriving absolute ZWDs from a combination of partial InSAR ZWD differences, GNSS ZWDs, and surface meteorological observations. The ZHD is modeled using the Saastamoinen model from Equation 2.18.

As surface meteorological measurements are often spatially poorly sampled, Dach et al. [2007] use standard atmospheres in order to deduce the surface pressure  $p_0$ , the surface temperature  $T_0$ , and the partial pressure of water vapor at the surface  $e_0$  for the computation of the ZHD, see also Berg [1948]. Alternatively, the ZHD can be computed using surface meteorological observations. Based on pressure and temperature observations  $p_{\text{meteo}}$  and  $T_{\text{meteo}}$  of the closest synoptic site, Alshawaf et al. [2015b] and Alshawaf [2013] compute the pressure  $p_{\text{GNSS}}$  at a certain location, e.g. a GNSS site, using the hydrostatic equation

$$p_{\text{GNSS}} = p_{\text{meteo}} \left( 1 - L_T \cdot \frac{h_{\text{GNSS}} - h_{\text{meteo}}}{T_{\text{meteo}}} \right)^{\frac{g \cdot M_{\text{dry air}}}{R \cdot L_T}} \quad (2.23)$$

with  $h_{\text{GNSS}}$  resp.  $h_{\text{meteo}}$  standing for the height of the considered GNSS site resp. for the height of the used surface meteorology site. The temperature lapse rate  $L_T$  is equal to  $6.5^\circ/\text{km}$ , the quantity  $g = 9.80665 \text{ m/s}^2$  is the mean

Earth's gravity acceleration,  $R$  is the universal gas constant  $R = 8.31447\text{J}/(\text{mol} \cdot \text{K})$ , and  $M_{\text{dry air}}$  is the molar mass of dry air. When applying Equation 2.23, in the case of a smooth topography, a meteorological site may be up to 100km distant to the GNSS site for which the pressure shall be computed.

If multiple meteorological sites are available close to the location for which the ZHD shall be computed, the pressure resp. temperature at the height of the location can be obtained by means of a linear regression model. To account for spatial variations, Alshawaf et al. [2015b] use the residuals of this linear regression to calculate a correction value at any point by applying, for example, inverse distance weighting.

## 2.6.2 Determining the wet component of the delay content using GNSS PPP

Subtracting the ZHD obtained from surface meteorological observations as described in the previous section from GNSS ZTDs yields the both spatially and temporally highly variable and thus difficult to precisely model neutrospheric Zenith Wet Delay ZWD. The GNSS ZTDs are deduced from an overall least squares adjustment also yielding the receiver coordinates, the receiver clock correction, and the carrier-phase ambiguities.

In this work, the GNSS processing is performed using the PPP strategy introduced in 1997 by Zumberge et al. [1997]. In contrast to differential GNSS positioning approaches, PPP requires only one single receiver. As a reference station is no longer necessary, the efforts and the equipment cost within the field work are reduced. Moreover, the PPP solution refers to a global reference frame whereas differential GNSS solutions are relative to the local base station. Therefore, PPP enables a greater positioning consistency than differential positioning. Moreover, there are no common parameters between the different sites to be solved within the PPP processing, which makes it possible to process the data site by site. Consequently, large networks can be processed on distributed systems using PPP. In return, the initialization time in PPP is longer than in differential positioning, which renders the use of PPP challenging in the case of real time measurements. Moreover, when compared to differential GNSS, orbit deviations and satellite clock offsets are no longer minimized or eliminated by building differences (Teunissen and Montenbruck [2017], Rizos et al. [2012]). That is, precise ephemerides as well as satellite clock corrections, provided e.g. by the International GNSS Service (IGS), are absolutely necessary in the case of PPP.

Since the satellite clock offset is known precisely for PPP, Kouba and Héroux [2001] describe the ionosphere-free linear combinations  $l_{\rho_{\text{pseudo}}}$  resp.  $l_{\Phi}$  of dual-frequency GPS pseudo-range resp. carrier-phase observations  $\rho_{\text{pseudo}}$  and  $\Phi$  as a function of

- the geometric range between the satellite at its Earth Centered Earth Fixed (ECEF) coordinates  $(X_{\text{SV}}, Y_{\text{SV}}, Z_{\text{SV}})$  and the receiver at its ECEF coordinates  $(x_{\text{rec}}, y_{\text{rec}}, z_{\text{rec}})$ ,
- the receiver clock offset  $\Delta t$ ,
- the slant total delay STD obtained from the zenith total ZTD delay using  $\text{STD} = \text{mf} \cdot \text{ZTD}$  with a mapping function mf,
- the ambiguity  $\mathcal{N}$  of the carrier-phase at a carrier-phase wavelength  $\lambda_{\Phi}$ ,
- the constant speed of light  $c$ ,
- and measurement noise on the pseudo-range resp. on the carrier-phase signal  $\epsilon_{\rho_{\text{pseudo}}}$  resp.  $\epsilon_{\Phi}$

as shown in the following equation:

$$\begin{aligned} l_{\rho_{\text{pseudo}}} &= \rho_{\text{geom}} + c \cdot \Delta t + \text{mf} \cdot \text{ZTD} + \epsilon_{\rho_{\text{pseudo}}} \\ l_{\Phi} &= \rho_{\text{geom}} + c \cdot \Delta t + \text{mf} \cdot \text{ZTD} + \mathcal{N} \cdot \lambda_{\Phi} + \epsilon_{\Phi}, \end{aligned} \quad (2.24)$$

In Equation 2.24, the Euclidean distance  $\rho_{\text{geom}}$  between the satellite and the receiver is

$$\rho_{\text{geom}} = \sqrt{(X_{\text{SV}} - x_{\text{rec}})^2 + (Y_{\text{SV}} - y_{\text{rec}})^2 + (Z_{\text{SV}} - z_{\text{rec}})^2}. \quad (2.25)$$

By means of a linearization of the observation equation from Equation 2.24 around some a-priori parameters, a linear system of equations can be established that can then be solved using a least squares adjustment. The design

matrix for this linear problem is obtained by computing the partial derivatives of the single terms of Equation 2.24 w.r.t. the unknown parameters, i.e. w.r.t. the receiver position, the receiver clock offset, the slant total delay, and the carrier-phase ambiguities.

In this work, the goal of the GNSS processing consists in the determination of SWDs as input for the tomographic reconstruction of water vapor. Based on the ZTDs deduced from the GNSS adjustment, SWDs can be computed. This is done by means of subtracting the ZHDs derived from the Saastamoinen model in Equation 2.18 from the ZTDs

$$\text{ZWD} = \text{ZTD}^{\text{estimate}} - \text{ZHD}^{\text{model, meteo}}. \quad (2.26)$$

and by means of mapping these ZWDs to the azimuth angles  $A$  and to the elevation angles  $\epsilon$  of the respective GNSS rays. The mapping is performed using a wet mapping function  $\text{mf}_{\text{wet}}$ . In order to obtain complete GNSS SWDs, horizontal neutrospheric gradients in Northing  $\Delta N$  and Easting  $\Delta E$  are added as proposed by Chen and Herring [1997]:

$$\begin{aligned} \text{SWD} = & \left( \text{ZTD}^{\text{estimate}} - \text{ZHD}^{\text{model, meteo}} \right) \cdot \text{mf}_{\text{wet}} + \\ & + \Delta N \cdot \frac{\cot \epsilon}{\sin \epsilon} \cdot \cos A + \Delta E \cdot \frac{\cot \epsilon}{\sin \epsilon} \cdot \sin A, \end{aligned} \quad (2.27)$$

When including all components considered in this work, the SWDs are composed of

- ZWDs, i.e. the difference between estimated ZTDs using the Berner GPS Software (Bernese) software and the dry model deduced e.g. from surface meteorology as described in Section 2.6.1
- the mapping of the ZWDs according to the elevation angle of the considered GNSS satellite, and
- horizontal gradients in Northing and Easting  $\Delta N$  and  $\Delta E$  estimated by Bernese.

While the horizontal gradients are mapped by a simple  $1/\sin \epsilon$  mapping function, the SWDs are mapped using the wet Niell mapping function given in Equation A.1. The Niell mapping function has been selected for two main reasons. For high elevation angles, the curvature of the atmosphere can be ignored and a simple  $1/\sin \epsilon$  mapping is sufficient. At lower elevation angles, this is not the case anymore. At the cutoff elevation angle  $\epsilon_{\text{cut}} = 7^\circ$  used in this work, the error in the SWD obtained by a simple  $1/\sin \epsilon$  mapping is over 3%, thus not negligible. If the curvature of the atmosphere were ignored in the tomographic method, i.e. if the voxels were cuboids, such a simple mapping function might still be acceptable, but within the ellipsoidal shape of the atmosphere considered in this work, the SWD should also contain the effect of the ellipsoidal voxel shape. A simple  $1/\sin \epsilon$  mapping is not appropriate for low elevation rays and consequently, i) a more advanced mapping function is required. Yet, ii) as meteorological parameters might be unavailable in some cases, the wet Niell mapping function depending only on the site latitude was selected for this work. More details about the Niell mapping function are given in Appendix A.2.

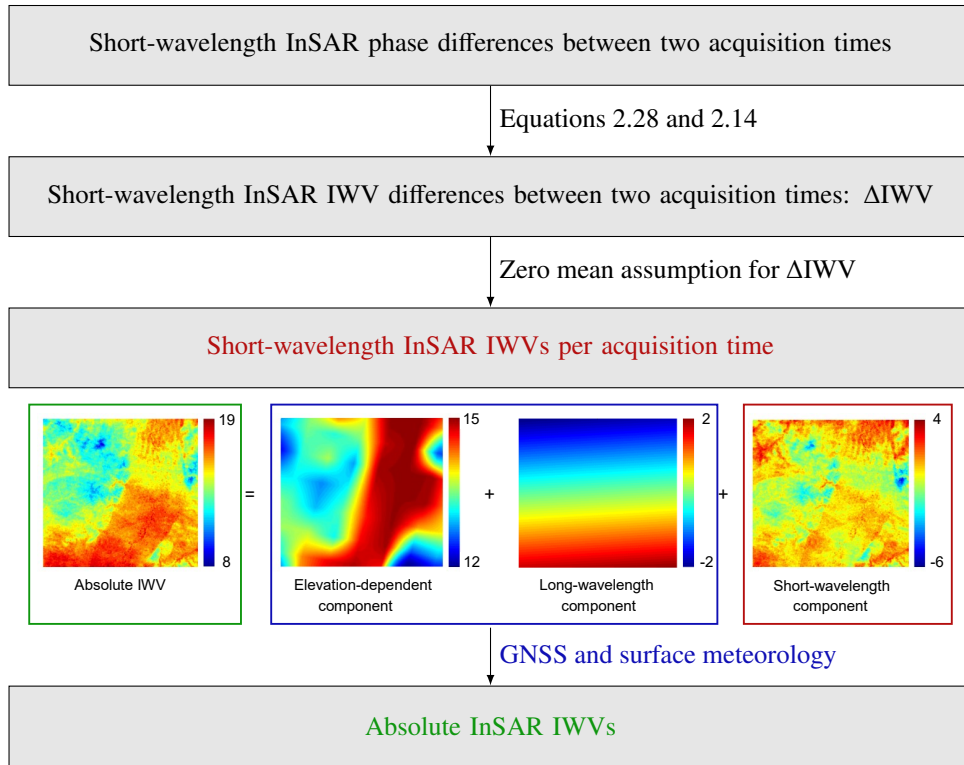
### 2.6.3 Determining the wet component of the delay using PS-InSAR

Due to the emission of pulses of electromagnetic radiation, RADAR sensors can be classified as *active sensors*. According to NTIA [2015], an active sensor is a measuring instrument in the Earth exploration-satellite service or in the space research service by means of which information is obtained by transmission and reception of radio waves. In contrast to passive sensors, active sensors are detection devices requiring input energy from a source other than that which is being sensed.

Having its own illumination, radar sensors have a day-and-night measuring capability. In addition, the radar microwave signals are hardly affected by clouds, dust, fog, wind and bad weather conditions. However, as shown in Zebker et al. [1997], radar signals are delayed by atmospheric water vapor. Such signal propagation delays are observed both for Side-Looking Airborne Radar (SLAR) and for SAR. Due to the better spatial resolution and the possibility of using persistent scatterers on the ground, SAR is used in this work rather than SLAR.

Each SAR image consists of the two complex components amplitude and phase. The amplitude represents the brightness of the backscatterers, whereas the phase of each resolution cell is composed of a scattering phase and a component proportional to the distance traveled by the signal between the SAR satellite and the backscattering surface. When aiming at an analysis of the atmospheric effects on SAR, the scattering phase is eliminated by means of building interferograms. When compared to SAR, InSAR always considers the interferometric difference between two acquisitions. This difference can be either spatial (single pass InSAR) or, as in this work, temporal (repeat pass InSAR).

Yet, only in the case of constant backscattering characteristics of a resolution cell, the scattering phase of a resolution cell cancels out when building complex interferograms by multiplying the first Single Look Complex (SLC) SAR image with the conjugate complex of the second SLC SAR image. In order to obtain constant backscattering characteristics within the analyzed resolution cells, a PSI processing considering only coherent scatterers is performed. Thus, for each persistent scatterer within the interferogram, an interferometric phase is obtained. Within this interferometric phase, the contributions of topography, displacement, the neutrosphere and the ionosphere, the orbits, the Earth's reference phase, and noise are superposed. In order to separate the InSAR neutrospheric phase from the other phase components, a collection of spatio-temporal filtering routines is performed, as explained e.g. in Alshawaf [2013]. Figure 2.8 summarizes in more detail how, for each PS point, absolute InSAR ZWD estimates are obtained from short-wavelength InSAR ZWD differences, GNSS ZWDs, and surface meteorological data.



**Figure 2.8:** Generation of absolute InSAR ZWD estimates based on short-scale neutrospheric phase differences from InSAR as well as additional information form GNSS ZWDs and surface meteorological data, as proposed in Alshawaf et al. [2015a] and Alshawaf et al. [2015c]. The elevation-dependent, long-wavelength, and short-wavelength components of the neutrospheric signal are illustrated according to Alshawaf [2013]. As the different components of the water vapor distribution are introduced in terms of IWVs in Alshawaf [2013], the IWV in mm is plotted in this figure instead of ZWDs.

Per definition, InSAR phases are relative measurements given, in the case of a temporal baseline between the two acquisitions, as a temporal difference between two dates. Therefore, Alshawaf et al. [2015a] analyzed the differences between GNSS and InSAR estimates of the neutrospheric delays. They point out that GNSS and InSAR show complementary features. As shown by Alshawaf [2013], the neutrospheric phase differences contained in

the SAR interferograms correspond to the short-wavelength component of the wet delay experienced by radio wave signals, while GNSS delays contain the long-wavelength and the elevation dependent component of the neutrospheric delay. By means of

$$\Delta ZWD = -\frac{4 \cdot \pi}{\lambda_{\text{radar}}} \cdot \Phi_{\text{neu}} \cdot \cos \theta_{\text{inc}} \quad (2.28)$$

from Alshawaf [2013], the InSAR phase differences  $\Phi_{\text{neu}}$  can be transformed into short-scale ZWD differences  $\Delta ZWD$ . The quantity  $\lambda_{\text{radar}}$  in Equation 2.28 corresponds to the wavelength of the SAR system used in this work, and  $\theta_{\text{inc}}$  is the incidence angle. In the case of the Envisat SAR system, the incidence angle corresponds to about  $23^\circ$ . Assuming the average neutrospheric phases over the whole stack to equal zero e.g. according to the research of Rocken and Braun [2000], short-scale InSAR neutrospheric phases per acquisition time are deduced. As shown in Alshawaf [2013], the remaining two components of the total neutrospheric delay, i.e. the long-wavelength component as well as the elevation-dependent component, can be modeled based on additional information from GNSS and from surface meteorology. Adding these two components to the short-scale delay from InSAR yields an absolute wet delay value at each PS point. In the following, the term InSAR ZWD will stand for the described, absolute wet delays from the combination of InSAR, GNSS, and surface meteorology.

## 2.7 Summary

This chapter gives an overview on common subdivisions of the atmosphere in order to well classify this work on water vapor tomography focusing on the electrically neutral lower  $\approx 10$ km of the atmosphere. Various measurement techniques contributing to the quantification of the neutrospheric water vapor are presented and basic knowledge on numerical weather prediction is given. Moreover, the path delay modeling for the measurement techniques GNSS and InSAR used in this research is detailed so as to prepare the reader to i) notice the differences between this work and previous research when reading the state of the art in Section 3 and to ii) understand how SWD observations can be deduced from GNSS and InSAR, when considering Chapter 6 that explains the generation of the real GNSS and InSAR data set used in this work.

## 3 State of the art in tomographic water vapor reconstruction

Current approaches to tomographic water vapor reconstruction can be distinguished e.g. by means of the methodology and the data sets used for the solution of the tomographic model. The different solution strategies e.g. vary in their ways of regularizing the tomographic problem. Much research has been carried out on regularizing the tomographic reconstruction by means of constraints or prior knowledge. In addition, the effect of the observing geometry on the quality of the tomographic reconstruction has been analyzed within current research on tomographic water vapor reconstruction. Depending on the geometric settings and the chosen solution methodology, the uncertainty of the wet refractivity reconstruction varies. In the following, each of the described components of the domain of tomographic water vapor reconstruction is explained in more detail.

### **Methodologies:**

Current methodologies for the solution of the tomographic model based on space-geodetic SWD observations are mainly based on a discretization of the analyzed atmospheric volume into voxels. The existing methodologies applied to such a discretized atmosphere can be distinguished into iterative and non-iterative techniques. While Bender et al. [2011] apply different iterative Algebraic Reconstruction Techniques (ART), Champollion et al. [2004], Flores et al. [2000], Hirahara [2000], Notarpietro et al. [2008], Troller [2004], Song et al. [2006], and Rohm [2013] propose different non-iterative approaches to solving the tomographic system by means of LSQ. As ART algorithms iteratively process observation by observation, no matrix inversion is necessary, and thus, ART algorithms are particularly profitable for large data sets. Due to its good capability to estimate dynamically changing parameters, Flores et al. [2000] and Gradinarsky and Jarlemark [2004] propose a Kalman filter approach. Hirahara [2000] applies a damped least squares adjustment known from seismic tomography to the tomographic system. Xia et al. [2013] present a combination of iterative and non-iterative techniques. They firstly apply iterative reconstruction algorithms in order to determine a refractivity field that is then introduced as initial value for a non-iterative tomography approach. Adavi and Mashhadi-Hossainali [2015] use a hybrid regularization method based on the LSQ QR and Tikhonov regularization techniques deriving a regularized solution independent of initial values. In their approach, the water vapor content is only estimated for voxels that are crossed by enough rays. The refractivities of voxels which are not adequately constrained are averaged based on the refractivities estimated in the surrounding voxels.

While most studies on tomographic water vapor reconstruction are mainly based on slant wet delay estimates from GNSS, Hurter and Maier [2013] include both slant wet delays from GNSS as well as wet refractivity profiles from radio occultation and radiosonde measurements into a combined least squares collocation. Instead of using slant wet delay observations, Nilsson and Elgered [2007] propose an approach relying directly on GPS phase observations. In their approach, the resulting ill-posed system of equations is regularized by relating the refractivity at different points according to atmospheric turbulence theory and by using a Kalman filter modeling the variations in time as related to the wind speed.

### **Regularization:**

Independently of the reconstruction strategy and of the introduced data sets, the ill-posed tomographic system of equations can be regularized i) by introducing constraint equations to the tomographic system, which can be considered as pseudo observations, ii) by including additional observations from other sensors, from models, or from simulations, or iii) by increasing the percentage of voxels crossed by rays. An increased percentage of voxels crossed by rays can be obtained e.g. by adapting the voxel sizes to the ray density. Alternatively, instead of using

only rays entering the volume on its top, both rays entering the study area on its top and on its side can be introduced into the tomographic system. In addition, as stated by Rohm [2013], when considering a general inverse approach based on singular value decomposition, the inverse system can be stabilized by carefully selecting the meaningful singular values. The following gives more details on the state of the art in regularizing the tomographic system of equations.

i) Constraints:

In order to regularize the solution by means of constraints, both Flores et al. [2000] and Gradinarsky and Jarlemark [2004] add horizontal and vertical smoothing constraints to the tomographic system and introduce a boundary constraint assuming zero refractivity above a certain height. E.g. in Flores et al. [2000], the horizontal constraints are implemented by means of additional observation equations imposing the refractivity in a voxel to be a weighted average of its neighbors' refractivities. In contrast, Song et al. [2006] generate horizontal smoothing constraints by assuming a certain degree of correlation between neighboring voxels using Gaussian weighted mean with controllable width. According to Gradinarsky and Jarlemark [2004], this Gaussian weighted mean can also be used for the implementation of vertical constraints. Alternatively, as proposed in Elosegui et al. [1998], the refractivity decay with increasing height can be assumed to follow an exponential law. Rohm [2013] presents an unconstrained approach for water vapor tomography, which is based on a combination of consecutive epochs of data. The work in Heublein et al. [2018] uses the sparsity of the signal in a specific, predefined transform domain as a prior for regularization and then reconstructs the signal by means of  $L_1$ -norm minimization. While helping a lot in regularizing the solution, both geometric constraints and exponential decay in most cases do not reflect the real atmospheric state.

ii) Prior knowledge:

In addition to the constraints, the ill-posed tomographic system can be regularized by means of prior knowledge. While Flores et al. [2000] add radiosonde profiles as profile-wise a priori knowledge, Champollion et al. [2004] use a standard atmosphere as a priori field. Champollion et al. [2004] and Xia et al. [2013] introduce water vapor profiles above 2 km from radio occultation into the tomographic system, that are e.g. obtained from the Constellation Observing System for Meteorology, Ionosphere, and Climate (COSMIC). In addition, Champollion et al. [2004] suggest the inclusion of surface meteorological observations for the regularization of the refractivity estimates in the lowest layer. Besides, Song et al. [2006] use a priori knowledge from numerical weather prediction. According to Chen and Liu [2016], water vapor radiometers and sun photometers can also yield observations that can regularize the tomographic system. With the aim of minimizing the smoothing effects of geometrical constraints, Benevides et al. [2016] include InSAR SWD difference maps as a constraint to GNSS tomography.

iii) Reduction of the number of voxels without crossing rays:

Besides introducing constraints and prior knowledge for regularization of the tomographic system, Yao et al. [2016], Yao and Zhao [2016] and Yao and Zhao [2017] suggest approaches to regularize the tomographic solution by means of a reduction of the number of voxels without crossing rays. Using the support of radiosonde data, Yao et al. [2016] increase the number of crossed voxels by including rays leaving the test region at the side. According to Yao and Zhao [2016], the utilization rate of SWD observations can not only be increased by the use of rays entering the study area both on its top and on its side, but also by a careful selection of the upper boundary of the analyzed atmospheric volume. First of all, Yao and Zhao [2016] select a reasonable vertical tomography boundary based on several years of radiosonde observations. Then, they suggest a two-step refractivity estimation that optimally uses GNSS rays entering the study area both on its top and on its side. In the first step, they define a study area larger than the analyzed study area and estimate the refractivities of this study area by using only the rays entering the area on its top. In the second step, they reduce the study area to the size of the final tomographic grid. In this step, the refractivities determined within the larger study area are used in order to compute a scale factor describing, for each ray, the ratio of SWD situated within or outside of the study area. This scale factor is then applied for reducing the total SWDs of side rays to the portion of SWDs corresponding to the tomographic grid. The reduced side ray SWDs can then be appended to the observation equation. In spite of the increased number of voxels



passed by rays in Yao and Zhao [2016], horizontal smoothing constraints and vertical a priori conditions remain necessary for the solution of their tomographic system. Yao and Zhao [2017] present a tomography approach that is based on a non-uniform symmetrical division of the horizontal voxels, such that the available information is distributed more evenly among all voxels than in the case of regular voxel divisions.

A similar idea of decreasing the number of voxels without crossing rays is pursued by Rohm [2013]. Yet, in contrast to Yao and Zhao [2016] and Yao and Zhao [2017] focusing on geometric aspects of the tomographic model, Rohm [2013] focuses on introducing a large amount of data, namely a combination of consecutive epochs of synthetic observations of at least three interoperable GNSS into a synthetic data set. As long as the consecutive epochs of observations are linked with one state of the atmosphere, Rohm [2013] combines them within an unconstrained approach for water vapor tomography. As SWD estimates from various GNSS are included, the number of crossed voxels is automatically increased. Yet, due to the cone shaped GNSS observing geometry, the ray geometry remains limited, and there remain voxels that are not crossed by any rays at all. As a consequence, many voxels are still underdetermined and the tomographic system needs to be regularized. Rohm [2013] solves the tomographic system by means of computing a general inverse. I.e. the system is regularized by means of a careful selection of the singular values used for the computation of the general inverse.

#### **Vertical resolvability:**

Besides the ill-posedness of the mathematical model, one of the main challenges encountered in GNSS water vapor tomographies consists in the vertical resolvability of the 3D wet refractivity field. According to Flores et al. [2000], the interchange of two different layers in a refractivity field that does not have any horizontal variation would lead to the same integrated delay measurements and therefore, a distinction between the refractivities of the two layers would not be possible. This challenge has to be faced in particular if no rays entering the study area on its side are included within the tomographic model. In addition to the interchangeability of different height layers, Perler et al. [2011] state that their tomographic approach has a limited ability to resolve vertical structures above the highest GNSS site of the used network. Troller et al. [2006] recommend large height differences between the GPS stations for better vertical resolvability and for an accurate tomographic solution.

#### **Including InSAR SWDs into water vapor tomography:**

Including InSAR SWD difference maps as a special case of additional data into the tomographic system as presented in Benevides et al. [2016] reduces the smoothing effects of geometric constraints. Yet, the approach proposed in Benevides et al. [2016] considers temporal changes of PW only. Moreover, they do not carefully distinguish between the different components of the precipitable water. According to Alshawaf et al. [2015c], the PW is composed of a stratified (elevation-dependent) component, a turbulently mixed short-scale component, and a long-wavelength component. Parts of the elevation-dependent component as well as the long-wavelength PW risk to be canceled out within the InSAR processing. These drawbacks of the InSAR processing for water vapor analyses are overcome in Alshawaf et al. [2015c] by combining PW values estimated at GNSS sites and PW difference maps extracted from SAR interferograms to maps of absolute PW at high spatial resolution. Alternatively, Alshawaf et al. [2015a] propose a data fusion of InSAR, GNSS, and simulations of the WRF model in order to derive PW maps. In this work, absolute PW maps from InSAR as described by Alshawaf et al. [2015c] are introduced into the tomographic system.

#### **Compressive Sensing in water vapor tomography:**

In addition to a classical LSQ approach as encountered in many previous studies on water vapor tomography, CS and sparse reconstruction are exploited in this work for 3D tomographic water vapor reconstruction. Initially proposed by Candès et al. [2006], Donoho [2006], Baraniuk [2007], and Candès and Wakin [2008] for the reconstruction of images or signals from a number of samples which is far smaller than the desired resolution or than required by the Nyquist rate, CS has been, in the meantime, applied to various remote sensing problems in which sparse signals occur. While Potter et al. [2010] and Alonso et al. [2010] present CS applications in SAR imaging, Aguilera et al. [2013], Budillon et al. [2011], Zhu and Bamler [2010] and Zhu and Bamler [2014] apply CS for SAR tomography. Alternatively, Prunte [2010] presents an approach to CS-based ground moving target identification, and Grohnfeldt et al. [2013], Jiang et al. [2014], Li and Yang [2011], Zhu et al. [2016], and Zhu

and Bamler [2013] apply CS for pan-sharpening and hyperspectral image enhancement. Compared to classical LSQ adjustments, possibly including  $L_2$ -norm regularization, Compressive Sensing and sparse reconstruction led to promising results. For general applications of CS, Rauhut [2010] states that involving randomness in the acquisition step enables one to utilize the minimum number of measurements. When considering image reconstructions based on frequency data, e.g. Candès et al. [2006] advise to randomly measure frequency coefficients in order to sense sparse objects by taking as few measurements as possible.

### **Voxel discretization:**

Most approaches to tomographic water vapor reconstruction discretize the study area into voxels of constant inner-voxel refractivity. Perler et al. [2011] parameterize the inner-voxel refractivities by means of trilinear and spline functions in ellipsoidal coordinates, yielding a substantially smaller maximal error than in non-parameterized voxels. They state that the parameterization of the voxels reduces the effects of discretization while increasing only negligibly the computational costs. Moreover, Perler et al. [2011] point out that the interpolation of in situ measurements to voxel central points becomes obsolete in the case of parameterized voxels. Instead of discretizing the analyzed atmospheric volume into voxels, Ding et al. [2018] propose a node parameterization approach using a combination of three meshing techniques to dynamically adjust both the boundary of the tomographic region and the position of nodes at each tomographic epoch. Zhao et al. [2018] try to reconstruct the water vapor field without imposing horizontal and vertical constraints referring to a voxel grid. They do not discretize the study area in the horizontal direction, but only distinguish different vertical layers. A horizontal parameterized quadratic function is introduced for each layer to describe the continuous water vapor change within that layer. I.e. their approach is, in some degree, similar to the approach of Heublein et al. [2018] parameterizing the wet refractivity field by iDCT, Euler, and Dirac base functions. However, in contrast to Heublein et al. [2018] performing the parameter estimation in a transform domain and thus benefiting of the sparsity of the refractivity field in that transform domain, Zhao et al. [2018] directly use the quadratic parameterization for the refractivity estimation. While Zhao et al. [2018] avoid a discretization of the study area in the horizontal direction, Heublein et al. [2018] still distinguish voxels horizontally.

In case of a discretization of the analyzed refractivity field into voxels, depending on the number and on the distribution of the available GNSS sites, different voxel sizes have been used in previous research. A decrease of the spatial resolution increases the discretization error. In turn, an increase of the spatial resolution makes the accurate estimation of the inner-voxel refractivity more challenging. Hirahara [2000], Gradinarsky and Jarlemark [2002], Nilsson and Elgered [2007], Champollion et al. [2009], and Yao and Zhao [2017] discretize the refractivity field into non-uniform voxels with horizontal resolutions between a few and some tens of kilometers. Yao and Zhao [2017] propose a non-uniform symmetrical division of horizontal voxels in order to increase the number of voxels crossed by rays. In contrast, most other approaches discretizing the refractivity field into voxels horizontally define uniform voxel sizes, i.e. voxels with constant sizes in longitude and latitude within the whole study area. Nilsson et al. [2005] and Nilsson and Elgered [2007] suggest to use small horizontal voxel sizes of a maximum of 10 km in order to minimize the voxel discretization error. According to Chen and Liu [2014], the optimal voxel resolution for a specific region should be determined before conducting water vapor tomography for that region. Therefore, they optimize the horizontal boundaries of their study area by moving in  $0.001^\circ$  large steps in longitude and latitude and maximizing the number of ray-crossing voxels. However, as this approach needs to relocate the tomographic area before every tomography experiment, the approach is hard to operate in practice. Notarpietro et al. [2011] also define the voxel grid based on several tests. They state that voxel grid rotations of  $\pm 15^\circ$  w.r.t. the North-South resp. the East-West direction produced changing tomography results. Champollion et al. [2004] state that the optimal horizontal size of a voxel should be equal to the mean inter-site distance between the available GNSS sites. In their approach, the thickness of the layers should be such that rays starting from a GNSS site situated at the center of a voxel are able to cross neighboring voxels, given a certain cutoff elevation angle. Rohm [2012] advises to use increasing height layer thicknesses with increasing altitude, because the wet refractivities in the upper layers are very small and the tomographic solutions may thus be sensitive to errors in the input data. While e.g. Troller et al. [2006], Perler et al. [2011], Chen and Liu [2014], and Ding et al. [2017] use non-uniform layer thicknesses, e.g. Flores et al. [2000], Gradinarsky and Jarlemark [2002], Champollion et al. [2004], Notarpietro et al. [2011], Bender et al. [2011], Xia et al. [2013], and Zhao et al. [2018] decided for uniform layer thicknesses. Table 3.1

gives the mean number of GNSS sites per  $100\text{km}^2$  within the tomographic approaches of different research groups. Depending on the network, the site densities range from less than one site in mesoscale tomography approaches like Bi et al. [2006], Perler et al. [2011], or Heublein et al. [2018] to about 20 sites per  $100\text{km}^2$  in a special water vapor tomography campaign at the Onsala Space Observatory described in Flores et al. [2001]. In this work, the site density corresponds to that of Heublein et al. [2018] and is comparably low w.r.t. most of the other tomographic approaches shown in Table 3.1. Therefore, the selection of an appropriate voxel size is of particular importance in this work, and the effect of the voxel discretization on the tomographic results will be analyzed.

**Table 3.1:** GNSS site densities encountered in the selected tomographic approaches discretizing the refractivity field into horizontally uniform voxels.

approach	number of sites	size of study area	horizontal voxel number	sites / $100\text{km}^2$
Flores et al. [2000]	16	$400\text{km}^2$	$4 \times 4$	4.0
Flores et al. [2001]	7	$0.05^\circ \times 0.1^\circ$	$4 \times 4$	20.8
Champollion et al. [2004]	17	$20 \times 20\text{km}^2$	$3 \times 5$	4.3
Bi et al. [2006]	4	$100 \times 100\text{km}^2$	$5 \times 6$	< 0.1
Troller et al. [2006]	73	$300 \times 150\text{km}^2$	$6 \times 3$	0.2
Song et al. [2006]	14	$60 \times 60\text{km}^2$	$4 \times 4$	0.4
Nilsson and Elgered [2007]	17	$20 \times 20\text{km}^2$	$3 \times 5$	4.3
Perler et al. [2011]	46	$3^\circ \times 1.5^\circ$	$6 \times 3$	0.2
Notarpietro et al. [2011]	11	$18 \times 26\text{km}^2$	$4 \times 4$	2.4
Heublein et al. [2018]	8	$117 \times 122\text{km}^2$	$5 \times 5$	< 0.1

Besides others, the uncertainty of the tomographic reconstruction depends on the aspects mentioned above, i.e. on the solution strategy, the sizes of possible discrete voxels, the type and the size of the included SWD data sets, the type of constraints applied for regularization, the characteristics of the possibly introduced prior knowledge, and the number of rays per voxel. Rohm [2012] asserts that he takes all factors influencing the final uncertainty into account, mainly the uncertainties of ZWD, pressure and temperature, antenna phase center height, satellite elevation angle, mapping function, and the tomography solution. He realizes that the mathematical properties of the design matrix largely influence the uncertainty of the tomographic solution. Yet, he does not discuss the influence of the horizontal distribution of the GNSS sites within the study area, the prevailing orbit geometry or the voxel discretization that he selected. However, in case of a campaign or the installation of new permanent sites, a knowledge on a good site distribution for water vapor tomography applications may be important. In addition, when developing new solution methodologies to the tomographic equation, the effect of the orbit geometry and of the voxel discretization should be analyzed.

#### **Relation of the presented approach to water vapor tomography and the state of the art:**

Similarly to the LSQ approaches described in the state of the art, this work presents an approach to water vapor tomography relying on LSQ. Yet, in addition, a CS solution strategy is developed. This CS approach to water vapor tomography does not include any geometric smoothing constraints as applied in common LSQ approaches. Instead, the CS solution benefits of the sparsity of the parameters in an appropriate transform domain as a prior for regularization. The proposed CS approach is compared to a LSQ approach regularized by means of a horizontal smoothing constraint and a vertical constraint assuming the water vapor content to exponentially decrease with increasing height. Prior knowledge from surface meteorology is included in both the CS and the LSQ solution. In order to increase the number of voxels crossed by rays, InSAR SWDs are added to the GNSS SWDs commonly introduced into the tomographic system. Finally, as the rule of thumb of Champollion et al. [2004] suggesting horizontal voxel sizes corresponding to the mean inter-site distance between the available GNSS sites is only valid for LSQ, this work investigates the effect of the observing geometry (e.g. satellite position, GNSS site distribution, and voxel discretization) on the quality of the tomographic results.

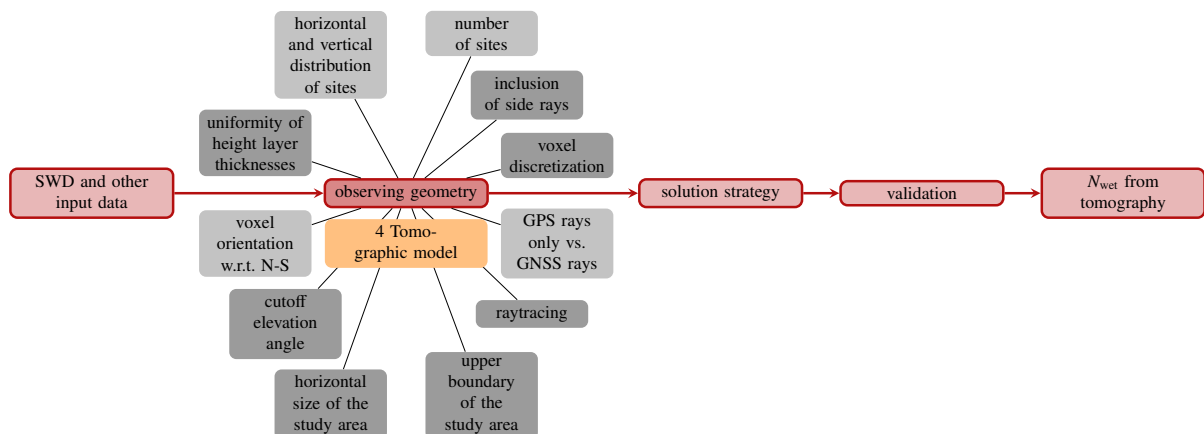


## 4 Tomographic model

The word *tomography* originates in the Ancient Greek *tomos*, ‘slice’, ‘section’, and *grafēin*, ‘to write’. This etymology already gives an idea of the basic principle of a *tomographic reconstruction*. A *tomographic reconstruction* consists in deriving the inner structure of some object, e.g. of the human body or of the Earth’s atmosphere, based on cross-sections of that object observed along different viewing angles.

E.g. this work aims at reconstructing the 3D water vapor field based on profile-wise observations of the slant wet delays at different directions. The radio wave signals originating e.g. from GNSS satellites and received at GNSS sites are integrated along their raypath. In order to derive the 3D water vapor distribution based on these integrated slant wet delays, a *tomographic model* needs to be defined.

*Tomographic models* in water vapor tomography commonly subdivide the refractivity field into discrete voxels of a certain size. In addition, based on the voxel definition and a raytracing of the GNSS rays within these voxels, a functional model relating the 3D refractivity field to the observed slant wet delays is established (Section 4.1). As described in Section 4.2, both rays entering the analyzed atmospheric volume on its top resp. on its side may be included into the tomographic system. As emphasized in Section 4.3, the careful definition of the voxel discretization within the tomographic model is important because the voxel discretization significantly effects both the spatial resolution and the quality of the reconstructed refractivities. Section 4.4 describes the raytracing approach yielding, in this work, the distances passed by straight GNSS rays within voxels defined by ellipsoidal upper and lower boundaries as well as planes of constant longitudes and constant latitudes. I.e., as illustrated in Figure 4.1, the tomographic model including the voxel discretization and the raytracing within the defined voxels is required in order to link the 1D slant wet delay observations and the 3D refractivity parameters in the functional model.



**Figure 4.1:** The tomographic model links the 1D slant wet delay observations and the 3D refractivity parameters. Besides the voxel discretization and the raytracing within the defined voxels, the tomographic model e.g. needs to define the horizontal size and the upper boundary of the study area as well as the cutoff elevation angle determining down to which elevation the GNSS rays are included into the tomography. The highlighted aspects of the tomographic model are explained in more detail in this section.

## 4.1 Tomographic equation within a discretized refractivity field

In February 2000, Flores et al. [2000] were the first to publish research on neutrospheric tomography using GNSS SWDs. They introduce the functional model

$$\text{SWD}_{i,\text{cont}} = 10^{-6} \cdot \int_{\text{sp}_i} N_{\text{wet}} dL, \quad (4.1)$$

already given in Equation 2.10, where  $\text{SWD}_{i,\text{cont}}$  are the slant wet delay observations integrated from a certain GNSS site to a certain satellite. The variable  $\text{sp}_i$  stands for the  $i$ th slant raypath, i.e. for a slant raypath of the radiowave signal between a certain receiver and a certain satellite, and the variable  $N_{\text{wet}}$  is the wet refractivity along this path. The index  $i$  is defined as

$$i \in \mathbb{N} \text{ with } 1 \leq i \leq N, \quad (4.2)$$

where  $N$  corresponds to the number of observations available within this work between any receiver and any satellite. When aiming at a tomographic reconstruction of the wet refractivity, however, the continuous functional model from Equation 4.1 is commonly replaced by a discrete functional model

$$\text{SWD}_{i,\text{disc}} = 10^{-6} \cdot \sum_{j=1}^L N_{\text{wet } j} \cdot d_{ij}, \quad (4.3)$$

see Section 2.2. That is, the 3D water vapor field is subdivided into  $L = \mathcal{P} \times \mathcal{Q} \times \mathcal{K}$  voxels in longitude, latitude, and height, in which the refractivity is e.g. assumed to be constant. In this work, a uniform voxel discretization is performed in the horizontal directions, whereas the voxel sizes in the vertical direction may increase with increasing height. Horizontally, the voxels are limited by constant longitudes and latitudes. In the vertical direction, the voxels dispose of ellipsoidal upper and lower boundaries. The numbers and the sizes of the voxels are essential for the quality of the tomographic reconstruction. Therefore, they are specified for each of the different settings analyzed in Section 7. Summarizing all observations  $\text{SWD}_{i,\text{disc}}$  in an observation vector  $\text{SWD} = \mathbf{y} \in \mathbb{R}^{N \times 1}$ , all unknown refractivities  $N_{\text{wet } j}$  for

$$j \in \mathbb{N} \text{ with } 1 \leq j \leq L \quad (4.4)$$

in a parameter vector  $\mathbf{x} \in \mathbb{R}^{L \times 1}$ , and all distances  $d_{ij}$  in a design matrix  $\Phi \in \mathbb{R}^{N \times L}$ , the linear system of equations from Equation 4.3 can be reformulated in the form

$$\mathbf{y} = \Phi \cdot \mathbf{x} \quad (4.5)$$

with

$$\Phi(i, j) = \begin{cases} d_{ij} & \text{if ray } i \text{ crosses voxel } j \\ 0 & \text{otherwise.} \end{cases} \quad (4.6)$$

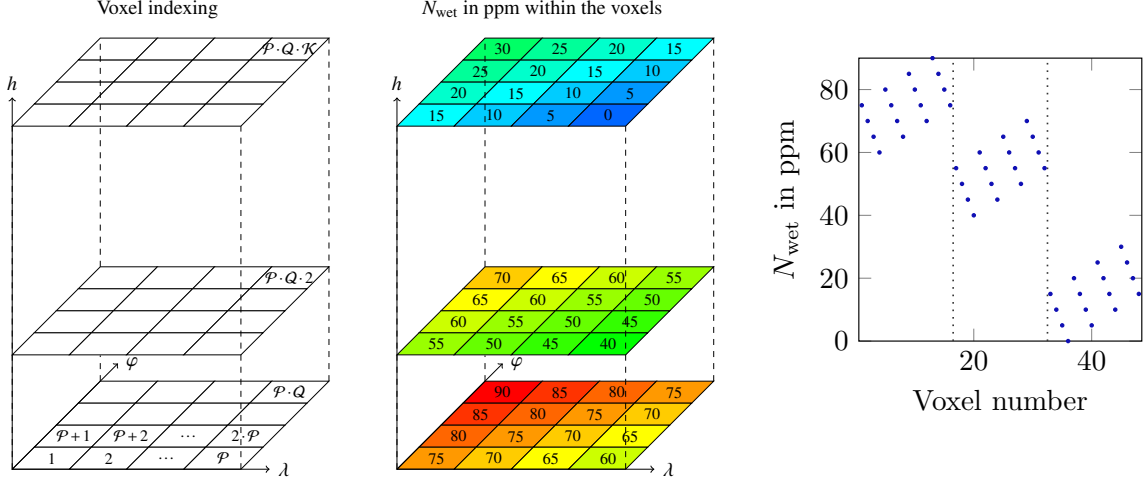
Figure 2.4 in Section 2.2 illustrates the distances passed by a ray within the voxel grid. As each ray only crosses a small subsection of the voxel grid, the matrix  $\Phi$  contains many zero elements and just a few non-zero elements (e.g. only about 4.5% of the entries of  $\Phi$  are non-zero in the case of  $5 \times 5 \times 5$  voxels, seven GNSS sites, and ten rays per site). Voxels that are not crossed by any rays yield a zero column in  $\Phi$ .

Figure 4.2 shows how the voxels are numbered within the tomographic grid. The voxel numbering is performed layer-wise, starting from the lowest height layer. Within each layer, the voxel numbering starts at the voxel with the smallest longitude and latitude and then runs row-wise through the considered height layer, firstly along increasing longitudes  $\lambda$ , then along increasing latitudes  $\varphi$ . Each element of the 3D parameter vector  $\mathbf{x}_{3\text{D}}$  is related to a distinct entry of the 1D parameter vector  $\mathbf{x}$  by means of the vec operator:

$$\mathbf{x} = \text{vec}(\mathbf{x}_{3\text{D}}) \quad (4.7)$$

The vec operator converts the 3D parameter vector  $\mathbf{x}_{3\text{D}}$  into a 1D column vector by means of stacking the elements of  $\mathbf{x}_{3\text{D}}$  on top of one another:

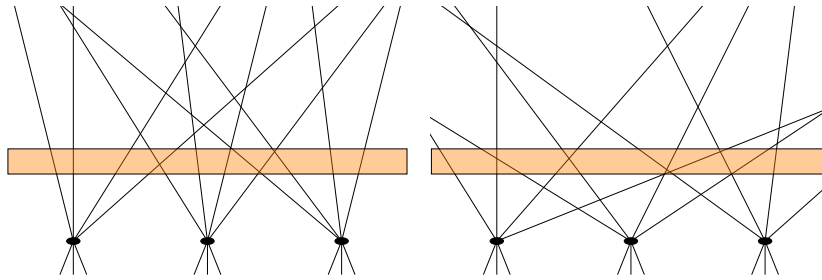
$$\text{vec}(\mathbf{x}_{3\text{D}}) = \left[ x_{3\text{D}(1,1,1)}, \dots, x_{3\text{D}(\mathcal{P},1,1)}, \dots, x_{3\text{D}(\mathcal{P},\mathcal{Q},1)}, \dots, x_{3\text{D}(\mathcal{P},\mathcal{Q},\mathcal{K})} \right] \quad (4.8)$$



**Figure 4.2:** Schematic representation of the voxel numbering within the tomographic model. The left plot exemplarily shows the indices of the  $L = \mathcal{P} \times \mathcal{Q} \times \mathcal{K}$  voxels in longitude  $\lambda$ , latitude  $\varphi$ , and height  $h$ . The middle plot shows refractivities that linearly decrease resp. increase with increasing longitude resp. latitude and that decrease with increasing height within a  $\mathcal{P} \times \mathcal{Q} \times \mathcal{K} = 4 \times 4 \times 3$  voxel discretization. While the numbers within the left plot stand for the indices  $[1, L]$  of the respective voxels, the numbers in the middle plot correspond to the refractivities within the respective voxels, that are also indicated by the color coding. Considering both the left and the middle plot of the figure shall help to understand how the refractivities of the respective height layers can be represented w.r.t. their voxel numbers in the two dimensional (2D) plot on the right, in which dashed gray lines separate the height layers.

## 4.2 Side rays for better vertical resolution

If all rays integrate the whole neutrospheric water vapor between the top and the bottom of the study area, as illustrated in the left plot in Figure 4.3, whole layer disturbances (e.g. in the case of inversion weather) can be hardly reconstructed. This also means that the vertical refractivity decay cannot be well reconstructed based on the input SWD data sets. If no side rays are and no flat-angle rays from GNSS radio occultation (GNSS RO) are included, the vertical water vapor distribution is mainly determined in a model-based way. Any horizontal refractivity layers could be interchanged without a change in the SWD estimates.



**Figure 4.3:** Left: whole layer disturbance without side rays, right: whole layer disturbance with side rays

Possible ways to model the portion of the side rays situated outside the study area could be i) approximating the missing part of the ray by means of radiosonde profiles, ii) computing the portion of the ray outside the study area using NWMs, iii) assuming the refractivity to decay exponentially with increasing height, or iv) solving the inverse system in a two-step estimation as described in Yao and Zhao [2016]. Without side rays, they firstly deduce a pure top ray solution from a larger study area than the area of interest, in order to then reconstruct the wet refractivity field in the area of interest by means of including side rays and by means of computing the portion of these side rays situated outside of the area of interest based on the top ray solution obtained within the larger study area.

### 4.3 Effects of the voxel discretization

When analyzing the effect of discretizing the refractivity fields into voxels of constant refractivity, different aspects need to be discussed. First of all, the assumption that the inner-voxel refractivity is constant should be questioned. In addition, the relation between the functional model from Equation 4.5 and the chosen voxel discretization has to be analyzed. The description of the observations  $\mathbf{y}$  as a product of the parameters  $\mathbf{x}$  and the design matrix  $\Phi$  within the forward model corresponds to a weighting of the voxel's refractivities by means of the distances passed within the voxels. Depending on the selected voxel sizes, the smoothing of the parameters induced by the voxel discretization varies. Finally, within the synthetic data set, both the precision of the observation vector  $\mathbf{y}$  and the precision of the refractivity values  $N_{\text{wet}}$  used for validation depend on the voxel discretization. Thus, the voxel discretizations used for determining the observation vector and the refractivity values for validation should be carefully selected.

In the following, firstly, the error committed by approximating constant inner-voxel refractivities is analyzed. This is done by comparing constant inner-voxel refractivities with exponentially decaying inner-voxel refractivities, assuming the exponential decay approximation to better represent the true atmospheric state than the approximation of constant inner-voxel refractivities. In the real, continuous atmosphere, the refractivity  $N_{\text{wet, const, cont}}$  within a voxel of constant refractivity, limited by lower and upper voxel boundaries  $h_k$  and  $h_{k+1}$  indexed by  $k$ , can be approximated by

$$N_{\text{wet, const, cont}}(h_k, h_{k+1}) = \frac{1}{h_{k+1} - h_k} \cdot \int_{h_k}^{h_{k+1}} N_{\text{wet}}(h_0) \cdot \exp\left(-\frac{h - h_0}{H_{\text{scale}}}\right) dh \quad (4.9)$$

with a scale height  $H_{\text{scale}}$  of the local neutrosphere and with  $h_0$  standing for some reference height with known  $N_{\text{wet}}(h_0)$ , e.g.  $h_0 = 0$  m at the surface, with  $N_{\text{wet}}(h_0)$  corresponding to the surface refractivity. In the discrete case, the exponential refractivity decay with height can be approximated by the Trapezoidal rule

$$N_{\text{wet, const, disc}}(h_k, \dots, h_{k+\text{nb\_h}}) = \frac{1}{h_{k+\text{nb\_h}} - h_k} \cdot \sum_{k=2}^{\text{nb\_h}} \frac{1}{2} \cdot [N_{\text{wet}}(h_{k-1}) + N_{\text{wet}}(h_k)] \cdot (h_k - h_{k-1}) \quad (4.10)$$

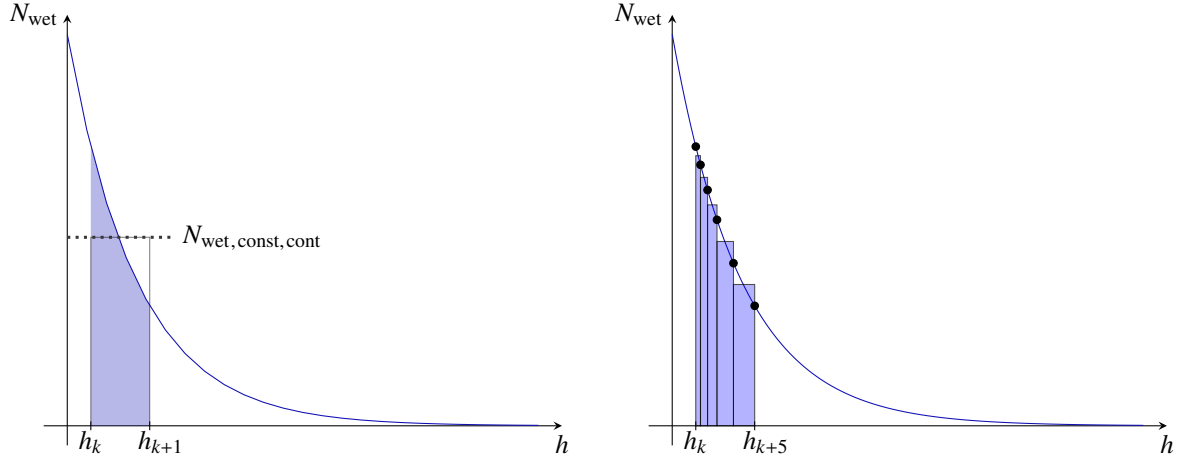
with  $\text{nb\_h}$  corresponding to the number of discrete refractivity values available along one vertical profile through the considered voxel between its lower and upper voxel boundaries  $h_k$  and  $h_{k+\text{nb\_h}}$ .

The left plot within Figure 4.4 illustrates the approximation of a voxel's refractivity to a constant value based on continuous refractivity information using Equation 4.9. The right plot in the same figure shows how a constant inner-voxel refractivity could be deduced from discrete refractivity information according to Equation 4.10. E.g. constant inner-voxel refractivities can be deduced from the discrete WRF refractivity information in order to generate a synthetic SWD data set as well as a validation data set for this synthetic data set. However, for the sake of simplicity, instead of using Equation 4.10, the constant inner-voxel refractivities in the synthetic data set described in Section 6.3.2 are determined by a simple averaging of the refractivities of all those WRF cells situated within the considered voxel. The deviation w.r.t. the discrete formulation from Equation 4.10 showed to be less than 0.1 ppm and was judged to be insignificant.

In order to illustrate the approximation errors generated by the assumption that the inner-voxel refractivity is constant, an exemplary setting with a surface refractivity of 41 ppm is defined. This value for the surface refractivity corresponds to the atmospheric behavior at the exemplarily chosen meteorological site 4444 (Karlsruhe Nordwest) on 2005-08-01 at 12 pm, with a temperature of 21.8°C, a pressure of 1006.00 hPa, and a dew point temperature of 8.8°C. In such a setting, the approximation error between a constant inner-voxel refractivity and refractivity values drawn from an exponential approximation to the refractivity decay corresponds to about 9.5 ppm resp. to about 7.2 ppm on the bottom resp. on the top of a voxel extending from, e.g. 500 m to 1750 m. In general, the maximum error between an exponential decay approximation and a constant inner-voxel refractivity approximation to such an exponential decay is equal to

$$\begin{aligned} \Delta N_{\text{wet, exp vs. disc}} &= N_{\text{wet}}(h_0) - N_{\text{wet, const, disc}}(h_0, \dots, h_1) \\ &= N_{\text{wet}}(h_0) - \frac{1}{h_1} \cdot N_{\text{wet}}(h_0) \cdot H_{\text{scale}} \cdot \left[ -\exp\left(-\frac{h_1}{H_{\text{scale}}}\right) + 1 \right] \end{aligned} \quad (4.11)$$

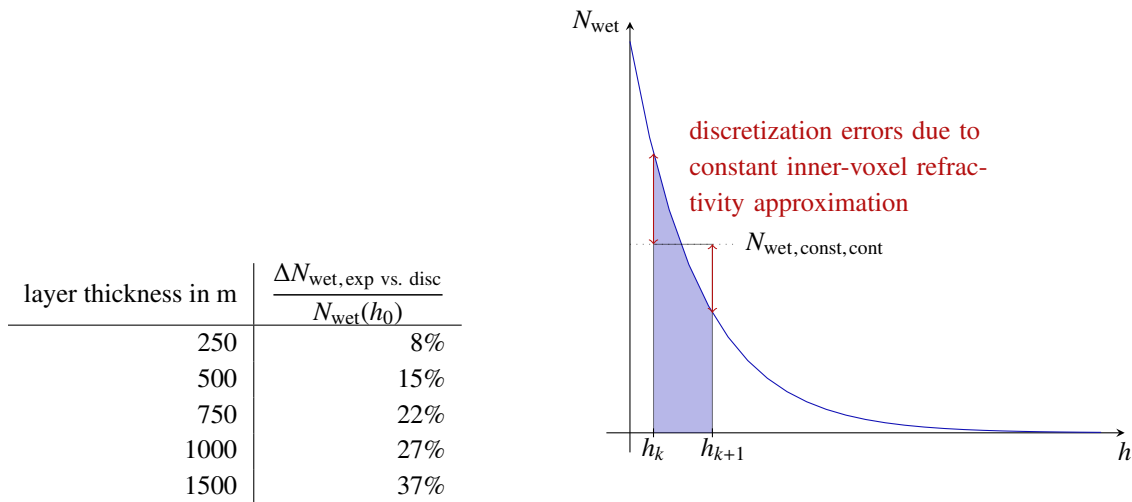




**Figure 4.4:** Schematic illustration of the approximation of constant inner-voxel refractivities from continuous (left) resp. discrete (right) refractivity information. In the left plot, the blue curve shows an exponential approximation to the refractivity decay with increasing height, whereas the top of the gray rectangle corresponds to the constant refractivity attributed to the voxel extending from  $h_k$  to  $h_{k+1}$ , i.e. to the value resulting from Equation 4.9. Thus, the area of the gray rectangle equals the integral of the exponential decay curve. In the right plot, the same exponential decay approximation is shown, but the constant refractivity value deduced for a certain voxel extending from the discrete height  $h_k$  to the discrete height  $h_{k+5}$  is deduced according to Equation 4.10, i.e. according to the Trapezoidal rule. In the case of constant inner-voxel refractivities deduced from WRF as in this work, the discretization steps are given by the discrete height levels of e.g. the WRF cell layers. The black dots schematically illustrate discrete refractivity values, e.g. taken from WRF.

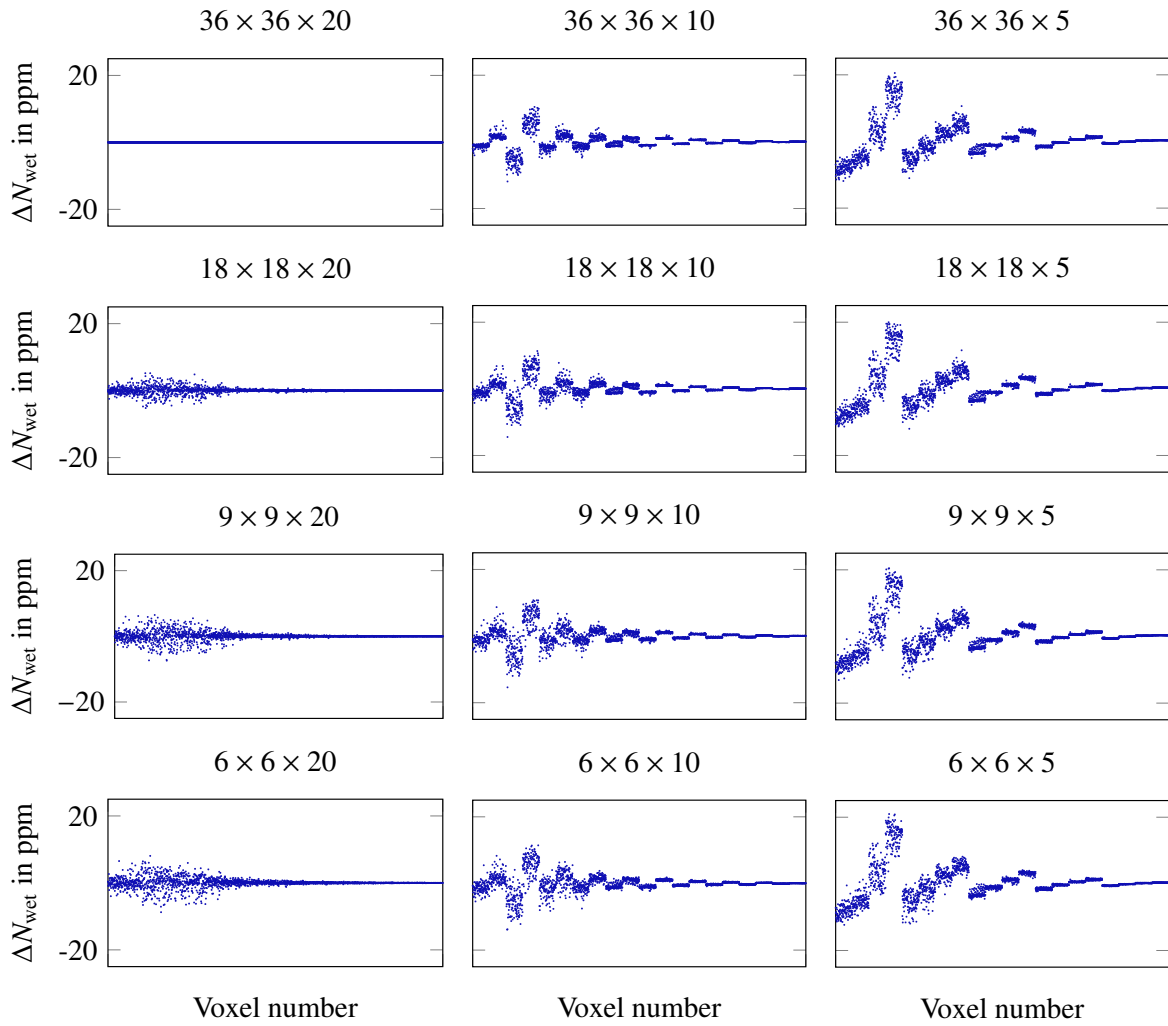
Assuming a scale height of  $H_{\text{scale}} = 1480\text{m}$ , Table 4.1 gives the ratio between the maximum error  $\Delta N_{\text{wet,exp vs. disc}}$  and the surface refractivity  $N_{\text{wet}}(h_0)$ . Figure 4.5 schematically shows the approximation error committed by assuming a constant inner-voxel refractivity instead of an exponential decay approximation to the refractivity.

Figure 4.6 shows refractivity differences between the refractivities deduced from WRF at a  $36 \times 36 \times 20$  voxel discretization and the refractivities deduced from WRF at different other, coarser voxel discretizations originating from merging neighboring small voxels of the  $36 \times 36 \times 20$  voxel discretization.



**Table 4.1:** Ratio between the maximum error  $\Delta N_{\text{wet,exp vs. disc}}$  and the surface refractivity  $N_{\text{wet}}(h_0)$ , given a scale height of  $H_{\text{scale}} = 1480\text{m}$ .

**Figure 4.5:** Approximation error committed by assuming a constant inner-voxel refractivity instead of an exponential decay approximation to the refractivity decay with increasing height. The approximation error depends on the height and on the thickness of the considered voxel.



**Figure 4.6:** Refractivity differences between refractivities deduced from WRF at different voxel discretizations and the refractivities deduced from WRF at a  $36 \times 36 \times 20$  voxel discretization. On the axis of abscissae, the voxel number within the  $36 \times 36 \times 20$  discretization is given, the axis of ordinates corresponds to the shown refractivity difference in ppm. The subfigure titles indicate the respective discretization of a  $100 \times 100 \text{ km}^2$  study area of 10 km height. The height layers are uniformly distributed, i.e. all height layers are equally thick. Assigning the  $36 \times 36 \times 20$  voxel indices of the finest discretization to the voxels in which they are situated within the coarser discretizations enables a comparison of the refractivity values at  $36 \times 36 \times 20$  voxel level. E.g. if the first voxel of the  $18 \times 18 \times 20$  voxel discretization contains the sub-voxels 1 to 18, 37 to 54, ..., and 613 to 630 of the  $36 \times 36 \times 20$  voxel discretization, then, the first voxel value of the  $18 \times 18 \times 20$  voxel discretization is assigned to all these  $36 \times 36 \times 20$  sub-voxels.

The plot shows that the horizontal and the vertical effect of the voxel discretization differ. If the difference between the refractivities in a reference  $36 \times 36 \times 20$  voxel discretization and the refractivities within a selected  $\mathcal{P} \times \mathcal{Q} \times \mathcal{K}$  voxel discretization is to be small, a fine voxel discretization in the vertical direction is more important than a fine horizontal voxel grid.

When analyzing the influence of the voxel discretization on the functional model

$$y = \Phi \cdot x \tag{4.12}$$

from Equation 4.5, the voxel discretization shows to have an effect on the smoothing of the refractivity parameters. The coarser the voxel grid, the smoother the refractivity parameters. On the one hand, when increasing the voxel

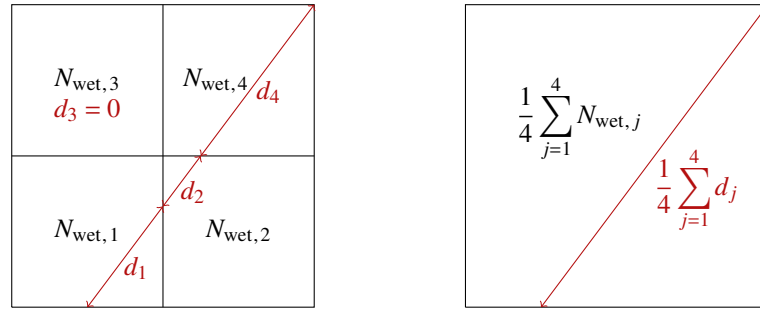
sizes from a fine to a coarse voxel grid by merging neighboring voxels to coarser voxels, the observations within the coarser grid are obtained by an unweighted average of the refractivities within the finer grid, i.e. the distances  $d_{ij}$  passed within the fine grid are simply summed up:

$$y_i = \frac{1}{4} \cdot \sum_{j=1}^4 N_{\text{wet},j} \cdot \sum_{j=1}^4 d_{ij} \quad (4.13)$$

However, on the other hand, the functional model is not consistent with such an unweighted average. As shown in Figure 4.7, the functional model weighs the different refractivities with respect to the distances passed within the voxels of the current voxel grid:

$$y_i = \sum_{j=1}^4 N_{\text{wet},j} \cdot d_{ij} \quad (4.14)$$

Therefore, the observations deduced based on different voxel discretizations differ in the synthetic data set. When taking into consideration the inverse model determining the parameters based on the observations, the functional model is only exact if the same voxel discretization is taken as a basis both for the computation of the observations and for the estimation of the parameters. I.e. in the real data set, where the observations correspond to the integrated wet delay along the signal path, the functional model would only be exact if infinitesimally small voxels were used for the tomographic reconstruction. Similarly, in the synthetic data set, the functional model can only describe the exact relation between the estimated parameters and the observations, if the observations are deduced from the same voxel discretization, i.e. at the same spatial resolution, at which the parameters are estimated.

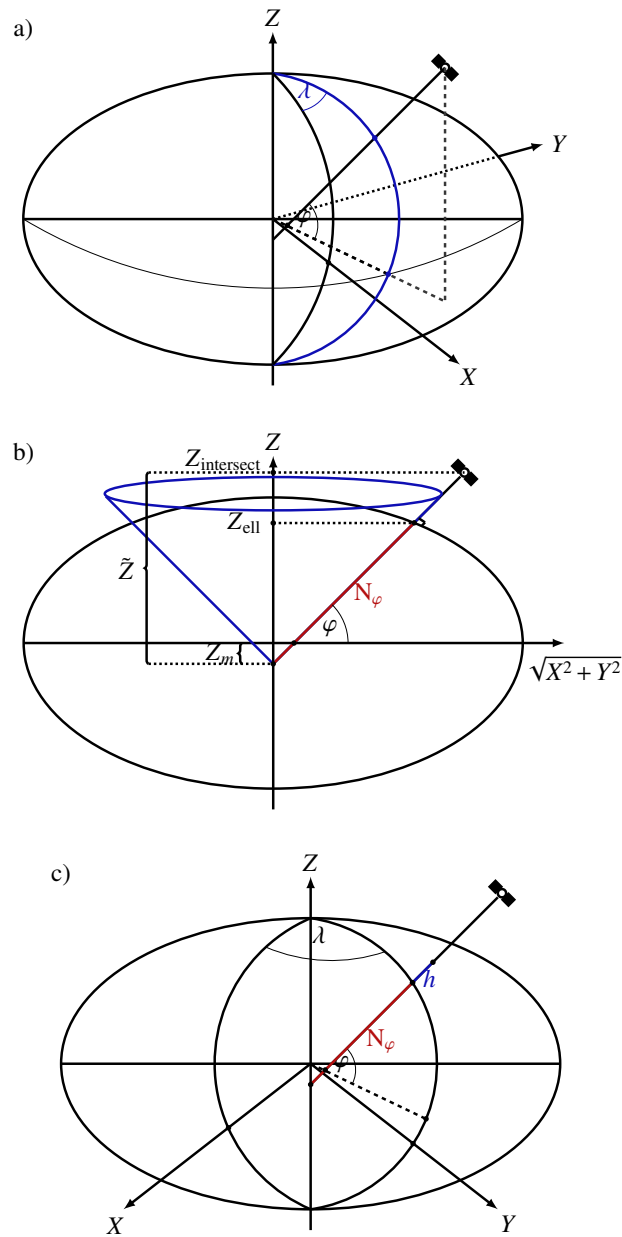


**Figure 4.7:** Relation of slant wet delay observations, refractivity parameters, and distances passed within the respective voxels within, on the left, a fine voxel grid, and on the right, a coarse voxel grid. In the case of the fine resp. coarse voxel grid, the observations are related to the parameters and to the distances according to Equation 4.14 resp. according to Equation 4.13

## 4.4 Raytracing on the ellipsoid

For this work, the raypaths from the satellites to the receivers are traced within the voxels described in Section 4.1. As these voxels are limited by ellipsoidal upper and lower boundaries, and as their sides are defined by constant ellipsoidal longitudes and latitudes, the raytracing is performed in an ellipsoidal coordinate system. More specifically, since the GNSS coordinates are given w.r.t. the World Geodetic System 1984 (WGS84), all ellipsoidal coordinates used within this work refer to WGS84, and thus, the raytracing is performed w.r.t. the WGS84 reference ellipsoid. Yet, without loss of generality, the presented raytracing algorithm can be transferred to any ellipsoid.

As shown in Figure 4.8, the raytracing w.r.t. any ellipsoid is based, for a straight ray path, a) on an intersection of the ray with constant latitudes corresponding to an intersection of the ray with an unbent plane, b) on an intersection of the ray with constant latitudes implemented by means of an intersection of the ray with a cone, and c) on an intersection of the ray with constant ellipsoidal heights obtained by solving a linear system of equations. In the following, these three main steps are explained in more detail.



**Figure 4.8:** Schematic representation of the basic principle of a raytracing in ellipsoidal coordinate systems: The intersections of a straight ray path with the voxels must be situated a) at the intersection of the ray path with an un bent plane of constant longitude corresponding to a voxel border in longitude, b) on a cone of constant latitude corresponding to a voxel border in latitude, and c) on one of the layers of constant height representing the vertical voxel borders. The variables  $\lambda$ ,  $\varphi$ ,  $h$ , and  $N_\varphi$  stand for ellipsoidal longitudes, latitudes, and heights as well as for the radius of curvature of a reference ellipsoid. The remaining variables shown in the figure are auxiliary quantities explained within the mathematical details on the raytracing.

a) **Intersection with constant longitudes**

The intersection  $\mathbf{X}_{\text{intersect}}$  of a straight ray path

$$g(r) = \mathbf{X}_0 + r \cdot \Delta \mathbf{X} = \begin{pmatrix} X_0 \\ Y_0 \\ Z_0 \end{pmatrix} + r \cdot \begin{pmatrix} \Delta X \\ \Delta Y \\ \Delta Z \end{pmatrix} \quad (4.15)$$

with origin  $\mathbf{X}_0$  and direction  $\Delta \mathbf{X}$  with an unbent limiting plane of constant longitude  $\lambda$

$$f_\lambda(s, t) = s \cdot \begin{pmatrix} \cos \lambda \\ \sin \lambda \\ 0 \end{pmatrix} + t \cdot \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \quad (4.16)$$

is obtained by setting

$$g(r) = f_\lambda(s, t) \quad (4.17)$$

yielding

$$\begin{pmatrix} X_0 \\ Y_0 \\ Z_0 \end{pmatrix} + r \cdot \begin{pmatrix} \Delta X \\ \Delta Y \\ \Delta Z \end{pmatrix} = s \cdot \begin{pmatrix} \cos \lambda \\ \sin \lambda \\ 0 \end{pmatrix} + t \cdot \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \quad (4.18)$$

solving for  $r$ , and computing

$$\mathbf{X}_{\text{intersect}} = \begin{pmatrix} X_{\text{intersect}} \\ Y_{\text{intersect}} \\ Z_{\text{intersect}} \end{pmatrix} = \mathbf{X}_0 + r \cdot \Delta \mathbf{X}. \quad (4.19)$$

In this context,  $r$ ,  $s$ , and  $t$  are free parameters describing the ray path and the face of the plane. The vector

$$\Delta \mathbf{X} = \begin{pmatrix} \Delta X \\ \Delta Y \\ \Delta Z \end{pmatrix} = \frac{\begin{pmatrix} X_{\text{SV}} - X_0 \\ Y_{\text{SV}} - Y_0 \\ Z_{\text{SV}} - Z_0 \end{pmatrix}}{\left\| \begin{pmatrix} X_{\text{SV}} - X_0 \\ Y_{\text{SV}} - Y_0 \\ Z_{\text{SV}} - Z_0 \end{pmatrix} \right\|} \quad (4.20)$$

obtained based on the satellite position

$$\mathbf{X}_{\text{SV}} = \begin{pmatrix} X_{\text{SV}} \\ Y_{\text{SV}} \\ Z_{\text{SV}} \end{pmatrix} \quad (4.21)$$

and the ray origin  $\mathbf{X}_0$  describes the direction of the ray. The  $Y$ -component of Equation 4.18

$$Y_0 + r \cdot \Delta Y = s \cdot \sin \lambda \quad (4.22)$$

can be solved for  $s$

$$s = \frac{Y_0 + r \cdot \Delta Y}{\sin \lambda} \quad (4.23)$$

and can be inserted in the  $X$ -component of Equation 4.18:

$$X_0 + r \cdot \Delta X = \frac{(Y_0 + r \cdot \Delta Y) \cdot \cos \lambda}{\sin \lambda} \quad (4.24)$$

Multiplying by  $\sin \lambda$  and expanding yields

$$X_0 \cdot \sin \lambda + r \cdot \Delta X \cdot \sin \lambda = Y_0 \cdot \cos \lambda + r \cdot \Delta Y \cdot \cos \lambda \quad (4.25)$$

and, after sorting by  $r$  and after factoring out  $r$ ,

$$X_0 \cdot \sin \lambda - Y_0 \cdot \cos \lambda = r \cdot (\Delta Y \cdot \cos \lambda - \Delta X \cdot \sin \lambda). \quad (4.26)$$

Therefore, the possible intersection points of the ray with the planes of constant longitude of the voxel grid can be obtained by means of Equation 4.15 with

$$r = \frac{X_0 \cdot \sin \lambda - Y_0 \cdot \cos \lambda}{\Delta Y \cdot \cos \lambda - \Delta X \cdot \sin \lambda}. \quad (4.27)$$

If applied to all planes of constant longitudes limiting voxels of the tomographic model, this approach may yield a high number of theoretic intersection possibilities. However, only those intersection possibilities are kept that are not only situated on any voxel faces within the voxel grid, but on faces of those voxels that the considered ray passes.

**b) Intersection with constant latitudes**

As described in Perler [2011], the intersection  $\mathbf{X}_{\text{intersect}}$  of the ray with constant latitudes corresponds to an intersection of the ray with a cone obtained by rotating, at the considered latitude, a normal vector on the Earth's surface around the Earth's rotation axis. The ray is parameterized as in Equation 4.15, whereas the cone of radius  $R_{\text{cone}}$  can be described by the parameterization of a circle

$$X_{\text{intersect}}^2 + Y_{\text{intersect}}^2 = R_{\text{cone}}^2. \quad (4.28)$$

The cone's radius  $R_{\text{cone}}$  depends on the height  $\tilde{Z}$  and on the constant latitude  $\varphi$  of the cone

$$R_{\text{cone}} = \frac{\tilde{Z}}{\tan \varphi}. \quad (4.29)$$

This height  $\tilde{Z}$  corresponds, as illustrated in Figure 4.8, to the distance between the cone's vertex and the  $Z$ -coordinate  $Z_{\text{intersect}}$  of the intersection point within the ECEF coordinate system associated to the ellipsoidal coordinate system used for defining the voxels.

Inserting Equation 4.29 in Equation 4.28 and replacing  $\tan \varphi$  by  $\sin \varphi / \cos \varphi$ ,

$$(X_{\text{intersect}}^2 + Y_{\text{intersect}}^2) \cdot \sin^2 \varphi = \tilde{Z}^2 \cdot \cos^2 \varphi \quad (4.30)$$

is obtained. With Equation 4.19 for the  $X$ - and  $Y$ -components,

$$\left[ (X_0 + r \cdot \Delta X)^2 + (Y_0 + r \cdot \Delta Y)^2 \right] \cdot \sin^2 \varphi = \tilde{Z}^2 \cdot \cos^2 \varphi \quad (4.31)$$

can be deduced. According to Heck [2003], the ellipsoid's radius of curvature in the prime vertical  $N_\varphi$  is

$$N_\varphi = \frac{a_{\text{ell}}}{\sqrt{1 - e_{\text{ell}}^2 \cdot \sin^2 \varphi}} \quad (4.32)$$

with a semi-major axis  $a_{\text{ell}}$  and a first numeric eccentricity  $e_{\text{ell}}$  of the considered ellipsoid, and

$$Z_{\text{ell}} = (1 - e_{\text{ell}}^2) \cdot N_\varphi \cdot \sin \varphi \quad (4.33)$$

holds for the relation between ECEF Cartesian coordinates and ellipsoidal coordinate systems. Labeling the  $Z$ -coordinate of the raypath's intersection with the considered reference ellipsoid with  $Z_{\text{ell}}$  and the distance of the cone's vertex from the ellipsoid's center with  $Z_{\text{m}}$ , Figure 4.8 b) and the Euclidean geometry show that based on

$$\frac{Z_{\text{ell}} + Z_{\text{m}}}{N_\varphi} = \sin \varphi, \quad (4.34)$$

the value of

$$Z_{\text{m}} = N_\varphi \cdot \sin \varphi - Z_{\text{ell}} \quad (4.35)$$

can be computed. Introducing

$$\tilde{Z} = Z_{\text{intersect}} + Z_{\text{m}} \quad (4.36)$$

into Equation 4.31, this yields a quadratic equation of the form

$$\alpha \cdot r^2 + \beta \cdot r + \gamma = 0, \quad (4.37)$$

namely:

$$\left[ (X_0 + r \cdot \Delta X)^2 + (Y_0 + r \cdot \Delta Y)^2 \right] \cdot \sin^2 \varphi = (Z_{\text{intersect}} + Z_m)^2 \cdot \cos^2 \varphi \quad (4.38)$$

Introducing the third line of Equation 4.19 for  $Z_{\text{intersect}}$

$$\left[ (X_0 + r \cdot \Delta X)^2 + (Y_0 + r \cdot \Delta Y)^2 \right] \cdot \sin^2 \varphi = (Z_0 + r \cdot \Delta Z + Z_m)^2 \cdot \cos^2 \varphi, \quad (4.39)$$

expanding the right side, and sorting by the power of  $r$

$$\begin{aligned} r^2 \cdot & \left[ (\Delta X^2 + \Delta Y^2) \cdot \sin^2 \varphi - \Delta Z^2 \cdot \cos^2 \varphi \right] + \\ r^1 \cdot & 2 \cdot \left[ (X_0 \cdot \Delta X + Y_0 \cdot \Delta Y) \cdot \sin^2 \varphi - (Z_0 \cdot \Delta Z + Z_m \cdot \Delta Z) \cdot \cos^2 \varphi \right] + \\ r^0 \cdot & (X_0^2 + Y_0^2) \cdot \sin^2 \varphi - (Z_0^2 + 2 \cdot Z_0 \cdot Z_m + Z_m^2) \cdot \cos^2 \varphi = 0, \end{aligned} \quad (4.40)$$

the coefficients  $\alpha$ ,  $\beta$ , and  $\gamma$  of Equation 4.37 can be deduced. Labeling

$$\tilde{Z}_0 = Z_0 + Z_m, \quad (4.41)$$

the coefficients are

$$\begin{aligned} \alpha &= (\Delta X^2 + \Delta Y^2) \cdot \sin^2 \varphi - \Delta Z^2 \cdot \cos^2 \varphi \\ \beta &= 2 \cdot \left[ (X_0 \cdot \Delta X + Y_0 \cdot \Delta Y) \cdot \sin^2 \varphi - \tilde{Z}_0 \cdot \Delta Z \cdot \cos^2 \varphi \right], \\ \gamma &= (X_0^2 + Y_0^2) \cdot \sin^2 \varphi - \tilde{Z}_0^2 \cdot \cos^2 \varphi \end{aligned} \quad (4.42)$$

deviating (by a factor 2 in the first term of  $\beta$ ) from the coefficients given without derivation by Perler [2011].

A stable numerical solution of Equation 4.37 can be obtained using the numerical solver *vpasolve.m* instead of *solve.m* in MATLAB<sup>®</sup>. Again, each value for  $r$  yields a possible intersection point. Intersection possibilities that are not situated on faces of the voxels passed by the considered ray are rejected.

### c) Intersection with constant heights

In the case of an intersection with constant height layers of height  $h_{\text{ell}}$  delimiting the tomographic voxels, based on Equation 4.15 and based on the relation between ECEF and ellipsoidal coordinates

$$\begin{aligned} X_{\text{intersect}} &= (N_\varphi + h_{\text{ell}}) \cdot \cos \varphi \cdot \cos \lambda \\ Y_{\text{intersect}} &= (N_\varphi + h_{\text{ell}}) \cdot \cos \varphi \cdot \sin \lambda \end{aligned} \quad (4.43)$$

from Heck [2003], a non-linear system of equations

$$\begin{aligned} (N_\varphi + h) \cdot \cos \varphi \cdot \cos \lambda - X_0 - r \cdot \Delta X &= 0 \\ (N_\varphi + h) \cdot \cos \varphi \cdot \sin \lambda - Y_0 - r \cdot \Delta Y &= 0 \end{aligned} \quad (4.44)$$

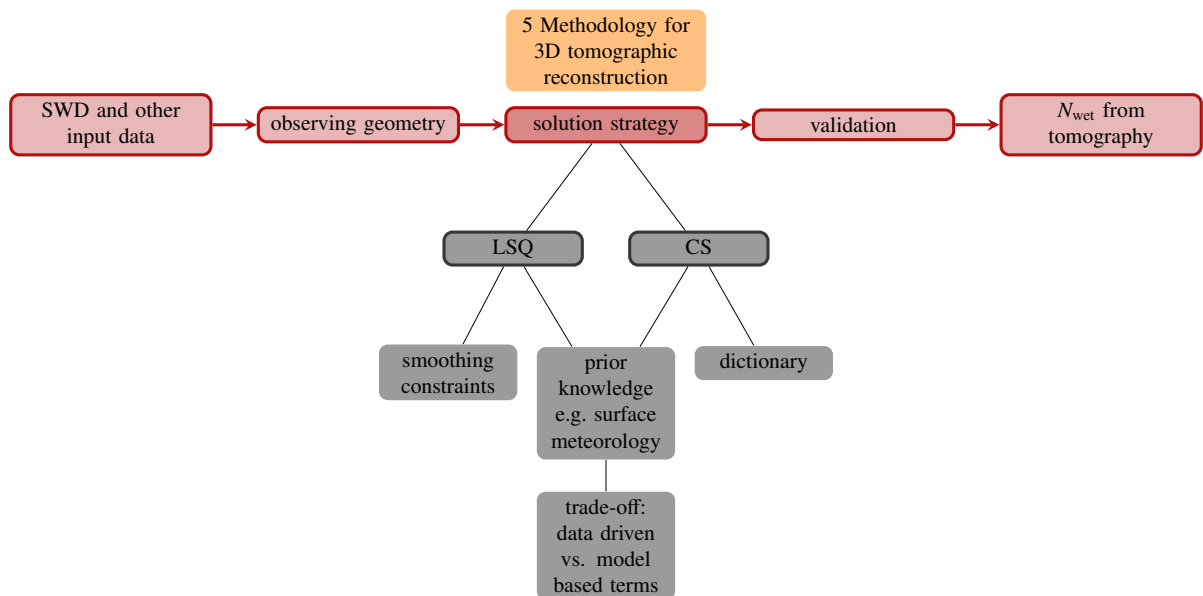
can be established. As no closed solution is available for Equation 4.44, the intersection with constant height layers is solved using an iterative method, e.g. the Newton-Raphson algorithm.





# 5 Methodology for 3D tomographic reconstruction

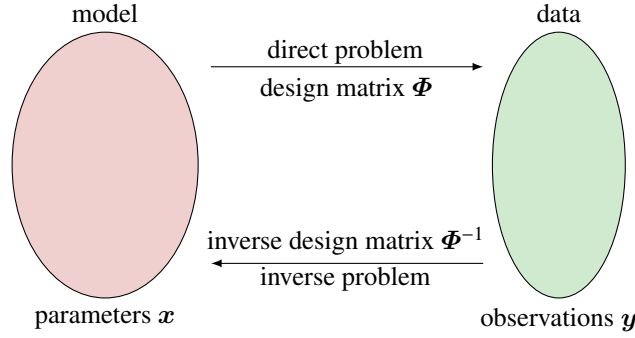
Tomographic approaches are always *inverse* problems, and as the number of viewing angles is often limited, many tomographic problems are *ill-posed* and difficult to solve. Therefore, this work develops a new solution strategy to water vapor tomography, disposing of a different way of regularizing the tomographic system. As shown in Figure 5.1, this work compares the proposed Compressive Sensing solution to a common Least Squares solution to water vapor tomography. Section 5.1 clarifies the terms *inverse problems* and *ill-posed problems* and introduces additional mathematical basics necessary for the solution of the linear system of equations encountered, independently of the solution strategy, in the tomographic reconstruction of the neutrospheric refractivity. Section 5.2 explains the terms  $L_1$ -norm and  $L_2$ -norm solution. Thereafter, the Sections 5.3 and 5.4 contrast the classical  $L_2$ -norm Least Squares method and a sophisticated  $L_1$ - $L_2$ -norm Compressive Sensing approach for tomographic refractivity reconstruction. Finally, Section 5.5 summarizes the chapter on 3D tomographic reconstruction techniques.



**Figure 5.1:** This thesis proposes a Compressive Sensing solution to water vapor tomography. Based on SWD input data, both a common Least Squares solution and the proposed Compressive Sensing solution are applied to different observing geometries.

## 5.1 Mathematical basics for the solution of inverse problems

For any physical system to be studied, Tarantola [2005] states that a *parameterization* of the system, considering a given point of view, is the discovery of a minimal set of *model parameters* whose values completely characterize the system. According to Tarantola [2005], a *direct* or *forward* model is encountered, if by means of some known *physical laws* and for given values of the model parameters, the results of measurements on some *observable parameters* can be predicted. Inverting this reasoning as illustrated in Figure 5.2, a model is said to be *inverse*, if the results of some measurements of the observable parameters at some state of the system are used to derive the model parameters at that state. This is exactly the case in tomographic applications: based on the observations



**Figure 5.2:** Direct resp. inverse problem: the observations  $\mathbf{y}$  can be predicted by means of the design matrix  $\Phi$  for given values of the model parameters  $\mathbf{x}$  resp. the model parameters are derived based on observations. The design matrix  $\Phi$  contains the physical relation between the parameters  $\mathbf{x}$  and the observations  $\mathbf{y}$ .

of the object's cross-sections, the inner structure of the object, i.e. its model parameters, is reconstructed. In the case of the human body and a computer tomography, X-ray absorption profiles are generated along many different viewing angles in order to reconstruct the inner structure of the examined body. The densities of the radiographed body sections represent the model parameters and can be deduced from the absorption profiles, which can be considered as measurements of the observable parameters. Similarly, in a neutrospheric tomography, the unknown wet refractivity values represent the model parameters, which are inferred from slant wet delay observations e.g. from GNSS corresponding to measurements of the observable parameters.

The introduction to Chapter 5 already states that many tomographic problems are *ill-posed*. Based on a study on the solution of differential equations, Hadamard [1902] defines a *well-posed* problem by

- the existence of a solution,
- the uniqueness of the solution, and by
- the *well-conditioning* of the problem, given if the solution continuously depends on the input data and a small change in the input observations does not lead to a large change in the output parameters.

In contrast to *well-posed* problems, an *ill-posed* problem is a problem that is not well-posed.

The existence and the uniqueness of an exact solution of a linear system of equations as given in Equation 4.5, with observations  $\mathbf{y} \in \mathbb{R}^{N \times 1}$ , parameters  $\mathbf{x} \in \mathbb{R}^{L \times 1}$ , and a design matrix  $\Phi \in \mathbb{R}^{N \times L}$

$$\mathbf{y} = \Phi \cdot \mathbf{x} \quad (5.1)$$

can be analyzed by means of the rank of the matrix  $\text{rank}(\Phi)$  as well as the rank of an augmented matrix  $(\Phi|\mathbf{y})$ . The rank of the augmented matrix  $\text{rank}(\Phi|\mathbf{y})$  can attain values greater or equal than  $\text{rank}(\Phi)$

$$\text{rank}(\Phi|\mathbf{y}) \geq \text{rank}(\Phi), \quad (5.2)$$

but a solution exists, if and only if

$$\text{rank}(\Phi|\mathbf{y}) = \text{rank}(\Phi). \quad (5.3)$$

If a solution exists, the dimension  $\dim(\mathcal{S})$  of the solution space  $\mathcal{S}$  of the augmented matrix  $(\Phi|\mathbf{y})$  can be computed using

$$\dim(\mathcal{S}(\Phi|\mathbf{y})) = \min(N, L) - \text{rank}(\Phi). \quad (5.4)$$

Hence, the solution is unique if

$$\text{rank}(\Phi) = \min(N, L) \quad \text{and} \quad \text{rank}(\Phi) = \text{rank}(\Phi|\mathbf{y}). \quad (5.5)$$

As  $\text{rank}(\Phi|\mathbf{y}) > \text{rank}(\Phi)$  for most overdetermined problems, such problems usually don't have an exact solution.

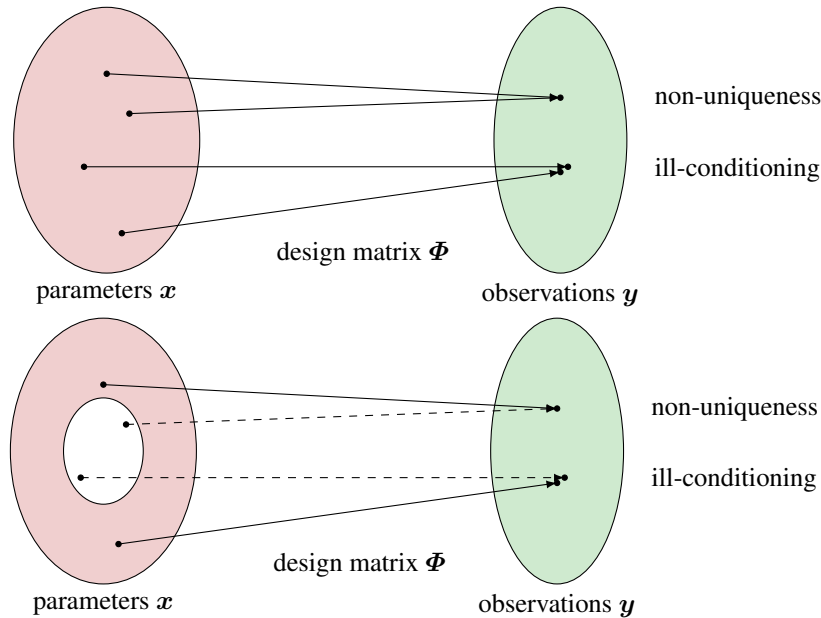
Besides the existence and the uniqueness of a solution, the well-conditioning of the linear problem is a requirement for a well-posed linear problem. The conditioning of a problem is characterized by its sensitivity with respect to small errors in the measurement data  $\mathbf{y}$ . In the case of an ill-posed problem, small errors in the input data may cause large effects in the solution. In general, the *condition number*  $\text{cond}(\Phi)$  of a linear inverse problem is considered as a measure for the degree of well-posedness. The condition number of  $\Phi$  can be computed as the ratio between the largest and the smallest singular value  $\sigma_{\max}$  and  $\sigma_{\min}$  of  $\Phi$

$$\text{cond}(\Phi) = \frac{\sigma_{\max}(\Phi)}{\sigma_{\min}(\Phi)}. \quad (5.6)$$

If the condition number is large, the problem is said to be *ill-conditioned* and therefore it is ill-posed.

There are two main ideas on how to overcome the non-uniqueness and the ill-conditioning illustrated in Figure 5.3. On the one hand, *prior knowledge* on the parameters can be introduced, on the other hand, *constraints* can be added. Such constraints can e.g. be introduced as additional observations in order to *regularize* the inverse problem. Both the Least Squares approach presented in Section 5.3 and the Compressive Sensing reconstruction within Section 5.4 regularize the ill-posed tomographic system of equations. However, as the regularization methods and the reconstruction algorithms in the two sections differ, the estimated solutions are not equal.

There are two main motivations in this work for developing a Compressive Sensing solution to the tomographic problem and for comparing it to a Least Squares solution. Firstly, the Compressive Sensing solution may be more robust and less sensitive to errors in the input SWD observations. Secondly, as the Compressive Sensing solution is expected to be sparse, less parameters need to be determined, while the number of observations remains the same as in the case of a Least Squares solution. While the Least Squares solution only involves  $L_2$ -norm minimizations, the Compressive Sensing also makes use of an  $L_1$ -norm minimization. Therefore, the following section introduces the terms  $L_1$ -norm and  $L_2$ -norm and explains the main differences between the  $L_1$ -norm solution and the  $L_2$ -norm solution to an inverse problem.



**Figure 5.3:** Non-uniqueness vs. ill-conditioning: The same observation might point towards different parameters, and small errors in the observations might cause large effects in the determination of the parameters. If at least one of these two characteristics, non-uniqueness or ill-conditioning, is fulfilled, then the problem is ill-posed. The upper resp. lower plot illustrates a linear system of equations without resp. with prior knowledge reducing the ill-posedness of the problem. The arrows indicate the operation performed by the matrix  $\Phi$ . Observations pointing towards parameters could be represented by arrows in the contrary direction, corresponding to the operation that  $\Phi^{-1}$  performs, if  $\Phi^{-1}$  exists. Constraints can be considered as additional observations limiting the solution space.

## 5.2 $L_2$ solution vs. $L_1$ solution to an inverse problem

Section 5.1 showed in which cases a linear inverse problem can be exactly solved. For linear inverse problems without an exact solution, the linear Least Squares solution aims at estimating the best approximate solution. The term *best* means that the estimated parameters  $\hat{x}$  shall minimize the  $L_2$ -norm observation residuals between the observed data  $\mathbf{y}$  and the data predicted based on the estimated parameters  $\Phi \cdot \mathbf{x}$  using Equation 5.1:

$$\hat{x} = \underset{x}{\operatorname{argmin}} \left\{ \|\mathbf{y} - \Phi \cdot \mathbf{x}\|_2^2 \right\} \quad (5.7)$$

In the following, *Least Squares* resp. LSQ will be written for *linear Least Squares*. An approximate Least Squares or  $L_2$ -norm solution may be necessary e.g. in the case of an overdetermined linear system of equations disposing of more observations than parameters, i.e. in a system of equations with  $N > L$ . Besides an  $L_2$ -norm minimization of the observation residuals, other cost functions could be applied in order to solve the tomographic system of equations. E.g. a cost function minimizing the  $L_2$ -norm of the observation residuals may be extended by an  $L_2$ -norm minimization of the parameters as in the case of a Tikhonov regularization (Menke [2012]):

$$\hat{x} = \underset{x}{\operatorname{argmin}} \left\{ \|\mathbf{y} - \Phi \cdot \mathbf{x}\|_2^2 + \Gamma_{\text{Tikhonov}}^2 \cdot \|\mathbf{x}\|_2^2 \right\} \quad (5.8)$$

Alternatively, the cost function minimizing the  $L_2$ -norm of the observation residuals may be extended by an  $L_1$ -norm minimization of the parameters, similar to the case of Compressive Sensing (Candès and Wakin [2008]):

$$\hat{x} = \underset{x}{\operatorname{argmin}} \left\{ \|\mathbf{y} - \Phi \cdot \mathbf{x}\|_2^2 + \Gamma_{\text{CS}}^2 \cdot \|\mathbf{x}\|_1 \right\} \quad (5.9)$$

The parameters  $\Gamma_{\text{Tikhonov}}$  and  $\Gamma_{\text{CS}}$  in the Equations 5.8 and 5.9 are trade-off parameters weighing the penalty terms, i.e. the  $L_1$ -norm resp.  $L_2$ -norm minimizations of the parameters, w.r.t. the observing residuals generating fidelity of the solution  $\mathbf{x}$  w.r.t. the input observations  $\mathbf{y}$ .

In order to better understand the difference of  $L_2$ -norm minimizations as applied to the observation residuals in Section 5.3 and solutions obtained using both an  $L_1$ -norm minimization of the parameters and an  $L_2$ -norm minimization of the observation residuals as those in Section 5.4, this section gives basic information on the terms  $L_1$ - and  $L_2$ -norm and on the characteristics of the solutions obtained when minimizing the respective norms of the parameters.

For a real number  $p \geq 1$ , the  $L_p$ -norm of some  $\mathbf{x} \in \mathbb{R}^{\ell \times 1}$  is defined by

$$\|\mathbf{x}\|_p = (|x_1|^p + |x_2|^p + \dots + |x_\ell|^p)^{1/p}. \quad (5.10)$$

The  $L_\infty$ -norm of  $\mathbf{x} \in \mathbb{R}^{\ell \times 1}$  is defined as

$$\|\mathbf{x}\|_\infty = \max \{|x_1|, |x_2|, \dots, |x_\ell|\}. \quad (5.11)$$

Yet, for  $0 < p < 1$ , the triangle inequality

$$\|\mathbf{x} + \mathbf{w}\|_p \leq \|\mathbf{x}\|_p + \|\mathbf{w}\|_p \quad (5.12)$$

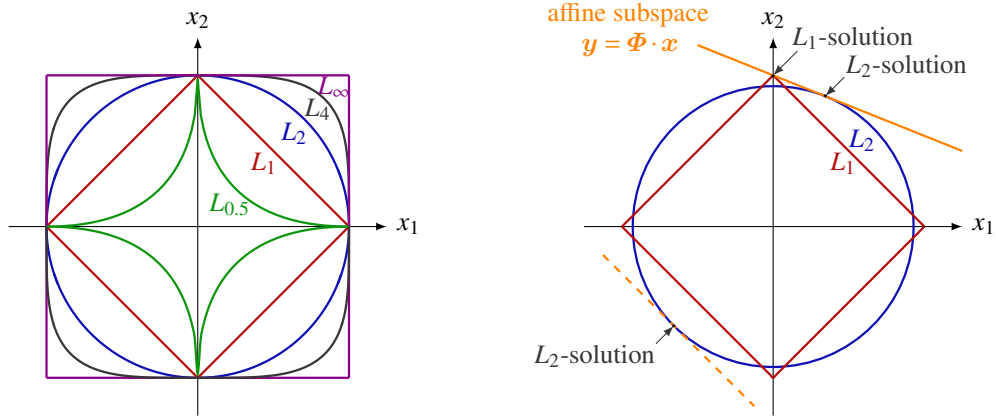
considering some second variable  $\mathbf{w} \in \mathbb{R}^{L \times 1}$  besides  $\mathbf{x}$  does not hold. Therefore, Equation 5.10 does not define a norm, but just a quasi norm for these values of  $p$ . Decreasing  $p$  even further, i.e. considering the limit of the  $p$ th power of the  $L_p$ -quasi norm of  $\mathbf{x}$  for  $p \rightarrow 0$ , a quantity

$$\|\mathbf{x}\|_0 = \operatorname{card}(\{j : x_j \neq 0\}) \quad (5.13)$$

often called the  $L_0$ -norm of  $\mathbf{x}$ , is obtained. However, it is neither a norm nor a quasi norm. Rauhut [2010] gives precise definitions of these notations.

Geometrically, linear inverse problems as that from Equation 5.1 can be imagined, in the case of  $\mathbf{x} \in \mathbb{R}^{2 \times 1}$  resp.  $\mathbf{x} \in \mathbb{R}^{3 \times 1}$ , as a line equation resp. as a plane equation. For larger parameter vectors than  $\mathbf{x} \in \mathbb{R}^{3 \times 1}$ , the Euclidean

geometry is exceeded, and Equation 4.5 generally represents an affine subspace. Finding a solution to Equation 4.5 means intersecting this affine subspace with the solution space, which depends on the solution strategy. In the case of an  $L_2$ -norm minimization of the parameters, the solution lies on a so-called  $L_2$ -ball, whereas the  $L_1$ -norm minimization of the parameters is found on an  $L_1$ -ball. Selected  $L_p$ -balls, computed based on the above definition of  $L_p$ -(quasi) norms, are schematically shown in Figure 5.4 for the case of a parameter vector  $\mathbf{x} \in \mathbb{R}^{2 \times 1}$ . In addition, Figure 5.4 shows the  $L_1$ -norm and  $L_2$ -norm solutions to a linear inverse problem.



**Figure 5.4:** Left:  $L_p$ -balls represent the solution spaces for  $L_p$ -norm minimization problems. Right: The affine subspace  $\mathbf{y} = \Phi \cdot \mathbf{x}$  intersects the  $L_1$ - resp.  $L_2$ -norm solution space in a most likely sparse resp. most likely non-sparse solution. In the plot, sparse solutions are situated on the axes because they are defined as disposing only of  $p$  non-zero or significant elements, where  $p$  is required to be smaller than the number of parameters  $L$ . The continuous resp. dashed orange lines show that in very unlikely cases, the  $L_1$ -norm solution may be non-sparse.

As both  $x_1$  and  $x_2$  in Figure 5.4 are non-zero in the solution obtained by means of  $L_2$ -norm minimization of the parameters, the solution is called *non-sparse*. In contrast, the solution derived using  $L_1$ -norm minimization of the parameters is called *sparse* because it is non-zero only for some elements, i.e. only for  $x_2 \neq 0$ .

In general, a signal  $\mathbf{x}$  is called  $\kappa$ -sparse in a basis resp. w.r.t. a transform matrix  $\Psi$  if the coefficient vector  $\mathbf{s}$  with

$$\mathbf{x} = \Psi \cdot \mathbf{s} \quad (5.14)$$

has only  $\kappa$  non-zero or significant elements, where  $\kappa$  is required to be smaller than the number of parameters  $L$ . This already indicates that the  $L_1$ -norm minimization of the parameters does not necessarily need to take place in the original domain of  $\mathbf{x}$ . Instead, the  $L_1$ -norm minimization regularizing a linear inverse problem can be performed in a transform domain in which the signal  $\mathbf{x}$  can be sparsely represented as  $\mathbf{s}$ .

Applying an  $L_1$ -norm minimization to the parameters is said to most likely yield a sparse solution. In contrast, minimizing the  $L_2$ -norm of the parameters does not promote sparseness. This is due to the fact that in an  $L_2$ -norm minimization of the parameters, large weights are assigned to large parameters, because each parameter is squared. Consequently, the  $L_2$ -norm minimization unlikely reduces small values to even smaller values or to zero. In contrast, the  $L_1$ -norm minimization gives the same weight to all parameters, which implies that in the case of an  $L_1$ -norm minimization, setting a parameter to zero is favorable.

Figure 5.4 illustrates that all  $L_p$ -norms with  $p \in [0, 1]$  are suitable for the estimation of sparse solutions. However, Baraniuk et al. [2011] state that for  $p < 1$ , the inverse problem is Non-deterministic Polynomial-time (NP) hard. Therefore, when aiming at a sparsity of the parameters as a prior for regularization, an  $L_1$ -norm minimization is commonly applied. In the case of an  $L_1$ -norm minimization, sparse solutions can be obtained by means of solving convex optimization problems.

In addition to influencing the sparsity of the solution, the application of an  $L_1$ -norm resp. of an  $L_2$ -norm minimization also effects the robustness of the solution. As stated in Menke [2012], successively higher norms give

the largest observation residual resp. the largest parameter value successively larger weights. Menke [2012] even proclaims that Least Squares weighs large errors so heavily that a single bad data point can completely throw off the result. In contrast,  $L_1$ -norm methods are expected to be more robust, i.e. yielding more reliable estimates tolerating a few bad data points.

### 5.3 Constrained Least Squares solution

As introduced in Equation 5.7, a Least Squares solution  $\hat{x}$

$$\hat{x} = \underset{x}{\operatorname{argmin}} \{ (\Phi \cdot x - y)^T \cdot P \cdot (\Phi \cdot x - y) \}, \quad (5.15)$$

to a linear functional model is obtained by minimizing the squares of the observation residuals where the matrix  $P$  in Equation 5.15 introduces, if desired, a weighting among the observations. In the unconstrained Gauß-Markov solution, the minimization of the squares of the observation residuals is done by means of

$$\hat{x} = (\Phi^T \cdot P \cdot \Phi)^{-1} \cdot \Phi^T \cdot P \cdot y. \quad (5.16)$$

In the case of a tomographic reconstruction of the neutrospheric refractivity field, small weights could e.g. be applied to low elevation rays suffering of site specific effects like multipath. Yet, in this work, all observations are equally weighted, yielding  $P = I$ . The design matrix  $\Phi$  obtained for water vapor tomography as described in Section 4.1 may be a fat, a square or a thin matrix. This is due to the fact that the shape of the matrix depends both on the number of available observations and the number of parameters, which is a function of the selected voxel discretization. I.e. depending on the number of observations and on the number of parameters, the system of equations may be underdetermined, even-determined, or overdetermined. In the case of underdetermined problems, Menke [2012] states that there can exist several exact solutions that have zero observation residuals. In contrast, in the case of even-determined problems, there is only one exact solution, and it has zero observation residuals. When considering overdetermined problems, too much information may be introduced into the system of equations to derive an exact solution and the Least Squares method approximates the best  $L_2$ -norm solution.

However, there may also occur scenarios that are neither completely overdetermined nor completely underdetermined. Namely, in this work, similarly to the example of an X-ray tomography shown in Menke [2012], there may on the one hand be voxels through which several rays pass. The refractivity within such voxels may be clearly overdetermined. On the other hand, some voxels may have been missed entirely. These voxels are completely underdetermined. In addition, there may exist voxels that cannot be individually resolved because every ray that passes through one of these voxels also passes an equal distance within another voxel. The refractivities within such two voxels are underdetermined, since only their mean refractivity can be determined. In typical water vapor tomography settings, such partly overdetermined and partly underdetermined problems, shortly called *mixed-determined* problems, have to be solved. I.e. this work does not deal with purely overdetermined problems, even if the number of observations may be larger than the number of parameters. Therefore, computing a solution to Equation 5.15 is not straightforward. Instead, prior knowledge on the parameters or additional observations constraining the solution need to be introduced into the system of equations.

Table 5.1 gives an idea of the number of parameters and of the number of observations encountered if nine satellites are assumed to be visible on average from each GNSS site. As detailed in Section 6.1, in the real data set, the observations of a total of seven GNSS sites are available per epoch. I.e. when still assuming that nine satellites were visible on average from each site, the number of observations in the real data set would only be larger than the number of parameters if the  $95 \times 99 \text{ km}^2$  large study area were discretized into only  $3 \times 3$  voxels of a size of about  $33 \times 33 \text{ km}^2$  in longitude and latitude, with only five height layers. This rough estimate of the number of observations resp. of parameters as well as the mixed-determined nature of the tomographic model emphasize the importance of regularizing the system of equations from Equation 5.15 by means of introducing prior knowledge or additional observations constraining the system.

**Table 5.1:** Number of observations  $N$  resp. number of parameters  $L$  as a function of the number of GNSS sites resp. as a function of the selected voxel discretization. The number of observations is computed by means of assuming nine satellites to be visible on average from each of the GNSS sites.

number of sites	number of observations $N$	number of parameters $L$	voxel discretization
7	63	45	$3 \times 3 \times 5$
9	81	80	$4 \times 4 \times 5$
16	144	180	$6 \times 6 \times 5$
25	225	405	$9 \times 9 \times 5$

For regularization purposes, in the case of the LSQ solution in this work, the refractivity of a voxel  $(a, b, k)$ , with voxel indices  $a$ ,  $b$ , and  $k$  in latitude, longitude, and height, is assumed to equal the mean refractivity of the remaining resp. of a certain number of surrounding voxels of the same  $k$ th height layer. As described in Heublein et al. [2018] to which the LSQ methodology described in this section refers, these horizontal smoothing constraints are applied for regularization of Equation 5.15:

$$N_{\text{wet } a,b,k} = \sum_{m,q} w_{m-a,q-b} \cdot N_{\text{wet } m,q,k} \quad (5.17)$$

Here, the voxel indices  $m$  and  $q$  correspond to the voxels in the  $k$ th height layer. The weights can be e.g. computed according to inverse distances

$$w_{m-a,q-b} = \begin{cases} \frac{1}{d_{m-a,q-b}} & \text{if } (a,b) \neq (m,q) \\ \frac{1}{\sum_{m,q} \frac{1}{d_{m-a,q-b}}} & \text{if } (a,b) = (m,q). \end{cases} \quad (5.18)$$

The distances  $d_{m-a,q-b}$  are the distances between the center of voxel  $(m, q)$  and the center of voxel  $(a, b)$  of the considered height layer.

Applying the horizontal constraints, the observation equation, Equation 5.15, can be extended to

$$\begin{pmatrix} \mathbf{P} \\ \mathbf{P}_{\text{hz}} \end{pmatrix} \begin{pmatrix} \mathbf{y} \\ \mathbf{0} \end{pmatrix} = \begin{pmatrix} \mathbf{\Phi} \\ \mathbf{\Phi}_{\text{hz}} \end{pmatrix} \cdot \mathbf{x}, \quad (5.19)$$

where  $\mathbf{\Phi}_{\text{hz}}$  contains  $w_{m-a,q-b}$  and  $\mathbf{P}_{\text{hz}}$  weighs the additional observation equations resulting from the horizontal constraint w.r.t. the unconstrained observation equation.

The horizontal constraint can be considered as approximating physical prior knowledge and helps to regularize the tomographic system of equations by increasing the rank of  $\mathbf{\Phi}$  to that of the column vector composed of both  $\mathbf{\Phi}$  and  $\mathbf{\Phi}_{\text{hz}}$ . However, the inverse problem may still be ill-posed. If the parameters can not yet be determined, further constraints or prior knowledge may be added, if available. In the case of the LSQ solution to the tomographic problem, additional constraints are included into the system of equations from Equation 5.15. According to Davis et al. [1993], an average refractivity profile can be approximated by an exponential decay with height:

$$N_{\text{wet}}(h_k) = N_{\text{wet}}(h_0) \cdot \exp\left(-\frac{h_k - h_0}{H_{\text{scale}}}\right) \quad (5.20)$$

In Equation 5.20, the variable  $h_k$  stands for the height of the  $k$ th layer with

$$k \in \mathbb{N} \text{ with } 1 \leq k \leq \mathcal{K} \quad (5.21)$$

and a total of  $\mathcal{K}$  height layers. The variable  $h_0$  in Equation 5.20 represents some reference height with known refractivity  $N_{\text{wet}}(h_0)$  and  $H_{\text{scale}}$  is the scale height of the local neutrosphere. Since the scale height is essential for the definition of the exponential refractivity decay with height, the value of  $H_{\text{scale}}$  is selected during the adjustment

from a set of possible values between 1000m and 2000m. The known wet refractivity  $N_{\text{wet}}(h_0)$  could originate e.g. from surface meteorological observations, but is, here, estimated within the adjustment. An estimation of  $N_{\text{wet}}(h_0)$  within the adjustment becomes possible, if  $N_{\text{wet}}(h_0)$  is considered as unknown parameter, i.e. if  $N_{\text{wet}}(h_0)$  is appended to the parameter vector  $\boldsymbol{x}$ :

$$\boldsymbol{x} = \begin{pmatrix} N_{\text{wet}}(h_0) \\ N_{\text{wet}}(\text{vx}_1) \\ \vdots \\ N_{\text{wet}}(\text{vx}_L) \end{pmatrix} \quad (5.22)$$

Then, Equation 5.20 needs to be adapted to the extended parameters and yields, for the refractivity  $N_{\text{wet}}(\text{vx}_j)$  of each voxel  $\text{vx}_j$ , as well as for the surface refractivity  $N_{\text{wet}}(h_0)$ , one line of equations for the vertical constraint. The voxel index range is defined as

$$j \in \mathbb{N} \text{ with } 1 \leq j \leq L = \mathcal{P} \cdot \mathcal{Q} \cdot \mathcal{K}, \quad (5.23)$$

with  $\mathcal{P}$  resp.  $\mathcal{Q}$  representing the voxel number in longitude resp. latitude. The design matrix for the vertical constraint then reads

$$\tilde{\Phi}_{\text{vert}} = \begin{pmatrix} \exp\left(-\frac{h_0-h_0}{H_{\text{scale}}}\right) & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \exp\left(-\frac{h_0-h_L}{H_{\text{scale}}}\right) & 0 & \cdots & 0 \end{pmatrix}. \quad (5.24)$$

In matrix notation, the relation of  $\boldsymbol{x}$  and  $\tilde{\Phi}_{\text{vert}}$  can be written as

$$\boldsymbol{x} = \tilde{\Phi}_{\text{vert}} \cdot \boldsymbol{x}. \quad (5.25)$$

After factoring out  $\boldsymbol{x}$  and setting  $\Phi_{\text{vert}} = \tilde{\Phi}_{\text{vert}} - \boldsymbol{I}$ , this reads

$$\Phi_{\text{vert}} \cdot \boldsymbol{x} = \mathbf{0}. \quad (5.26)$$

Including both the horizontal and the vertical constraints as well as prior knowledge from surface meteorology into the observation equation, Equation 5.15 can be extended to

$$\begin{pmatrix} P_{\text{data}} & \cdot & \boldsymbol{y}_{\text{data}} \\ P_{\text{hz}} & \cdot & \boldsymbol{y}_{\text{hz}} \\ P_{\text{vert}} & \cdot & \boldsymbol{y}_{\text{vert}} \\ P_{\text{meteo}} & \cdot & \boldsymbol{y}_{\text{meteo}} \end{pmatrix} = \begin{pmatrix} P_{\text{data}} & \cdot & \Phi_{\text{data}} \\ P_{\text{hz}} & \cdot & \Phi_{\text{hz}} \\ P_{\text{vert}} & \cdot & \Phi_{\text{vert}} \\ P_{\text{meteo}} & \cdot & \Phi_{\text{meteo}} \end{pmatrix} \cdot \boldsymbol{x}, \quad (5.27)$$

with matrices  $P_{\text{vert}}$  and  $P_{\text{meteo}}$  weighing the vertical constraint and the prior knowledge from surface meteorology w.r.t. the unconstrained observation equation and w.r.t. the horizontal constraint. The matrix  $\Phi_{\text{data}} \in \mathbb{R}^{N \times (L+1)}$  is composed of

$$\Phi_{\text{data}} = (\mathbf{0} \quad \Phi) \quad (5.28)$$

and the constraint matrices  $\Phi_{\text{hz}} \in \mathbb{R}^{L \times (L+1)}$

$$\Phi_{\text{hz}} = \begin{pmatrix} 0 & -1 & w_{1,2,1} & \cdots & w_{\mathcal{P},\mathcal{Q},1} \\ 0 & w_{1,1,2} & -1 & \cdots & \vdots \\ 0 & \vdots & \vdots & \ddots & w_{\mathcal{P},\mathcal{Q},\mathcal{K}-1} \\ 0 & w_{1,1,\mathcal{K}} & \cdots & w_{\mathcal{P},\mathcal{Q},\mathcal{K}} & -1 \end{pmatrix} \quad (5.29)$$

using  $w_{b,a,k} = w_{b,a}$  from Equation 5.18 and  $\Phi_{\text{vert}} \in \mathbb{R}^{(L+1) \times (L+1)}$  from Equation 5.26.

Both additional observations  $\boldsymbol{y}_{\text{hz}} \in \mathbb{R}^{L \times 1}$

$$\boldsymbol{y}_{\text{hz}} = \mathbf{0} \quad (5.30)$$



and  $\mathbf{y}_{\text{vert}} \in \mathbb{R}^{(L+1) \times 1}$

$$\mathbf{y}_{\text{vert}} = \mathbf{0} \quad (5.31)$$

as well as prior knowledge from surface meteorology

$$\mathbf{y}_{\text{meteo}} = \Phi_{\text{meteo}} \cdot \mathbf{x} \quad (5.32)$$

are included into the system of equations. The prior knowledge from surface meteorology is of dimension  $\mathbf{y}_{\text{meteo}} \in \mathbb{R}^{L \times 1}$  and the entries of the corresponding design matrix  $\Phi_{\text{meteo}} \in \mathbb{R}^{L \times (L+1)}$  are

$$\Phi_{\text{meteo } a,b,k} = \begin{cases} 1 & \text{if synoptic site in voxel } (b, a, k) \\ 0 & \text{elsewhere.} \end{cases} \quad (5.33)$$

No prior knowledge from surface meteorological sites situated in the vicinity of radiosonde sites used for validation is included.

As stated above, the matrix  $\mathbf{P}_{\text{data}}$  weighing the SWD estimates to each other is set to the identity matrix in this work. In contrast to  $\mathbf{P}_{\text{data}}$ , the matrices  $\mathbf{P}_{\text{hz}}$ ,  $\mathbf{P}_{\text{vert}}$ , and  $\mathbf{P}_{\text{meteo}}$  do not refer to the SWD estimates but to the additional observations introduced with the constraints. As each constraint shall have a similar effect on all voxels, and in order to equally weigh the a priori information from surface meteorology for all voxels in which prior knowledge is available, the entries of the matrices  $\mathbf{P}_{\text{hz}} \in \mathbb{R}^{L \times (L+1)}$ ,  $\mathbf{P}_{\text{vert}} \in \mathbb{R}^{(L+1) \times (L+1)}$ , and  $\mathbf{P}_{\text{meteo}} \in \mathbb{R}^{L \times (L+1)}$  only contain the effect of the horizontal and vertical constraints as well as of the prior knowledge from surface meteorology on the data fidelity term. Thus, Equation 5.27 can be reformulated as

$$\begin{pmatrix} \Gamma_{\text{hz}} & \cdot & \mathbf{y}_{\text{data}} \\ \Gamma_{\text{vert}} & \cdot & \mathbf{y}_{\text{hz}} \\ \Gamma_{\text{meteo, LSQ}} & \cdot & \mathbf{y}_{\text{vert}} \end{pmatrix} = \begin{pmatrix} \Gamma_{\text{hz}} & \cdot & \Phi_{\text{data}} \\ \Gamma_{\text{vert}} & \cdot & \Phi_{\text{hz}} \\ \Gamma_{\text{meteo, LSQ}} & \cdot & \Phi_{\text{vert}} \end{pmatrix} \cdot \mathbf{x}. \quad (5.34)$$

In Equation 5.34, a total of  $N + 3 \cdot L + 1$  lines of equations are used in order to determine  $L + 1$  parameters. Therefore, even if no SWD estimate at all were available, a total of  $3 \cdot L + 1$  lines of equations would be left in order to determine  $L + 1$  parameters. That is, a combined design matrix composed of  $\Phi_{\text{hz}}$ ,  $\Phi_{\text{vert}}$ , and  $\Phi_{\text{meteo}}$  would still be a thin matrix, meaning that the system of equations would still seem overdetermined.

The LSQ solution to Equation 5.34 is derived by solving the following minimization problem with the trade-off parameters  $\Gamma_{\text{hz}}$ ,  $\Gamma_{\text{vert}}$ , and  $\Gamma_{\text{meteo, LSQ}}$  for the constraints as well as for the prior knowledge:

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\operatorname{argmin}} \left\{ \|\mathbf{y}_{\text{data}} - \Phi_{\text{data}} \cdot \mathbf{x}\|_2^2 + \Gamma_{\text{hz}}^2 \cdot \|\mathbf{y}_{\text{hz}} - \Phi_{\text{hz}} \cdot \mathbf{x}\|_2^2 + \Gamma_{\text{vert}}^2 \cdot \|\mathbf{y}_{\text{vert}} - \Phi_{\text{vert}} \cdot \mathbf{x}\|_2^2 + \Gamma_{\text{meteo, LSQ}}^2 \cdot \|\mathbf{y}_{\text{meteo}} - \Phi_{\text{meteo}} \cdot \mathbf{x}\|_2^2 \right\}. \quad (5.35)$$

Simultaneously to the choice of the value of  $H_{\text{scale}}$ , the trade-off parameters  $\Gamma_{\text{hz}}$ ,  $\Gamma_{\text{vert}}$ , and  $\Gamma_{\text{meteo, LSQ}}$  are selected from a certain number of logarithmically scaled possible trade-off parameters. As described in Heublein et al. [2018], this selection is performed in two steps. In the first step, all those combinations of trade-off parameters are preselected that satisfy the eigenvalue cutoff criterion defined in Flores et al. [2000]. The work in Hajj et al. [1994] and Wiggins [1972] indicate that the input noise is amplified into the solution by a factor given by the smallest non-zero eigenvalue. Based on the assumption of standard deviations (std) of the observations  $\mathbf{y}$  resp. of the parameters  $\mathbf{x}$  to be estimated of  $\text{std}_{\mathbf{y}} = 5 \text{ mm}$  resp.  $\text{std}_{\mathbf{x}} = 3.5 \text{ mm/km} = 3.5 \text{ ppm}$  from Flores et al. [2000], following the approach proposed in Heublein et al. [2018], the cutoff value  $w$  for the eigenvalues is

$$w = \left( \frac{\text{std}_{\mathbf{y}}}{\text{std}_{\mathbf{x}}} \right)^2 = 2 \text{ km}^2. \quad (5.36)$$

The result of this preselection step yields a large set of stable solutions, which do not necessarily match equally well with the observations. As a consequence, in a second step, the combination of trade-off parameters resulting in minimum observation residuals is selected as final trade-off parameters.

## 5.4 Sparse Compressive Sensing solution

Section 5.2 introduced the terms  $L_1$ -norm and  $L_2$ -norm and stated that a Compressive Sensing solution to the tomographic system was obtained using a sparse  $L_1$ -norm minimization. If the signal to be reconstructed is *sparse*, its coefficients only have a small number of non-zeros. In this work, as described in Heublein et al. [2018], the wet refractivity signal  $x$  itself is not sparse, but the assumption is that a sparse representation  $s$  of it can be obtained after an appropriate transform

$$x = \Psi \cdot s. \quad (5.37)$$

Then, a Compressive Sensing solution as introduced, e.g. by Baraniuk et al. [2011] and Candès and Wakin [2008], can be applied in order to reconstruct a sparse signal  $s$  in an appropriate, sparsifying transform domain. Instead of estimating the parameters  $x$  in the original domain as proposed in Equation 5.9, sparse parameters  $s$  are estimated

$$\hat{s} = \underset{s}{\operatorname{argmin}} \left\{ \|y - \Phi \cdot \Psi \cdot s\|_2^2 + \Gamma_{\text{CS}}^2 \cdot \|s\|_1 \right\}. \quad (5.38)$$

The sparsity  $\kappa$  of a signal is, in this work, defined by the number of coefficients containing 99% of the total signal power.

The Least Squares solution described in Section 5.3 is obtained via  $L_2$ -norm minimization of the observation residuals and some geometric and meteorological regularization constraints. Instead of such a Least Squares solution, and instead of minimizing the observation residuals as in the case of Least Squares, but applying an  $L_1$ -norm as length measure, the Compressive Sensing solution is based on both an  $L_2$ -norm minimization of the observation residuals and an  $L_1$ -norm minimization of the sparse parameters  $s$ . The  $L_2$ -term ensures data fidelity, whereas the  $L_1$ -term regularizes the system by generating a sparse solution, if both terms are appropriately weighted w.r.t. each other.

Cai et al. [2010] state that  $\kappa$ -sparse signals can be recovered exactly resp. stably using  $L_1$ -norm minimization techniques, provided that the sensing matrix  $\theta$

$$\theta = \Phi \cdot \Psi \quad (5.39)$$

satisfies the Restricted Isometry Property (RIP) in the absence resp. in the presence of noise. The RIP is fulfilled if the restricted isometry constant  $\delta$  defined by Candès and Wakin [2008] as the smallest  $\delta \geq 0$  satisfying

$$(1 - \delta) \cdot \|\nu\|_2^2 \leq \|\theta \cdot \nu\|_2^2 \leq (1 + \delta) \cdot \|\nu\|_2^2 \quad (5.40)$$

for any vector  $\nu$  attains values smaller than  $\delta < 0.307$ . Yet, the RIP is difficult to verify for an infinite number of possible vectors  $\nu$ . Therefore, the mutual coherence between the design matrix  $\Phi$  and the transform matrix  $\Psi$

$$\mu(\Phi, \Psi) = \max_{1 \leq n_\Phi, n_\Psi \leq L} \frac{\left| \langle \varphi_{n_\Phi}, \psi_{n_\Psi} \rangle \right|}{\|\varphi_{n_\Phi}\|_2 \cdot \|\psi_{n_\Psi}\|_2} \quad (5.41)$$

is used in this work as a metric for assessing the CS recovery properties, although the coherence provides weaker reconstruction guarantees than the RIP. The vectors  $\varphi_{n_\Phi}$  resp.  $\psi_{n_\Psi}$  in Equation 5.41 stand for row  $n_\Phi$  of  $\Phi$  resp. for column  $n_\Psi$  of  $\Psi$ . The coherence can reach values within the interval  $[1/\sqrt{L}, 1]$ . The smaller the coherence, the less coherent are  $\Phi$  and  $\Psi$ , and the better the  $L_1$ -norm reconstruction is expected to perform.

When aiming at a Compressive Sensing solution to Equation 5.1, a transform matrix  $\Psi$  is required that represents the wet refractivity parameters  $x$  as sparse parameters  $s$ . The observation equation in the sparse domain then reads

$$y = \Phi \cdot \Psi \cdot s. \quad (5.42)$$

Section 5.4.1 describes the generation of 3D transform matrices  $\Psi$  using Kronecker multiplication as well as the definition of dictionaries for sparse representation. Thereafter, Section 5.4.2 shows how Compressive Sensing can be applied to the tomographic reconstruction of atmospheric water vapor and how Equation 5.38 can be extended by additional constraints.

### 5.4.1 Sparse dictionaries using Kronecker bases

In the following, some important terms are explained before investigating the generation of 3D sparsifying bases. The definition of 3D transform matrices using Kronecker multiplication is described and general aspects on transform matrices – or *dictionaries* – for sparse representation are given.

A *base function* resp. an *atom* corresponds to one column of the sparsifying basis  $\Psi$ . In the case of orthogonal  $\Psi$ , the terms *transform matrix* and *base function* are commonly used. More generally, when considering rectangular but non-square  $\Psi$ , the terms *dictionary* and *atom* are preferred because they do not require orthogonality of  $\Psi$  (Rauhut et al. [2008]). The expressions *dictionary* and *atom* can also be used in a generalizing way for the terms *transform matrix* and *base functions*. When thinking of languages, an *atom* would correspond to a word within a dictionary. As each word within a language dictionary is composed of different *letters*, each *atom* within the dictionary for sparse representation results from the Kronecker product of smaller items, that will be called *letters* in the following. To some extent, the explained terms may seem inconsistent, because the term *atom* does not match well with *letters* and *dictionaries*. The term *word* would be more appropriate. However, since the terms *dictionary* and *atom* are commonly used in the available literature on Compressive Sensing, this work also adopts the term *atom* instead of the term *word*. A better consistence and understanding could be obtained by imagining either the *molecules*, *atoms*, and *quarks*, or *dictionaries*, *words*, and *letters*.

In order to sparsely represent a 3D signal, a sparsifying basis jointly characterizing the signal structure present in each of the three dimensions is required. This is done by means of generating 3D transform matrices using Kronecker multiplication. According to Duarte and Baraniuk [2012], the restriction of a multidimensional signal to fixed indices for all but its  $\omega$ th dimension is called a  $\omega$ -section of the signal. Using this notation, Duarte and Baraniuk [2012] state that a single sparsifying basis  $\Psi$  can be obtained for a whole multidimensional signal by applying the Kronecker product of sparsifying bases for each of its  $\omega$ -sections. I.e. Kronecker products can be applied in order to design the sparsifying basis  $\Psi$ .

For the purpose of understanding the Kronecker product and its application for generating transform matrices, a closer look on the Kronecker product is necessary. Any matrices  $A \in \mathbb{R}^{m \times n}$  and  $B \in \mathbb{R}^{p \times q}$  can be represented as

$$A = \{a_{i_A j_A}\} = [a_1 \cdots a_n] = \begin{bmatrix} \alpha_1^T \\ \vdots \\ \alpha_m^T \end{bmatrix} \text{ and } B = \{b_{i_B j_B}\} = [b_1 \cdots b_q] = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_p^T \end{bmatrix}, \quad (5.43)$$

where Greek letters denote the rows, and bold Roman letters denote the columns of the matrices. Single matrix elements of  $A$  resp. of  $B$  are called  $a_{i_A j_A}$  resp.  $b_{i_B j_B}$ . According to Henderson and Searle [1981], the Kronecker product of the matrices  $A$  and  $B$  of Equation 5.43 is defined as the  $mp \times nq$  matrix

$$A \otimes B = \{a_{i_A j_A} \cdot B\} = \begin{bmatrix} a_{11} \cdot B & a_{12} \cdot B & \cdots & a_{1n} \cdot B \\ a_{21} \cdot B & a_{22} \cdot B & \cdots & a_{2n} \cdot B \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} \cdot B & a_{m2} \cdot B & \cdots & a_{mn} \cdot B \end{bmatrix}. \quad (5.44)$$

Applying the Kronecker product

$$\Psi = \Psi_1 \otimes \Psi_2 \otimes \Psi_3 \quad (5.45)$$

to the 3D separable bases  $\Psi_1$ ,  $\Psi_2$ , and  $\Psi_3$  of the first, second, and third dimension of the considered 3D signal yields a single basis  $\Psi$  for the entire 3D signal. According to Boussakta and Holt [1995], a basis is called separable if its base functions along the different dimensions can be represented as a product of the respective 1D base function, which holds for the most common bases like wavelets, iDCT base functions, and Fourier base functions as stated in Rivenson and Stern [2009]. If the signal  $x \in \mathbb{R}^{L \times 1}$  shall be sparsely represented by applying the transform matrix  $\Psi \in \mathbb{R}^{L \times L}$ , both the sparse signal  $s$  and the initial 3D parameters  $x_{3D} \in \mathbb{R}^{P \times Q \times K}$  must be represented by vectors. According to Equation 4.7,

$$x = \text{vec}(x_{3D}). \quad (5.46)$$

For the matrix  $\mathbf{A}$  of Equation 5.43, the  $\text{vec}$  operator is defined similarly to the  $\text{vec}$  operator transforming the 3D parameters  $\mathbf{x}_{3D}$  into a parameter vector  $\mathbf{x}$  in Equation 4.7 of Section 4.1. Namely, the  $\text{vec}$  operator stacks the columns of the matrix one on top of one another in order to yield a vector-reshaped signal representation:

$$\text{vec}(\mathbf{A}) = \begin{bmatrix} \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_n \end{bmatrix} \quad (5.47)$$

This yields

$$\mathbf{s} = \boldsymbol{\Psi} \cdot \text{vec}(\mathbf{x}_{3D}). \quad (5.48)$$

In this context, it is important to note that each element of the vector-reshaped form  $\text{vec}(\mathbf{x}_{3D})$  of the 3D signal  $\mathbf{x}_{3D}$  corresponds to a clearly defined element of  $\mathbf{x}_{3D}$ , which can be identified by an index triple along the three signal dimensions, see also Figure 4.2 for better understanding of the voxel indexing. Therefore, even though  $\text{vec}(\mathbf{x})$  is a 1D representation of the 3D signal  $\mathbf{x}_{3D}$ , the 3D characteristics of the signal  $\mathbf{x}_{3D}$  are preserved in  $\mathbf{x} = \text{vec}(\mathbf{x}_{3D})$ . In a similar way, each element of the transform matrix  $\boldsymbol{\Psi}$  from Equation 5.61 corresponds to a product of three clearly defined elements of the 1D transforms along the three different dimensions. Thus, these elements are again associated to the 3D position of the considered voxel. In this way, the 2D matrix  $\boldsymbol{\Psi}$  in Equation 5.48 performs a 3D multiplication of the 3D signal  $\mathbf{x}$ . According to Henderson and Searle [1981], this fact and the relation from Equation 5.48 can be understood based on

$$\text{vec}(\mathbf{A} \cdot \mathbf{x}_{3D} \cdot \mathbf{B}) = (\mathbf{B}^T \otimes \mathbf{A}) \cdot \text{vec}(\mathbf{x}_{3D}) \quad (5.49)$$

established for any three matrices  $\mathbf{A}$ ,  $\mathbf{x}$ , and  $\mathbf{B}$  for which the matrix product  $\mathbf{A} \cdot \mathbf{x} \cdot \mathbf{B}$  is defined.

For illustration, it might be easier to imagine a 2D transform instead of the 3D case. If a transform shall be applied to a 2D signal (e.g. an image), two 1D transforms  $\mathbf{A}$  and  $\mathbf{B}$  can be performed along each dimension of the image  $\mathbf{x}$ , thanks to Equation 5.49. Thus, a 1D transform matrix is first applied to the rows, and then, the same or some other 1D transform is applied to the columns of the image. In matrix notation, the transform  $\mathbf{x}_{AB}$  of  $\mathbf{x}$  is obtained via

$$\mathbf{x}_{AB} = \text{vec}(\mathbf{A} \cdot \mathbf{x}_{3D} \cdot \mathbf{B}) = (\mathbf{B}^T \otimes \mathbf{A}) \cdot \text{vec}(\mathbf{x}_{3D}). \quad (5.50)$$

As described above, a single transform matrix  $\boldsymbol{\Psi}$  can be obtained for the whole 3D refractivity signal by applying the Kronecker product of base functions for each of the signal's dimensions. Now, the notion of *transform matrices* is extended to that of *dictionaries* in order to be able to generate dictionaries for sparse representation. Mallat [1999] states:

*“In natural languages, a richer dictionary helps to build shorter and more precise sentences. Similarly, dictionaries of vectors that are larger than bases are needed to build sparse representations of complex signals.”*

As an example, Mallat [1999] chooses an orthogonal basis with the three basis vectors  $\mathbf{v}_1$ ,  $\mathbf{v}_2$ , and  $\mathbf{v}_3$

$$\boldsymbol{\Psi} = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 \end{bmatrix} \quad (5.51)$$

representing a dictionary of minimum size, also referred to as *complete* dictionary. This can for example be a 3D discrete Fourier basis, a 3D discrete cosine basis, or a 3D discrete wavelet basis. Using such a basis resp. dictionary, a sparse representation of simple signals might be achieved. However, complex signals might require a dictionary that is larger than a single basis, i.e. a dictionary which contains additional *atoms*  $\tilde{\mathbf{v}}_1$ ,  $\tilde{\mathbf{v}}_2$ , and  $\tilde{\mathbf{v}}_3$ :

$$\boldsymbol{\Psi} = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 & \tilde{\mathbf{v}}_1 & \tilde{\mathbf{v}}_2 & \tilde{\mathbf{v}}_3 & \dots \end{bmatrix}. \quad (5.52)$$

Such a dictionary containing more columns than rows is called *overcomplete*. In contrast to the *basis vectors* of a transform matrix, the elementary signals represented by the column vectors generating a dictionary are termed *atoms*. Dictionaries intend to represent a signal using a set of atoms to decompose the signal.

Obviously, as soon as a dictionary contains more columns than rows, the atoms cannot be mutually orthogonal anymore and the signal is represented in a redundant way. Then, the available atoms may allow multiple, equally well matching decompositions of the signal. Coming back to the natural languages Mallat [1999] is referring to, there are many different ways to express the content of some message adequately. However, it is often not easy to find the expression that fits best to the intention of the speaker. The same holds for sparse signal representations: an overcomplete dictionary implies that it is computationally hard to choose the best signal representation.

## 5.4.2 Application of CS to the tomographic reconstruction of atmospheric water vapor

When aiming at a tomographic reconstruction of atmospheric water vapor by means of Compressive Sensing, the sparse parameters  $\hat{s}$  are estimated by

$$\hat{s} = \underset{s}{\operatorname{argmin}} \left\{ \|\mathbf{y} - \Phi \cdot \Psi \cdot s\|_2^2 + \Gamma_{\text{CS}}^2 \cdot \|s\|_1 + \Gamma_{\text{meteo, CS}}^2 \cdot \|\mathbf{y}_{\text{meteo}} - \Phi_{\text{meteo}} \cdot \Psi \cdot s\|_2^2 \right\} \quad (5.53)$$

from Heublein et al. [2018], which serves as a basis for the approach described below. When compared to the LSQ solution in Equation 5.35, instead of adding horizontal and vertical constraints to the data fidelity term as in Equation 5.38, an  $L_1$ -norm regularization term is introduced here to promote sparse solutions for  $s$ . Here, minimizing the  $L_1$ -norm of the sparse parameters reduces the solution space. Subsequently, the wet refractivity estimates  $\hat{x}$  can be reconstructed by

$$\hat{x} = \Psi \cdot \hat{s} \quad (5.54)$$

with a dictionary  $\Psi \in \mathbb{R}^{L \times M}$ . The dimension  $M$  of the parameters  $s \in \mathbb{R}^{M \times 1}$  in the transform domain depends on the number of base functions resp. atoms defined in  $\Psi$ .

In the case of a tomographic reconstruction of water vapor, we assert that a sparse representation of the refractivity field can be obtained using e.g. a dictionary composed of Kronecker products of inverse iDCT letters in longitude and latitude directions and of Euler letters and Dirac letters in the height direction. Thus, the letters  $C$  schematically shown in Figure 5.5 represent iDCT letters in longitude and latitude and the letters  $\mathcal{D}$  and  $\mathcal{E}$  stand for Dirac letters and for Euler letters in the height direction.

In the context of a neutrospheric water vapor tomography, based on physical prior knowledge on the atmospheric water vapor distribution, the iDCT *letters* in longitude and latitude shall represent horizontal refractivity variations, the Euler letters model an exponential refractivity decay with height, and the Dirac letters describe deviations from a decay that could exactly be described by a linear combination of Euler letters. Examples for the mentioned letters are shown in Figure 5.6. An example for atoms of a 3D dictionary is shown in Figure 5.7. As the Dirac letters in Figure 5.7 are combined with constant letters in longitude and latitude, the resulting atoms only attain the values zero (in all but one height layer) or one (in one height layer).

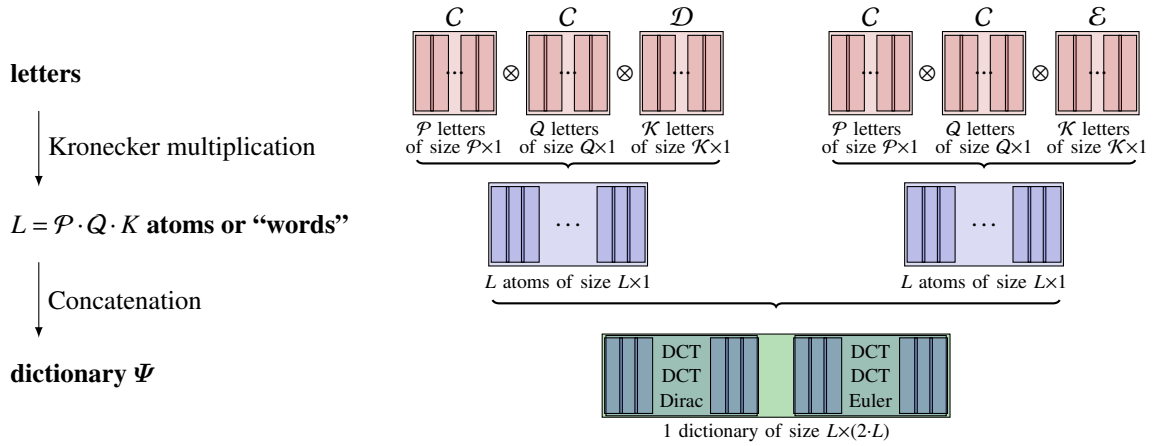
Applying Equation 5.45, the atoms of the sparsifying matrix  $\Psi$  in Equation 5.53 are built using the Kronecker product of inverse iDCT letters in the longitude and latitude directions ( $\Psi_1$  and  $\Psi_2$ ) and Euler and Dirac base functions in the height direction ( $\Psi_3$ ). The 3D refractivity distribution is assumed to be sparsely representable by means of linear combinations of these atoms.

In the following, the computation of the iDCT, Euler, and Dirac letters is explained. According to Annadurai [2007], a signal  $f(r_1)$  (e.g. the 3D refractivity distribution) is obtained based on its Discrete Cosine Transform  $F(r_2)$  (e.g. the sparse parameters) by means of

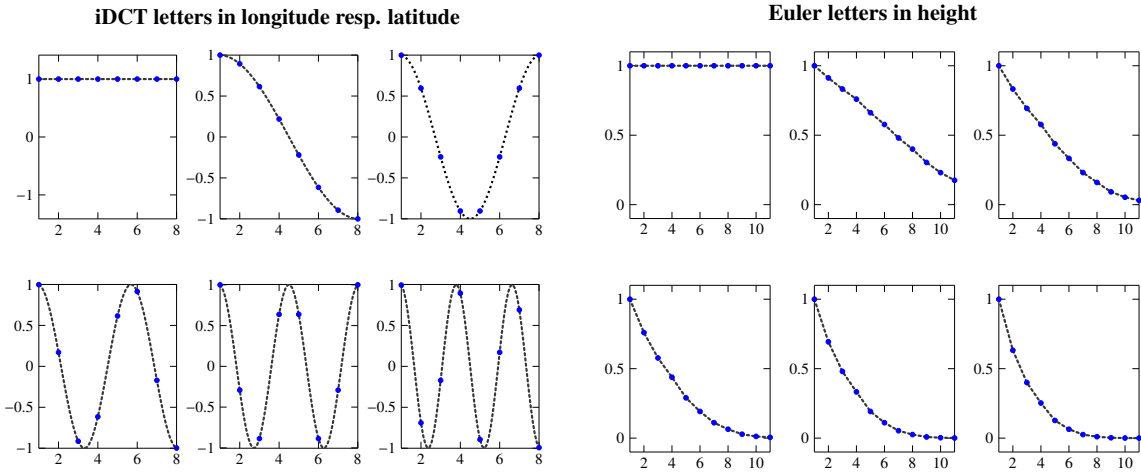
$$f(r_1) = \sum_{r_2=1}^R w(r_2) \cdot \cos\left(\frac{\pi}{2 \cdot R} \cdot (2 \cdot r_1 - 1) \cdot (r_2 - 1)\right) \cdot F(r_2) \quad (5.55)$$

with parameter indices  $r_1$  resp.  $r_2$  in the original domain resp. in the transform domain with

$$r_1, r_2 \in \mathbb{N} \quad \text{with} \quad 1 \leq r_1, r_2 \leq R \quad (5.56)$$

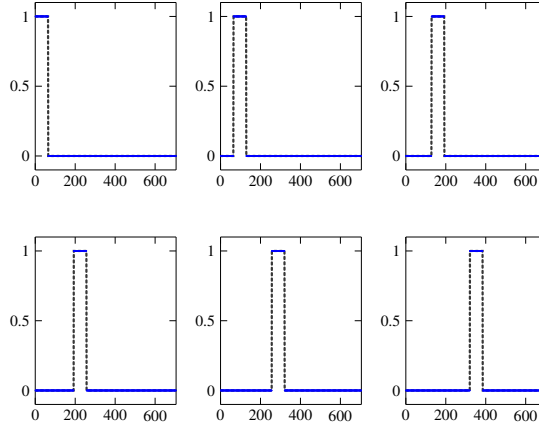


**Figure 5.5:** Relation of *letters*, *atoms*, and *dictionaries* from Heublein et al. [2018]. An *atom* corresponds to one column of the dictionary  $\Psi$ . When referring to languages, an *atom* would correspond to a word within a dictionary. As each word within a language dictionary would be composed of different *letters*, each *atom* within the dictionary for sparse representation is obtained by Kronecker multiplication of smaller items, called *letters* in the following. The dictionary  $\Psi$  is used in order to transform the coefficients in the sparse representation to the parameters in the original domain:  $x = \Psi \cdot s$ . In this work, the square  $\mathcal{C}$  summarizes  $\mathcal{P}$  iDCT letters of size  $\mathcal{P} \times 1$  in longitude or  $\mathcal{Q}$  letters of size  $\mathcal{Q} \times 1$  in latitude. The squares  $\mathcal{D}$  and  $\mathcal{E}$  summarize  $\mathcal{K}$  Dirac letters of size  $\mathcal{K} \times 1$  and the  $\mathcal{K}$  Euler letters of size  $\mathcal{K} \times 1$  in the height direction. In case of the iDCT letters and the Dirac letters, the number of letters is consciously chosen equal to the letters’ dimension in order to span the whole iDCT space and the whole Dirac space. In contrast, when considering the Euler letters, the number of letters could also differ from  $\mathcal{K}$ .



**Figure 5.6:** Representation of six 1D inverse Discrete Cosine Transform letters on the left resp. of six 1D Euler letters on the right describing the neutrospheric behavior in the longitude or latitude directions resp. modeling the refractivity decrease with height, according to Heublein et al. [2018]. Atoms for the 3D dictionary for sparse representation can be deduced by Kronecker multiplication of the iDCT letters in longitude with those in latitude and with the Euler and Dirac letters in the height direction. The gray dotted lines indicate the course of the function of the iDCT letters. The sampling points are highlighted in blue. The axis of abscissae shows the voxel number in longitude or in latitude.

Atoms of a Dirac basis in height



**Figure 5.7:** Representation of six atoms of a 3D Dirac basis dictionary, which might be used in order to correct deviations from a decay with height represented by linear combinations of atoms that are based on Euler letters in the height direction, according to Heublein et al. [2018]. The Dirac letters in the height direction are combined with constant letters in longitude and latitude. The gray dotted lines indicate the course of the function of the 3D Dirac basis. The sampling points are highlighted in blue. The axis of abscissae shows the parameter numbers 1 to  $M$  in the transform domain. The estimated solution in the transform domain is a linear combination of all atoms.

and by means of

$$w(r_2) = \begin{cases} \frac{1}{\sqrt{R}} & \text{if } r_2 = 1 \\ \sqrt{\frac{2}{R}} & \text{else.} \end{cases} \quad (5.57)$$

Thus, the iDCT letters for the longitude and latitude direction correspond to

$$C(r_1, r_2) = w(r_2) \cdot \cos\left(\frac{\pi}{2 \cdot R} \cdot (2 \cdot r_1 - 1) \cdot (r_2 - 1)\right). \quad (5.58)$$

As a function of the steepness  $\alpha_{CS} > 0$  of the exponential decay, the  $r_1$  1D Euler letters  $\mathcal{E}(r_1, r_h)$  for the vertical direction are given by

$$\mathcal{E}(r_1, r_h) = \exp\left(-\frac{r_1 \cdot (r_h - 1)}{\alpha_{CS}}\right) \quad (5.59)$$

where the  $k$ th element of  $r_h$ , again with

$$k \in \mathbb{N} \text{ with } 1 \leq k \leq \mathcal{K}, \quad (5.60)$$

is proportional to the height of the upper border of the  $k$ th voxel layer.

Each letter can be imagined to describe the neutrospheric behavior in one of the three signal dimensions longitude, latitude, and height. The parameter  $R$  in Equation 5.58 stands for the number of voxels in the respective dimension. For this work, different values of  $\alpha_{CS}$  out of the interval  $\alpha_{CS} \in [2; 10]$  that showed to yield reasonable results are introduced. The Dirac atoms have compact support: They deviate from zero only in a small interval. If Dirac letters in the height direction are considered, they can e.g. be zero for all but one height layer. Specifying Equation 5.45 for the tomographic reconstruction of water vapor using Compressive Sensing, the corresponding 3D dictionary is obtained based on a Kronecker product of many different 1D letters  $\Psi_{1D}$  for each of the three signal directions longitude, latitude, and height, represented by indices  $\lambda$ ,  $\varphi$ , and  $h$ .

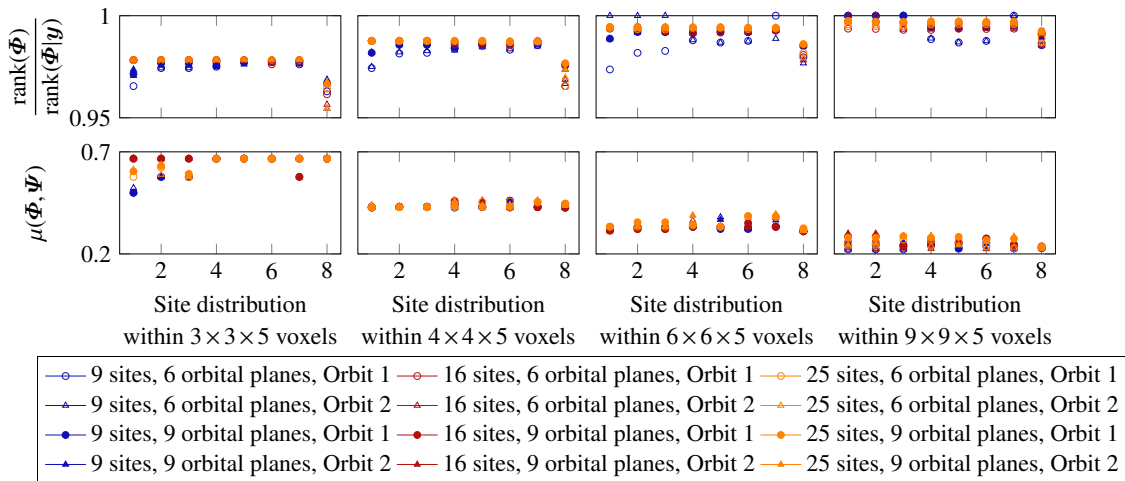
$$\Psi = \Psi_{1D,h} \otimes \Psi_{1D,\varphi} \otimes \Psi_{1D,\lambda} \quad (5.61)$$

Figure 5.7 shows atoms of a 3D dictionary resulting from the Kronecker product of two constant functions corresponding to letters in longitude and latitude and Dirac letters in the height direction.

The coefficients of  $s$  are obtained from solving Equation 5.53, where the minimization of the  $L_1$ -norm in the regularization term ensures that only a small number of atoms are selected and most of the coefficients are zero. In this work, the MATLAB `cvx` package for disciplined convex programming was used in order to solve Equation 5.53.

As in the case of LSQ, equal weights have been assigned to the observations, and prior knowledge from three surface meteorological sites is introduced. The solution is varied by varying the trade-off parameters  $\Gamma_{CS}$  and  $\Gamma_{\text{meteo}, CS}$  that weigh the sparse prior and the information from surface meteorology w.r.t. the data fidelity term. From the resulting large amount of theoretical solutions, a reasonable solution is determined by means of selecting the trade-off parameters  $\Gamma_{CS}$  and  $\Gamma_{\text{meteo}, CS}$  as carefully as the trade-off parameters  $\Gamma_{\text{hz}}$ ,  $\Gamma_{\text{vert}}$ , and  $\Gamma_{\text{meteo}, LSQ}$  in the case of LSQ in Section 5.3. Yet, instead of setting an eigenvalue cutoff criterion as in the LSQ case, a set of stable solutions is selected based on the sparsity of the solution. The number of sparse coefficients in the iDCT Euler Dirac domain containing 99% of the signal power is required to attain values between 5% and 15% of the total number  $M$  of coefficients in the transform domain, which ensures a sparse, yet not too sparse solution. Based on this preselection, the final trade-off parameter is then again chosen by minimizing the observation residuals. As  $H_{\text{scale}}$  is essential for the vertical constraint in the LSQ case, the definition of an appropriate steepness parameter  $\alpha_{CS}$  is essential for the parametrization of the vertical decay in CS. In contrast to the LSQ case, where  $H_{\text{scale}}$  is chosen during the selection of the trade-off parameters, in the case of CS,  $\alpha_{CS}$  is selected automatically by choosing appropriate atoms out of the dictionary within the minimization process. In addition, the scaling of the atoms is essential for an accurate solution. A very sparse solution with only a single, very high coefficient corresponding to one of the decreasing Euler atoms is not attractive. Instead, a linear combination of e.g. 5% to 15% of the atoms is searched that generates a more accurate solution than a single atom could produce. This is only possible if the scaling of the most prominent Euler atoms is reasonable.

While the rank of the design matrix  $\Phi$ , compared to the rank of the augmented design matrix  $(\Phi|y)$ , gives information on the existence and uniqueness of an exact solution to an inverse problem, the coherence introduced in this section is a measure for assessing the CS recovery properties. The smaller the coherence, the better should be the CS recovery. Figure 5.8 shows both the ratio of the ranks of the design matrix  $\Phi$  and of the augmented design matrix  $(\Phi|y)$  and values of the coherence  $\mu$  of  $\Phi$  and the sparsifying matrix  $\Psi$  for different numbers of GNSS sites, different distributions of these sites within the used study area, different voxel discretizations, different numbers of orbital planes, and different orbits. The figure mainly shows that the rank ratio resp. the coherence increases resp. decreases with increasing voxel number. The decreasing coherences suggest that Compressive Sensing should perform better within fine voxel grids than within coarse voxel grids.



**Figure 5.8:** Ratio of the ranks of the design matrix  $\Phi$  and of the augmented design matrix  $(\Phi|y)$  as well as values of the coherence  $\mu$  of  $\Phi$  and the sparsifying matrix  $\Psi$  for different numbers of GNSS sites, different distributions of these sites within the used study area, different voxel discretizations, different numbers of orbital planes, and different orbits.



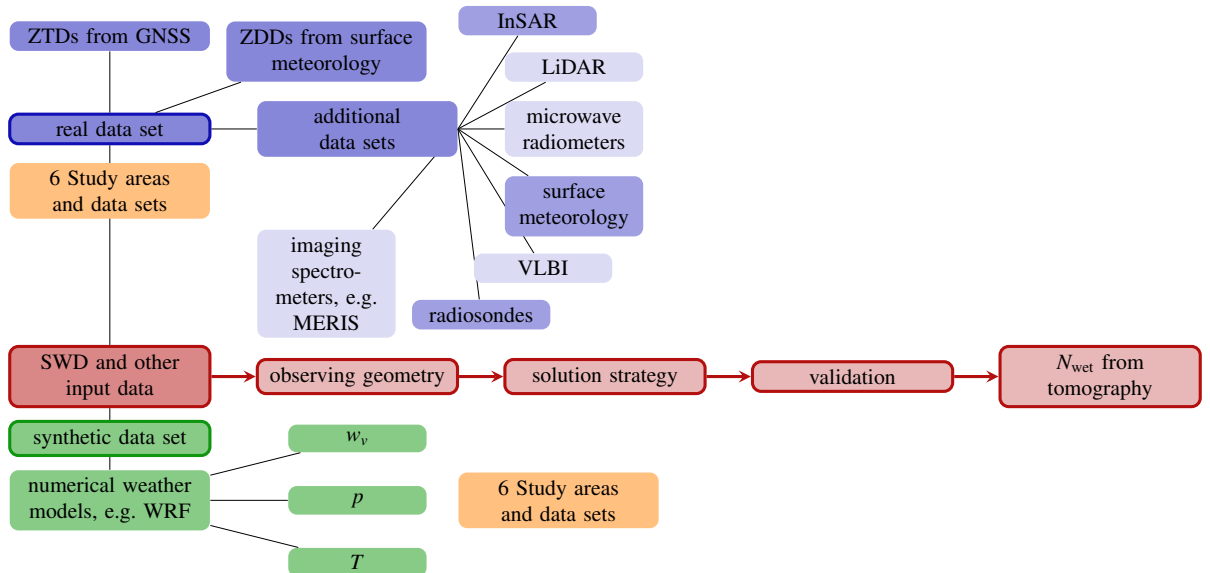
## 5.5 Summary

As the rank of the design matrix  $\Phi$  and the design matrix augmented by the observations ( $\Phi|y$ ) are never equal in this work, no exact solutions to the analyzed tomographic models exist. Consequently, the tomographic model needs to be regularized. In this work, LSQ and CS are distinguished as two main approaches for the regularization and the solution of the tomographic equation. In the case of LSQ, the tomographic equation is regularized by means of i) introducing horizontal smoothing constraints, ii) assuming the vertical refractivity decrease with height to follow an exponential decay, and iii) adding prior knowledge on the surface refractivity from surface meteorological sites. Instead of the horizontal and vertical constraints in LSQ, the CS solution benefits of the sparsity of the parameters  $s$  in some transform domain. This transform domain is defined by a dictionary composed of discrete iDCT letters in the longitude and latitude directions and by means of Euler and Dirac letters in the height direction. I.e. both the LSQ and the CS solution strategy make use of physical prior knowledge on the atmospheric water vapor distribution. The horizontal refractivity variations are expected to be much smaller than the vertical refractivity variations and the vertical refractivity decrease is approximated by means of an exponential decay with increasing height. There are two main motivations for regularizing the tomographic model by means of a sparse prior. Firstly, the CS solution may be more robust and less sensitive to errors in the input SWD observations. Secondly, as the  $L_1$ -norm solution is expected to be sparse, depending on the dictionary size, less parameters need to be determined, while the number of observations remains the same as in the case of an LSQ solution. Chapter 7 will compare the reconstruction quality of the LSQ and the CS solution within different tomographic settings resulting from both a real and a synthetic data set and from varying orbit geometries. The following section gives details on the analyzed study areas as well as on the data sets.



## 6 Study areas and data sets

In this chapter, the approaches for neutrospheric tomography from Chapter 4 are applied to both a real data set in the Upper Rhine Graben (URG) study area presented in Section 6.1 and to different synthetic data sets. As shown in Figure 6.1, for both the real and the synthetic data sets, SWD input data are necessary for the tomographic reconstruction of water vapor. The real data set in this work uses real GNSS SWD estimates as well as prior knowledge from surface meteorology. If available, real InSAR SWDs are added. In the case of the synthetic data sets, the SWD data originate from simulations, e.g. from the WRF model. The real resp. one of the synthetic data sets is used in order to investigate in how far the inclusion of InSAR SWDs resp. synthetic InSAR SWDs in addition to the GNSS SWDs resp. synthetic GNSS SWDs improves the accuracy and the precision of the reconstructed refractivity field. Additional synthetic data sets are generated with the purpose of investigating the effect of the observing geometry on the quality of the tomographic reconstruction. While the observing geometry in the real data set is mainly predetermined by the position of the GNSS sites and by the GNSS orbit around the considered acquisition times, the observing geometry in the synthetic data sets can be flexibly designed. In addition, tomographic water vapor reconstructions within the synthetic data set can be validated based on the 3D input WRF refractivities, whereas the real data set is only validated using a single radiosonde profile. Both the real and the synthetic data sets analyze the differences between the proposed CS solution and a common LSQ approach to water vapor tomography. Section 6.2 describes the data sources GNSS, InSAR, surface meteorology, and radiosonde for the real data set. Section 6.3 gives information on the data source WRF for the synthetic data set. Moreover, the derivation of the input SWD estimates for the real resp. for the synthetic data sets is explained in the Sections 6.2.5 and 6.3.2. The height systems associated to the different input data sets are presented in Section 6.4. A summary of the chapter is given in Section 6.5.

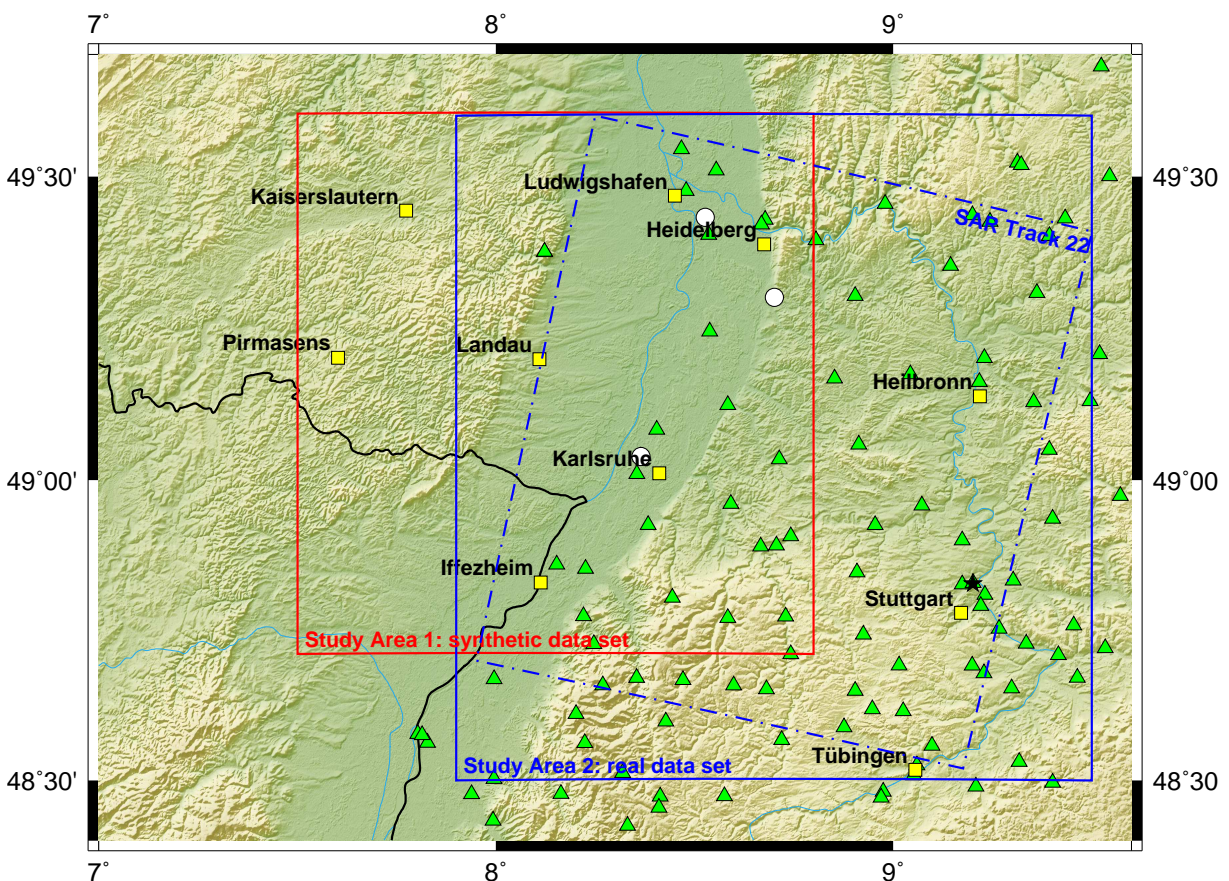


**Figure 6.1:** In this work, both a real and different synthetic data sets are used. While the observing geometry in the real data set is predetermined by the position of the GNSS sites and by the GNSS orbit geometry, the observing geometry can be flexibly selected in the real data sets. Tomographic water vapor reconstructions within the synthetic data set can be validated based on the 3D input WRF refractivities, whereas the real data set is only validated using a single radiosonde profile.

## 6.1 Study areas

For this work, two  $95 \times 99 \text{ km}^2$  resp.  $117 \times 122 \text{ km}^2$  large study areas (Study Area 1 resp. Study Area 2) in the URG in southern Germany and eastern France were chosen. According to Fuhrmann et al. [2013], the URG area is geophysically very stable within the considered time period, disposing of annual deformations in the order of  $0.5 \text{ mm}$  in the horizontal direction and about  $\pm 0.2 \text{ mm}$  in the vertical direction. This simplifies the distinction of InSAR atmospheric phases from InSAR deformation phases and makes the study area suitable for our research. The URG study area is characterized by the river Rhine flowing within an about  $35 \text{ km}$  large valley. The URG valley is surrounded by forested mountainous regions, e.g. the Black Forest in the East. As a result, the weather is mainly cold and dry in winter, but active and highly variable in summer.

As shown in Figure 6.2, GNSS observations of seven resp. eight GNSS sites are available within Study Area 1 resp. Study Area 2. However, due to the lack of radiosonde validation data, Study Area 1 is only used for the synthetic data set based on WRF data. Study Area 2 was selected for the real data set relying, for the tomographic reconstruction, on surface meteorological prior information, GNSS and InSAR data, and on radiosonde data for the validation of the estimated refractivities. The dates at which tomographic solutions are computed for Study Area 1 differ from those of Study Area 2, because the WRF data for Study Area 1 are available at other dates than the InSAR acquisitions for Study Area 2.

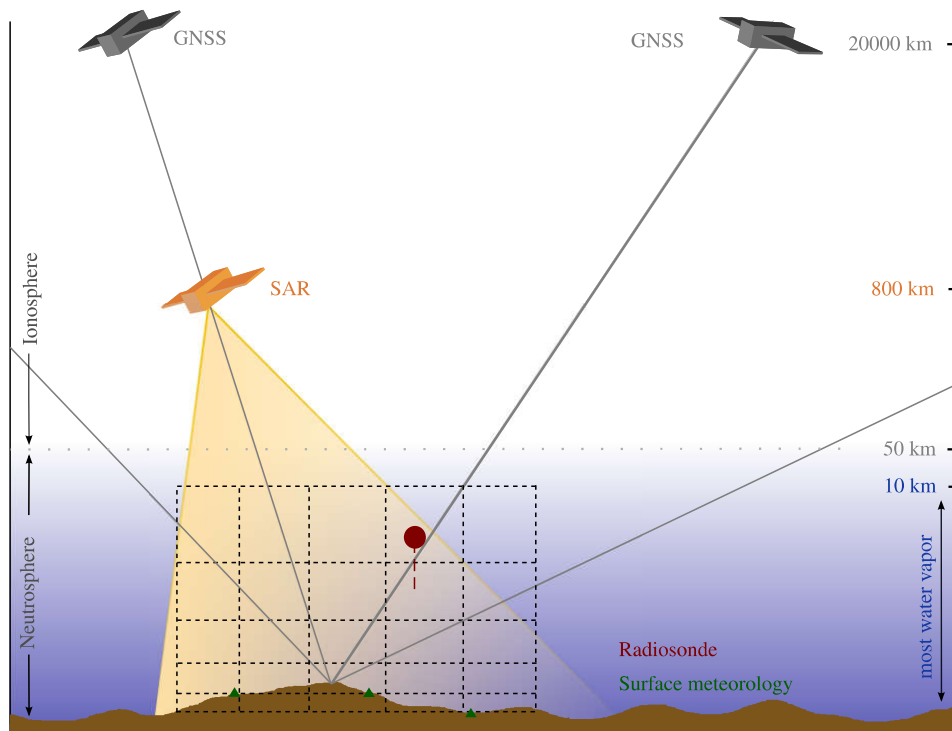


**Figure 6.2:** Study areas for the synthetic and the real data set according to Heublein et al. [2018]. The Study Area 2 (real data) contains the Envisat SAR frame of Track 22 in the Upper Rhine Graben. Yellow squares indicate the available GNSS sites. The black star indicates the radiosonde site 10739 in Stuttgart. Green triangles indicate the location of the surface meteorological sites available for the computation of the ZDDs, whereas the observations of surface meteorological sites illustrated by white circles are used in order to determine prior knowledge on the surface refractivity.

In the following, the GNSS observing geometries for the real and for the synthetic data set are described. The term *observing geometry* shall, here, comprise the position of the GNSS sites and the ray directions. Within Study Area 1, a realistic GNSS observing geometry defined by seven real GNSS sites, real orbits, and three real surface meteorological sites is used for generating the synthetic SWD data sets for the sensitivity analysis presented in Section 7.2.1. In contrast, artificial ray directions are defined from artificial site positions and from artificial satellite positions when analyzing the effect of the GNSS observing geometry on the tomographic results in the Sections 7.2.2 and 7.2.3. Within Study Area 2, the GNSS observing geometry is defined by eight real GNSS sites, real orbits, and three surface meteorological sites. A validation of the estimated 3D refractivities is possible along radiosonde profiles at the radiosonde site 10739 in Stuttgart. Further details on the generation of the real and of the synthetic SWD data set from GNSS as well as on the inclusion of synthetic InSAR data from WRF resp. of neutrospheric phases from InSAR are given in the Sections 6.2.5 and 6.3.2.

## 6.2 Real data set

As illustrated in Figure 6.3, the real data set contains GNSS and InSAR data as well as surface meteorological measurements. Therefore, the GNSS observing geometry within the real data set is predefined by the position of the GNSS sites and by the orbit geometry of the GNSS satellites. The section presents the GNSS, InSAR, and surface meteorology data introduced within the real data set. In addition, the radiosonde data used for validating the refractivity estimates are described.



**Figure 6.3:** Schematic illustration of the observing geometry of point-wise GNSS and area-wide InSAR observations. The real data set is composed of GNSS and InSAR data and prior knowledge from surface meteorology. Radiosonde observations are used for validating the 3D refractivity estimates from tomography.

### 6.2.1 GNSS data availability

The GNSS observations used in this work originate from the German satellite positioning service (SAPOS<sup>®</sup>) permanent GNSS sites and cover the period of time from 2002 to 2014. The data given within Receiver Independent

Exchange Format (RINEX) files is sampled at a temporal resolution of 15 s. On average, there is a GNSS site every 37 km resp. 42 km in Study Area 1 resp. in Study Area 2. Table 6.1 gives an overview over the GNSS sites of the two study areas. In this work, GNSS refers to GPS only. Using the Bernese PPP processing scheme given in A.2, ZTDs estimates were deduced for all SAR acquisition of the year 2005.

**Table 6.1:** GNSS sites within the study area: names, locations, coordinates, and assignment to the study areas

site name	locations	$\lambda$ in $^\circ$	$\varphi$ in $^\circ$	$h_{\text{ell}}$ in m	Study Area 1	Study Area 2
0384	Stuttgart	9.1709	48.7795	341.0		✓
0386	Heilbronn	9.2183	49.1385	234.8		✓
0387	Heidelberg	8.6753	49.3889	168.8	✓	✓
0388	Iffezheim	8.1126	48.8301	185.4	✓	✓
0394	Tübingen	9.0559	48.5177	386.0		✓
0520	Ludwigshafen	8.4506	49.4687	158.3	✓	✓
0521	Landau	8.1094	49.1998	208.0	✓	✓
0522	Pirmasens	7.6025	49.2021	448.4	✓	
0523	Kaiserslautern	7.7740	49.4441	307.4	✓	
KARL	Karlsruhe	8.4113	49.0112	182.9	✓	✓

## 6.2.2 InSAR: data availability, PSI processing, and absolute InSAR ZWD generation

A total of 17 C-band Envisat SAR acquisitions at a repeat cycle of 35 days is available for each of the two study areas. Track 294 corresponds to Study area 1, whereas Track 22 corresponds to Study Area 2. Table 6.2 shows the available acquisition dates for the respective tracks. As no radiosonde validation data are available for Study Area 1, only the SAR Track 22 situated within Study Area 2 is used. For Study Area 2, independently of the acquisition date, the SAR acquisition time in Track 22 is constant at 9h48 UTC. The SAR acquisition itself takes 29 seconds. A total of 332828 PS points are detected in the SAR swath of Track 22. Within the real data set, this work focuses on the seven PS-InSAR neutrospheric phase maps of Track 22 acquired in the year 2005, for which GNSS SWDs have been computed.

**Table 6.2:** InSAR acquisition dates covering the URG Study area in Track 294 resp. Track 22

index	date Track 294	date Track 22		index	date Track 294	date Track 22
1	2003-01-03	2003-08-13		10	2005-08-01	2006-03-15
2	2004-07-12	2005-01-19		11	2005-09-05	2006-05-24
3	2004-11-29	2005-02-23		12	2005-11-14	2006-06-28
4	2005-01-03	2005-03-30		13	2006-07-17	2006-09-06
5	2005-02-07	2005-05-04		14	2006-10-30	2007-01-24
6	2005-03-14	2005-08-06		15	2007-01-08	2008-07-02
7	2005-04-18	2005-07-13		16	2007-04-23	2008-09-10
8	2005-05-23	2005-08-17		17	2008-12-08	2009-01-28
9	2005-06-27	2005-10-26				

As explained in Section 2.6.3, a PSI processing is applied in order to identify stable points in the interferograms. At these persistent scatterers, the scattering phase cancels out and the targets thus show phase stability at all acquisition times. These stable points are then used to estimate the neutrospheric phase maps using a collection of spatial and temporal filtering routines. I.e. the different phase components contained in the interferometric phase are separated by assuming their spatial and temporal structures. The InSAR atmospheric phases used within this work were provided by Fadwa Alshawaf who carried out many research on fusing GNSS and InSAR data for deriving maps of PW (see, e.g. Alshawaf [2013], Alshawaf et al. [2015c], Alshawaf et al. [2015b], and Alshawaf et al. [2015a]).

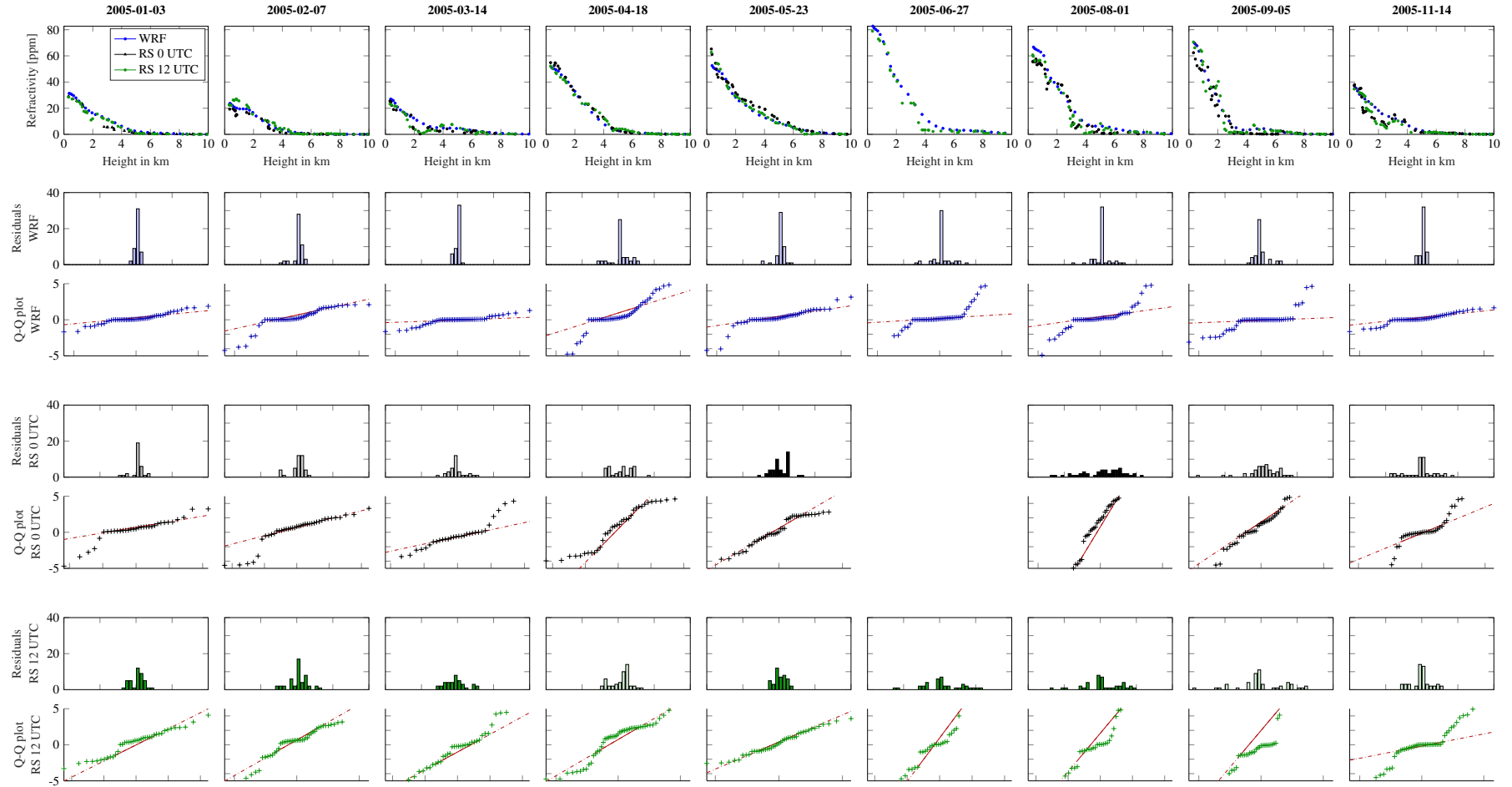
### 6.2.3 Surface meteorology

Meteorological data received from the Landesanstalt für Umwelt, Messungen und Naturschutz Baden-Württemberg (LUBW) are used to model the ZDD at the GNSS sites. The location of the LUBW surface meteorological sites are shown in Figure 6.2. At a temporal resolution of 30 min to 1 h, they measure air pressure, temperature and dew point temperature. Based on this information, ZDDs can be computed as described in Section 2.6.1. The height of the surface meteorological sites is given as orthometric height. In order to stay consistent with the GNSS ellipsoidal heights, the geoid undulation  $U$  has to be added to the orthometric heights when computing ZDDs for the GNSS sites according to the Saastamoinen model from Equation 2.18. Besides using the surface meteorological information for the ZDD modeling, prior knowledge on the surface refractivity can be derived from the surface meteorological sites by means of applying Equation 2.5. In order to derive the ZDD at a GNSS site, the closest surface meteorological site to that GNSS site, up to 100 km distant, is used.

### 6.2.4 Radiosonde data

Radiosondes are usually carried into the atmosphere by a weather balloon. They contain instruments that measure various atmospheric parameters and transmit the measured values to a receiver on the ground by means of radio waves. Global radiosonde data are freely available at a temporal resolution of 12 hours on <http://weather.uwyo.edu/upperair/sounding.html> (2018-04-26). Within the two study areas introduced in Section 6.1, a single permanent radiosonde site is available in Stuttgart, i.e. in Study Area 2. The radiosonde profiles associated to the radiosonde site Stuttgart 10739 are available at 0 UTC and at 12 UTC for most Track 22 SAR acquisition dates. Besides others, during their ascents, radiosondes capture pressure, height, temperature, dew point, relative humidity, and water vapor mixing ratios. Based on these data, wet refractivity profiles can be computed. In this work, the radiosonde refractivity profiles are assumed to extend vertically above the radiosonde ascent site, even though the radiosonde is driven by winds and may change direction during its ascent. For the nine dates for which WRF simulations are available, the first line of plots within Figure 6.4 compares refractivities from radiosonde profiles above Stuttgart with the simulated WRF refractivities above the radiosonde ascent site. In contrast to the simulated WRF refractivities, the radiosonde refractivity profiles are considered to represent the true atmospheric state. Depending on the acquisition date, the simulated and the radiosonde refractivities deviate significantly. On 2005-02-07 and at heights above about 3 km, the WRF refractivities simulated for 9h50 UTC seem to attain refractivity values between the two values deduced from radiosonde observations at 0 UTC and at 12 UTC. In lower heights, WRF overestimates resp. the radiosonde underestimates the refractivity at that date. In contrast, on 2005-05-23 and on 2005-08-01, the simulated refractivities from WRF are smaller than the radiosonde refractivities for all heights above 2 km.

Consequently, the available WRF data can neither represent good prior knowledge for the tomographic reconstruction of water vapor nor a means for validation. Moreover, Figure 6.4 motivates future research on alternatives to the exponential decay of the wet refractivity introduced as vertical constraint into the LSQ solution. As shown in this figure, the refractivity decay observed along the radiosonde ascent only follows an exponential law in eight out of the 17 plotted radiosonde acquisition times. The exponential character of the decay is determined as follows. For each setting (i.e. for WRF and for the two radiosonde ascents at 0 UTC and 12 UTC), an exponential fit to the WRF resp. to the radiosonde refractivity decay is computed. In a next step, the residuals between this exponential fit and the WRF resp. the radiosonde refractivity decay are derived. In case of an exponential refractivity decay, the residuals are assumed to be normally distributed. A Shapiro-Wilk test is used in order to statistically return a test decision for the null hypothesis that the residuals come from a distribution in the normal family, against the alternative that it does not come from such a distribution. The significance level of the test is set to 5%. The Shapiro-Wilk test of Shapiro and Wilk [1965] has been selected due to its suitability for sample sizes below 50 samples. The analyzed WRF and radiosonde height profiles dispose of such small sample sizes of about 30 to 50 values, depending on the acquisition date.

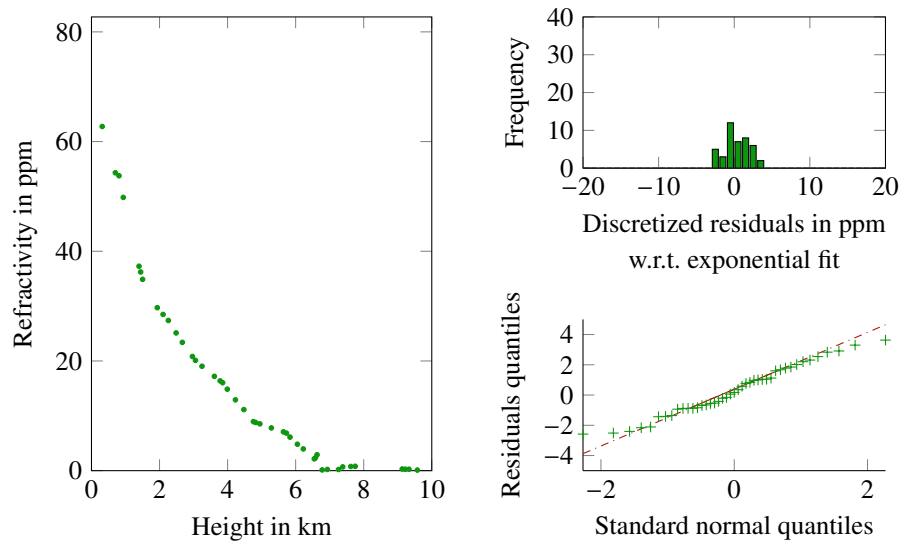


**Figure 6.4:** The first line of plots compares mean WRF refractivities per height layer for the SAR acquisition time with, if available, radiosonde profiles of the neutrospheric refractivity above Stuttgart at the radiosonde ascent times 0UTC and 12UTC for all SAR acquisition dates in 2005. The three lower pairs of lines show frequency distributions and quantile-quantile plots (Q-Q plots) of residuals between the WRF resp. radiosonde data shown in the first line and an exponential fit to these data. If the bar color in the histograms is opaque, a Shapiro-Wilk test judged the residual distribution to follow a normal distribution, otherwise, in case of transparent fill colors in the bars, the hypothesis test rejected the null hypothesis that the residuals follow a normal distribution. The frequency distributions of the residuals are plotted with 1 ppm wide bars within the range from  $-20$  ppm to  $20$  ppm. The axes of ordinate of the Q-Q plots covers the interval  $[-5, 5]$ , even though there are also dates at which this range is exceeded. For this dates, only a zoom to the interval  $[-5, 5]$  is shown. The axes of abscissae of the Q-Q plots cover the interval  $[-2.3, 2.3]$ .



In addition, quantile-quantile plots (Q-Q plots) are generated in order to assess visually whether the residual distribution comes from a normal distribution. The Q-Q plot orders the residual distribution from the smallest to the largest residual value and then plots these values against the values expected for a normal distribution at each *quantile* in the residual distribution. *Quantiles* of the residual distribution are points dividing the distribution into intervals with equal probabilities. The *2-quantile* of a distribution e.g. corresponds to the median of the distribution, i.e. to the value dividing the distribution into two equal-sized groups.

As illustrated in the zoom of Figure 6.4 in Figure 6.5, the axes of ordinate of the Q-Q plots in Figure 6.4 show the quantile values of the residual distribution, while the axes of abscissae contain the theoretical values of a normal distribution at the same quantiles at which the residual distribution is analyzed. If the Q-Q plot is linear as it is approximately the case in Figure 6.5, the residual distribution can visually be judged to come from a normal distribution, and thus, the refractivity decay is assumed to follow an exponential decay.



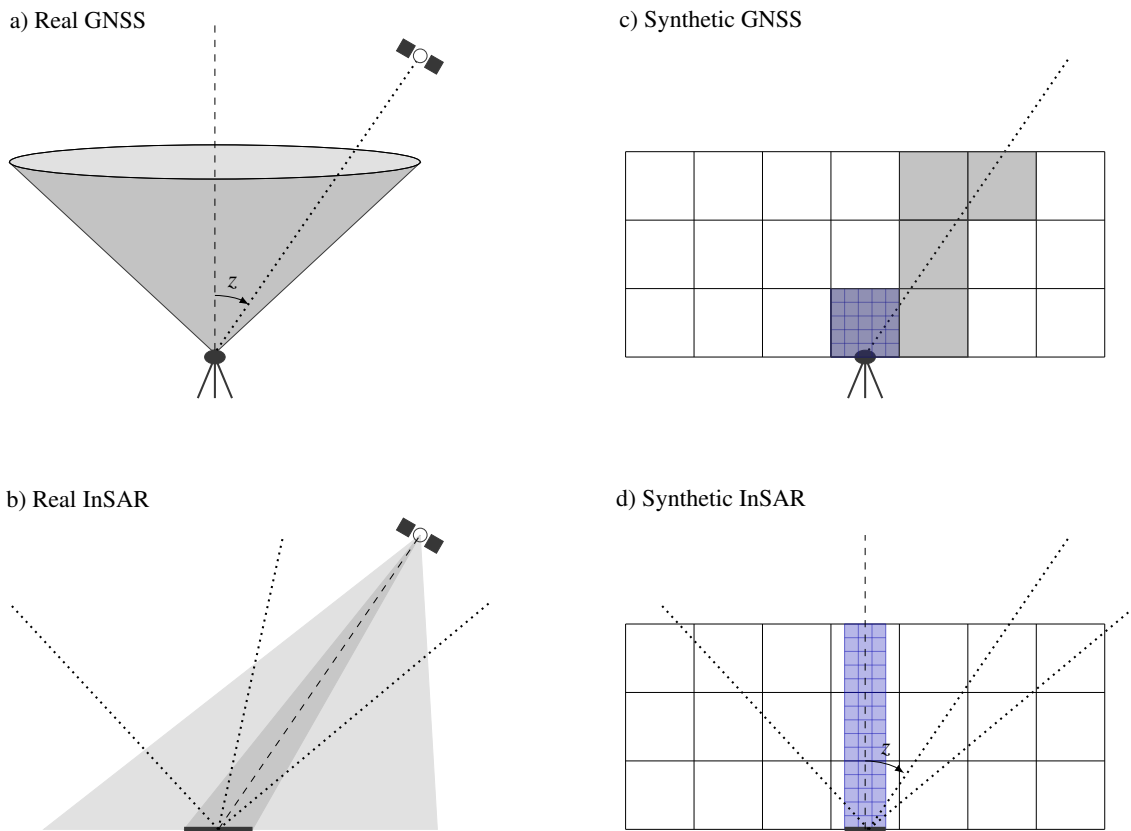
**Figure 6.5:** The left plot shows a zoom on the radiosonde refractivity decrease on 2005-05-23 at 12 UTC. On the top right, the frequency distribution of the residuals between the radiosonde refractivity decrease of the left plot and an exponential fit to this refractivity decrease is plotted. On the bottom right, the associated Q-Q plot is given.

### 6.2.5 Real SWD data set based on GNSS and InSAR

According to Heublein et al. [2018], to which this whole section refers, the tomographic reconstruction relies on total SWD estimates from GNSS PPP and PSI in the case of the real data set. GNSS SWD estimates are included into the system of equations. As described in Section 2.6.2, these ZWDs are separated from the ZTDs by means of subtracting the ZHDs derived from the Saastamoinen model from Equation 2.18 and including horizontal gradients in northing and easting. The GNSS SWDs are derived from the ZWD values as described in Section 2.6.2 and are mapped according to the azimuth and elevation angles of the real GPS satellites, as shown in Figure 6.6 a).

The neutrospheric model within the GNSS processing is composed of the Saastamoinen zenith model (hydrostatic, wet), the hydrostatic and wet Niell mapping functions, and a tilting gradient model. The estimation interval of the ZTDs corresponds to 15 min. Each set of total horizontal gradient parameters is estimated for 24 h.

The optimal scenario for building 3D wet refractivity fields using GNSS tomography is to have a dense GNSS network. At each GNSS site, an SWD estimate is available as input for the tomographic system. In reality, as stated in Section 6.2, the GNSS mean inter-site distance in Study Area 2 is about 42km, i.e. the site density is quite low. However, InSAR provides a dense network of PS points at which atmospheric phases are available. Consequently, while considering GNSS SWDs on the one hand, on the other hand, 2D absolute ZWD maps as



**Figure 6.6:** Observing geometry of real resp. synthetic GNSS and InSAR SWDs. The real GNSS observing geometry is shown in the subfigure a). Zenith wet delays averaged over all visible GNSS satellites within a cone defined by the cutoff elevation angle  $\epsilon_{\text{cut}}$  are mapped according to the zenith angle  $z$  into the direction of the considered GNSS satellite. In contrast, in the case of real InSAR data in subfigure b), the InSAR ZWDs of all those persistent scatterers are averaged that are situated within a radius  $r_{\text{average}}$  around the considered InSAR site. The obtained mean ZWD within  $r_{\text{average}}$  is then mapped into artificial ray directions. In the case of synthetic GNSS, mean refractivities per voxel are deduced, based on a certain voxel discretization, from the WRF cells' refractivities. As shown in subfigure c), the artificially defined rays are then traced within the voxel grid in order to obtain SWDs in the respective ray direction. Finally, subfigure d) illustrates how synthetic InSAR SWDs are obtained. For each WRF cell of the lowest WRF cell layer situated within  $r_{\text{average}}$  around the considered synthetic InSAR site, a ZWD value is integrated along the vertical column above the cell. Similarly to the averaging of the SWDs of all those PS points situated within  $r_{\text{average}}$  around an InSAR site in subfigure b), these ZWDs for the WRF cell columns surrounding the synthetic InSAR site are averaged, and the corresponding SWDs are obtained by means of mapping the cylindric average ZWDs into artificial ray directions.

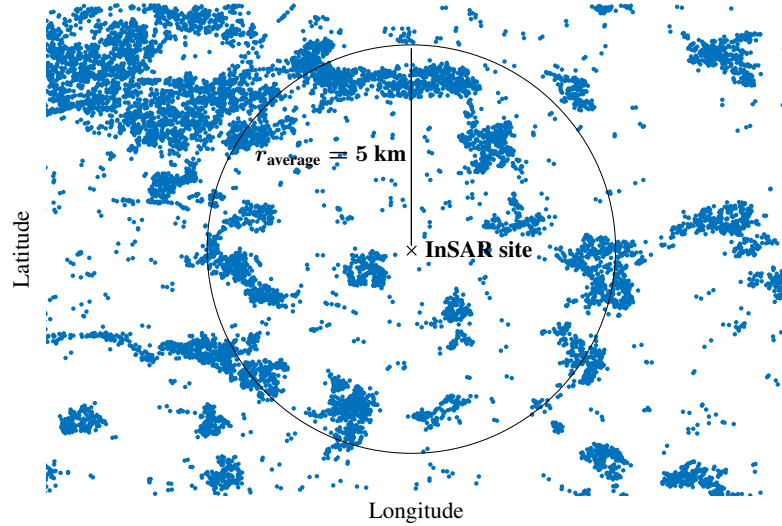
The small blueish grids in the figure schematically represent WRF cells, whereas the large grids stand for the voxel discretization. Dotted lines correspond to the ray directions introduced into the tomographic system, while dashed lines represent the directions in which the slant resp. zenith wet delays are primarily obtained. In case of satellites at the beginning of a ray, the ray corresponds to a real satellite position. If no satellite is drawn at the beginning of a ray, then the ray direction is artificial.

defined in Section 2.6.3 are introduced into the tomographic system. When compared to the  $\pm 15$  min of GNSS SWDs around the SAR acquisition time that are included within the real data set, the InSAR ZWD maps represent short snapshots of the atmospheric state. These InSAR ZWD estimates can be aggregated to derive real wet delay input data at given points, called *InSAR sites* in the following, as if corresponding to GNSS sites within the study area. Such InSAR ZWDs can be estimated for any *InSAR site* simulated within the InSAR swath. As shown in Figure 6.7, this is done by means of averaging the InSAR ZWDs of all those persistent scatterers that are situated within a radius  $r_{\text{average}}$

$$r_{\text{average}} = \frac{H_{\text{scale}}}{\tan \epsilon_{\text{cut}}} \quad (6.1)$$

around the InSAR site, where  $H_{\text{scale}}$  is set to some mean scale height for the considered study area, e.g., in this work, to 1480 m. The InSAR ZWDs averaged around the InSAR sites could be visualized as averaging cylinders, whereas in the case of GNSS, the water vapor distribution is averaged within conical shapes centered at the GNSS site of the GNSS observing geometry.

In this first study incorporating absolute InSAR ZWDs into the solution of a water vapor tomography, and as illustrated in Figure 6.6 b), the obtained mean ZWDs per cylinder are mapped into artificial directions that simulate a possible satellite geometry, e.g. with azimuth angles  $A \in [0^\circ, 360^\circ[$  and elevation angles  $\epsilon \in [7^\circ, 90^\circ]$ .



**Figure 6.7:** Schematic illustration of the PS distribution around an InSAR site from Heublein et al. [2018]. In the figure, the ZWDs of the PS points (blue dots) are averaged within a radius of 5 km (black circle) around the shown InSAR site.

Since these ray directions are artificially selected, the mapped SWDs do not necessarily represent the real atmospheric behavior in the considered direction. In the case of azimuthal refractivity variations or a large tilting of the local atmospheric horizon around the InSAR site, the artificial ray directions may corrupt the InSAR SWDs, even though the InSAR ZWDs contain a large component originating from azimuthally averaged GNSS ZWDs. Nevertheless, these InSAR SWDs are introduced into the tomographic system of equations in order to provide additional rays in many different directions.

In this work, one InSAR site is defined in the horizontal center of each voxel of the lowest tomographic layer and 20 artificial rays are defined per InSAR site. The heights of the InSAR sites are deduced from the height of the PS points situated within  $r_{\text{average}}$  around the InSAR site. Surface meteorological prior knowledge on the surface refractivity of the three synoptic sites indicated by white circles in Figure 6.2 is included into the tomographic system. The validation of the tomographic reconstruction using external data is only possible in Stuttgart. As in the case of GNSS, the mapping from the zenith direction to the slant direction is performed using the Niell mapping function.

## 6.3 Synthetic data set

When compared to the real data set, the synthetic data set fully originates from WRF simulations. For each component of the real data set, Table 6.3 shortly describes the equivalent component within the synthetic data set.

**Table 6.3:** Comparison of the components of the real and of the synthetic data set

	Real data set	Synthetic data set
GNSS SWDs	from each GNSS site to each visible GNSS satellite, deduced from real GNSS ZTDs as well as from ZDDs from surface meteorological observations	synthetically derived from WRF ZWDs, e.g. from each GNSS site of the real data set to each GNSS satellite that would be visible in the real data set, or, alternatively, based on artificial site locations and artificial ray directions
InSAR SWDs	obtained for artificially defined <i>InSAR sites</i> and artificially defined ray directions using absolute SWD maps from InSAR	synthetically derived from WRF SWD maps for artificially defined <i>InSAR sites</i> and artificially defined ray directions
prior knowledge on surface refractivity	from surface meteorology	from WRF
validation	1D radiosonde refractivity profiles	3D refractivity fields from WRF

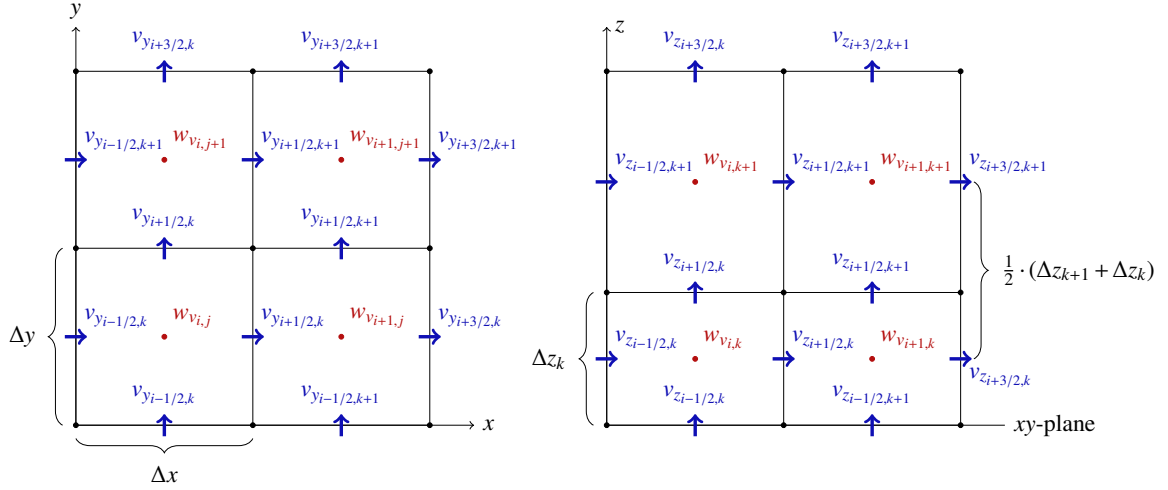
### 6.3.1 WRF data set

The SWD estimates of the synthetic data set are generated based on simulations of the WRF modeling system. According to Skamarock et al. [2008], WRF is a NWP and atmospheric simulation system serving both research and operational needs. NWP applies mathematical models of the atmosphere and oceans in order to predict the weather based on current weather conditions. Contrary to the voxels with ellipsoidal upper and lower boundaries in the tomographic model, the upper and lower boundaries of the WRF cells do not correspond to parallel ellipsoidal surfaces of constant ellipsoidal height. Instead, the vertical coordinate of the WRF output corresponds to terrain-following geopotential heights. Moreover, while the tomographic voxels are defined by their outer corners, the WRF cells are defined by the coordinates of their centers. As indicated in Skamarock et al. [2008] and shown in Figure 6.8, vectorial quantities like velocities  $v$  are staggered one-half cell length from scalar variables given at the cell centers.

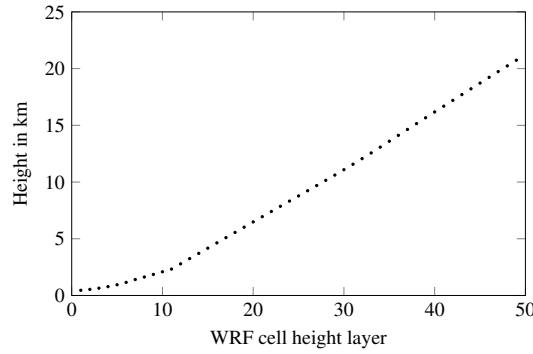
The WRF simulations are based on compressible, non-hydrostatic Euler equations extended by the effects of moisture in the atmosphere. In fluid mechanics, Euler equations govern adiabatic and inviscid flow, represent the conservation of mass, and balance momentum and energy.

For this work, 3D WRF simulations with a horizontal resolution of 900 m were received from Franz Ulmer (Remote Sensing Technology Institute, German Aerospace Center) for the nine Track 294 SAR acquisition dates within 2005. They contain, for example, simulations of temperature, pressure, and water vapor mixing ratios. The WRF model is run with three nested domains at horizontal resolutions of about 9 km, 3 km, and 900 m. With increasing height above the surface, the thickness of the WRF cells increases from about 75 m to about 500 m, as shown in Figure 6.9.

On the considered acquisition dates, the simulation was run at a temporal rate of ten minutes for the URG study areas centered at  $(\lambda, \varphi) = (8.15^\circ, 49.15^\circ)$ . The domain with the highest spatial resolution has a horizontal extension of  $200\text{km} \times 200\text{km}$  and contains a total of 49 height layers extending up to approximately 20 km. For all dates at which WRF simulations are available, Figure 6.4 shows the WRF height layer-wise refractivity decay averaged over the whole simulation area. The figure shows that the weather is dry and cold in winter time, and that the refractivity generally decreases with increasing height. If the exponential character of the WRF refractivity decay



**Figure 6.8:** Spatial discretization in WRF: The WRF cells are defined by the coordinates of their centers. Vectorial quantities  $v$  are staggered one-half cell length from scalar variables  $w$  given at the cell centers. The left plot shows a top view on a horizontal layer, whereas the right plot shows a vertical slice at some position within the  $xy$ -plane.



**Figure 6.9:** Thicknesses of the WRF cells at the constant longitude and latitude  $(\lambda, \varphi) = (6.93^\circ, 48.33^\circ)$  and different heights for 2005-05-23, deduced from the WRF simulation

is tested as shown in Figure 6.4 and described in Section 6.2.4, the WRF refractivity decay does not follow an exponential law. As the initial boundary conditions for the WRF simulations originate from global atmospheric models, locally and temporally occurring atmospheric phenomena cannot be expected to be accurately simulated.

### 6.3.2 Synthetic SWD data set based on WRF

Based on the observing geometry of the available GNSS and InSAR measurements, synthetic SWDs are calculated from the WRF data as described in Heublein et al. [2018], on which this whole section is based. Computing the synthetic SWDs from the WRF data enables a direct comparison of the later estimated 3D water vapor field with the reference data available from WRF.

The synthetic GNSS data set is generated based on WRF using the azimuth and elevation angles of real GNSS rays as well as real GNSS site coordinates in longitude and latitude. The heights of the sites for the synthetic data set correspond to the heights of the WRF DEM at the longitudes and latitudes given by the GNSS sites. The WRF simulation output (water vapor mixing ratios, pressure, temperature) is transformed into wet refractivities as shown in Figure 2.5. As described in more detail in Section 4.3, a single WRF refractivity value is derived for each tomographic voxel by averaging all WRF cells situated within that tomographic voxel. Thereafter, Equation 2.11

and a direct raytracing along the real GNSS rays within the tomographic voxels, as illustrated in Figure 6.6 c), yield the synthetic GNSS SWDs.

For the synthetic InSAR data set, additional sites and rays can be simulated at any point on the WRF DEM using artificial directions that emulate a possible satellite geometry, e.g. with azimuth angles  $A \in [0^\circ, 360^\circ[$  and elevation angles  $\epsilon \in [7^\circ, 90^\circ]$ . For these *synthetic InSAR sites*, the procedure generating synthetic SWDs is the following. All those WRF cells of the lowest WRF layer are determined that are horizontally situated within a radius  $r_{\text{average}}$  around the considered synthetic InSAR site. For each of the selected WRF cells of the lowest layer, a ZWD value is then integrated along the vertical column above the cell. As shown in Figure 6.6 d), similarly to the averaging of the SWDs of all those PS points situated within  $r_{\text{average}}$  around an InSAR site in Section 6.2.5, these ZWDs for the WRF cell columns surrounding the synthetic InSAR site are averaged, and the corresponding SWDs are obtained by means of mapping the cylindric average ZWDs into the artificial ray directions defined above. The mapping of the ZWDs to the slant direction is performed by dividing the ZWDs by the sine of the respective elevation angle. No gradients are considered in the synthetic data set. In this work, 35 synthetic InSAR sites are defined within the horizontal centers of the  $5 \times 5$  ground voxels, at a height given by the WRF DEM. A total of 20 rays per site is defined.

As described in Section 5, the tomographic system is regularized by means of horizontal and vertical constraints as well as prior knowledge from surface meteorology. In case of the synthetic data set, this prior knowledge is obtained from the WRF model. A 3D validation of the reconstructed refractivities within the tomographic voxels is done using the input WRF refractivities averaged in the tomographic grid.

## 6.4 Geodetic and meteorological height systems

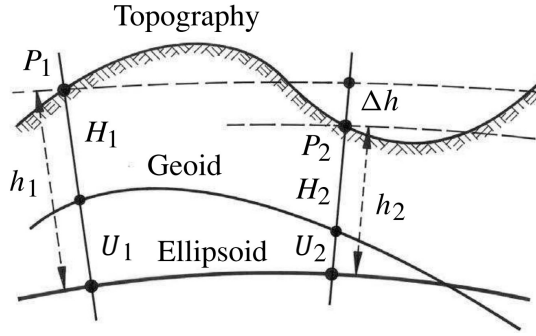
When dealing with several different data sources, the respective height systems have to be taken into account. In this work, *ellipsoidal* GNSS site heights are given with respect to the geodetic datum WGS84. The raytracing is also performed in ellipsoidal coordinates. In contrast, the PS points' heights as well as the heights of the surface meteorological sites refer to *orthometric heights*. As shown in Figure 6.10, the difference between ellipsoidal heights  $h_{\text{ell}}$  and orthometric heights  $H$  is the geoid *undulation*  $U$ :

$$U = h_{\text{ell}} - H \quad (6.2)$$

According to Seeber [2003], the undulations can reach up to 100m for a global reference ellipsoid. Since undulations of about 50m are observed in the URG study area, the differences between orthometric and ellipsoidal heights must be corrected before introducing data sources from satellite geodesy (with ellipsoidal heights) and from terrestrial geodesy (with orthometric heights defined w.r.t. the gravity field) into a combined adjustment. Applying Equation 2.11, the effect of a 50 m height error on the SWD corresponds, at a surface refractivity of 50ppm, to 2.5mm.

For the available surface meteorologic sites, undulations were provided by Kurt Seitz from the Geodetic Institute of the Karlsruhe Institute of Technology using the model European Gravity Geoid 1997 (EGG97). They are given with respect to the geodetic datum WGS84 which coincides with the Geodetic Reference System 1980 (GRS80). According to Seeber [2003], the gravity field associated to GRS80 is the Earth Gravitational Model 1996 (EGM96). EGM96 is a spherical harmonic model of the Earth's gravitational potential complete to degree and order 360.

In a similar way, undulations were derived for the whole study area from the difference between EGM96 and GRS80 at a grid spacing of  $0.0083^\circ$ . Based on these undulations, orthometric heights of any PS point within the study area can be transformed into ellipsoidal heights. Consequently, GNSS and InSAR data as well as surface meteorological information can be used in a combined adjustment.



**Figure 6.10:** Relation between geoid undulation  $U$ , orthometric height  $H$ , and ellipsoidal height  $h_{\text{ell}}$  according to Seeber [2003]

Besides ellipsoidal and orthometric heights, this work works with data given with respect to *geopotential* heights. Firstly, the WRF data refer to perturbation geopotential PH and base state geopotential PHB, which are related to the geopotential height  $H_{\text{geopotential}}$  by means of

$$H_{\text{geopotential}} = \frac{\text{PH} + \text{PHB}}{g} \quad (6.3)$$

with gravity  $g$ . Secondly, the radiosonde data used for validation often refer to the geopotential height  $H_{\text{geopotential}}$ .

According to Wallace and Hobbs [2006], geopotential heights are used as the vertical coordinate in most atmospheric applications in which energy plays an important role. They relate the geopotential height  $H_{\text{geopotential}}$  to a surface point's potential  $W_{P_0}$  difference to the geoid potential  $W_P$

$$C = W_{P_0} - W_P = - \int_{P_0}^P dW = \int_{P_0}^P g dv \quad (6.4)$$

by means of

$$H_{\text{geopotentialWallace}} = \frac{C}{g_0}, \quad (6.5)$$

where the value of  $g_0$  corresponds to the globally averaged acceleration due to the gravity at the Earth's surface. The geopotential number  $C$  is computed based on the potentials  $W_{P_0}$  and  $W_P$  of an arbitrary point  $P_0$  on the geoid and a point  $P$  on the surface. The potential difference can be derived from geometric leveling and gravity measurements. As described in Torge [2001], this is done by integrating the mean gravity  $g$  along leveling segments  $dv$  along any path between  $P_0$  and  $P$ .

In Kraus [2007] and in Lewis [2007], the geopotential height is defined as the geopotential number  $C$  divided by the normal gravity  $g_N = 9.80665 \text{ m/s}^2$  at a latitude of  $\varphi = 45^\circ$ , which corresponds, according to Yilmaz [2008], to the definition of dynamic heights:

$$H_{\text{geopotentialKraus}} = \frac{C}{g_N} \quad (6.6)$$

In contrast, orthometric heights are computed by means of

$$H = \frac{C}{\bar{g}}, \quad (6.7)$$

with  $\bar{g}$  standing for the actual average gravity along the plumb line. As described in Torge [2001], this definition of orthometric heights corresponds to the common understanding of *heights above sea level*. In meteorological literature like Kraus [2007] or Lewis [2007], orthometric heights defined w.r.t. a reference geoid are called *geometric heights*.

In order to get an idea of the differences between orthometric heights and geopotential heights, Kraus [2007] approximate the geopotential number  $C$  at a latitude  $\varphi$  by means of

$$C \approx g(\varphi, H) \cdot H \quad (6.8)$$

with

$$g(\varphi, H) = g_N \cdot \left(1 - 0.0026373 \cdot \cos(2 \cdot \varphi) + 0.0000059 \cdot \cos^2(2 \cdot \varphi)\right) \cdot \left(1 - 3.14 \cdot 10^{-7} \cdot H\right), \quad (6.9)$$

where  $H$  is divided by meter. For the URG study area situated around a latitude of  $\varphi \approx 49^\circ$ , differences between orthometric heights and geopotential heights are given in Table 6.4. Up to heights of about 2000m, geopotential and geometric heights correspond to each other on less than 1m. In higher altitudes around 10km, the geopotential heights are about 30m smaller than the orthometric heights. Taking also into account that the orthometric heights in our study area are about 50m smaller than ellipsoidal heights yields maximum height deviations between geopotential heights and ellipsoidal heights of about 80m, which cannot be neglected within this work.

In this work, geopotential heights are used within the synthetic data set and become relevant when validating the estimated refractivities by means of radiosonde observations. After conversion of the geopotential heights  $H_{\text{geopotential}}$  into ellipsoidal heights  $h_{\text{ell}}$ , the validation is performed within the tomographic voxel grid. As described by Wang et al. [2016], this is done by first transforming the geopotential heights into orthometric heights  $H$  and by then converting the orthometric heights into ellipsoidal heights according to Equation 6.2. Wang et al. [2016] state that the conversion from orthometric heights to geopotential heights can be performed using

$$H_{\text{geopotential}}(\varphi, H) = \frac{g_S(\varphi)}{g_N} \cdot \frac{R_{\text{Earth}}(\varphi) \cdot H}{R_{\text{Earth}}(\varphi) + H}, \quad (6.10)$$

where  $g_S(\varphi)$  corresponds to the normal gravitation acceleration at some latitude  $\varphi$  and  $R_{\text{Earth}}(\varphi)$  is the effective radius of the Earth at some latitude  $\varphi$ , both defined below. Consequently, orthometric heights  $H$  can be deduced from geopotential heights  $H_{\text{geopotential}}$  by means of

$$H(\varphi, H_{\text{geopotential}}) = \frac{\frac{g_S}{g_N} \cdot H_{\text{geopotential}} \cdot R_{\text{Earth}}(\varphi)}{R_{\text{Earth}}(\varphi) - \frac{g_S}{g_N} \cdot H_{\text{geopotential}}}. \quad (6.11)$$

The normal gravitation acceleration at some latitude  $\varphi$  on the surface of the ellipsoid,  $g_S$ , is computed using

$$g_S(\varphi) = 9.780325 \frac{\text{m}}{\text{s}^2} \cdot \sqrt{\frac{1 + 0.00193185 \cdot (\sin \varphi)^2}{1 - 0.00669435 \cdot (\sin \varphi)^2}}, \quad (6.12)$$

and the effective radius of the Earth at some latitude  $\varphi$  is derived from

$$R_{\text{Earth}}(\varphi) = \frac{6378.137 \text{ m}}{1.006803 - 0.006706 \cdot (\sin \varphi)^2}. \quad (6.13)$$

The numbers in the Equations 6.12 and 6.13 are given for converting geopotential heights to orthometric heights referring to the WGS84 ellipsoid.

**Table 6.4:** Differences between orthometric heights and geopotential heights

$H$ in m	$H_{\text{geopotential}}$ in m	$H - H_{\text{geopotential}}$ in m
0	0	0
500	500.1	-0.1
1000	1000.1	-0.1
2000	1999.5	0.5
5000	4994.0	6.0
10000	9972.3	27.7

Finally, if no metric height information was contained in the radiosonde output variables, the height could be approximated based on temperature and pressure information using the international height formula given in Burde et al. [2015] or Wiegleb [2016]

$$p_{H_{\text{geopotential}}} = p_0 \cdot \left(1 - \frac{0.0065 \text{ in } \frac{\text{K}}{\text{m}} \cdot H_{\text{geopotential}} \text{ in m}}{T_0 \text{ in K}}\right)^{5.255}, \quad (6.14)$$



reformulated as

$$H_{\text{geopotential}} \text{ in m} = \frac{T_0 \text{ in K}}{0.0065 \text{ in } \frac{\text{K}}{\text{m}}} \cdot \left( 1 - \left( \frac{p_{H_{\text{geopotential}}}}{p_0} \right)^{1/5.255} \right). \quad (6.15)$$

The values of  $p_0$  and  $T_0$  correspond to reference values of pressure and temperature at some reference height and the reference height is set to the start point of the radiosonde. The international height formula is valid for heights up to 11 km.

## 6.5 Summary

Chapter 6 gives an overview over the study areas selected for this work. Both study areas are situated in the URG. Due to the availability of WRF data and the lack of radiosonde validation data, Study Area 1 is only used for the synthetic data set. Study Area 2 is defined in order to perform a tomographic reconstruction of water vapor based on real GNSS and InSAR SWD estimates. A summary of the available data sets is given in Table 6.5. The next chapter will present the tomographic results obtained for the respective data sets.

**Table 6.5:** Summary of the available data sets

Data set	Spatial resolution	Temporal resolution	Availability	Height system
GNSS (GPS)	point-wise	15 seconds	2002-2014	ellipsoidal
PSI (Envisat)	here: $\varnothing$ 23 points per km <sup>2</sup>	35 days	2003-2009	orthometric
Surface meteorology	point-wise	30-60 min	since 1991	orthometric
Radiosonde	along point-wise profiles	12 h (site 10739)	as of 1973	geopotential
WRF 3.1	$3 \times 3$ km <sup>2</sup>	10 min	9 days in 2005	geopotential

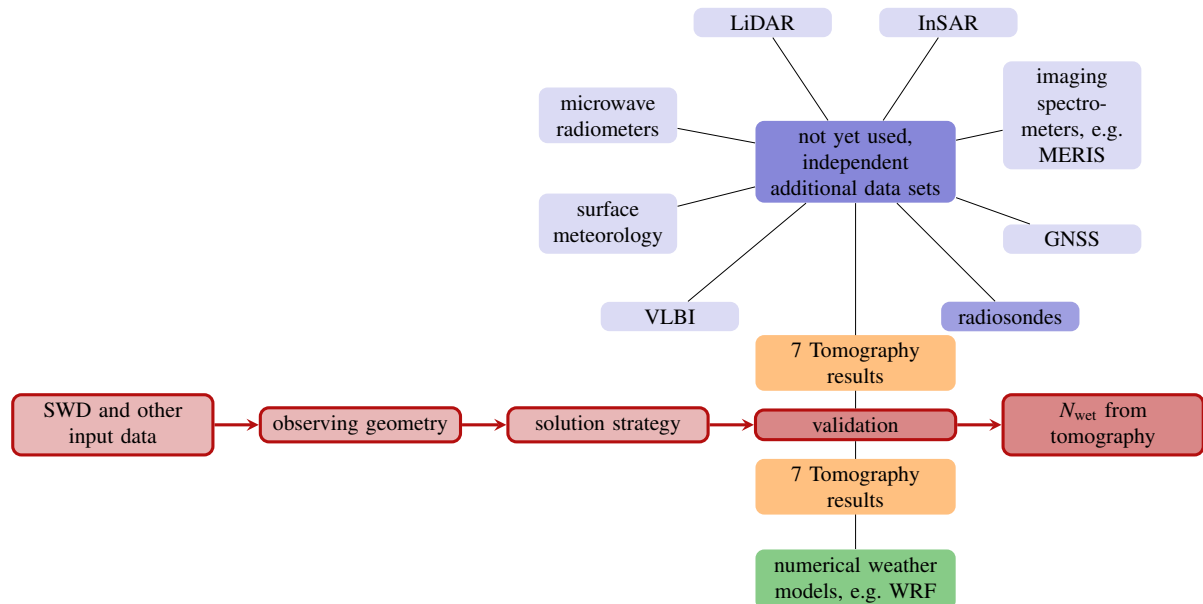


# 7 Tomography results

This chapter aims at

- analyzing in how far the inclusion of InSAR SWD estimates improves the tomographic reconstruction of the 3D neutrospheric refractivity field,
- comparing the reconstruction quality of the LSQ and the CS solution to the tomographic model, and
- investigating to what extent the observing geometry has an effect on the tomographic results.

As illustrated in Figure 7.1, the quality of the tomographic results is evaluated, in the real resp. in the synthetic data sets, by means of validating using a single radiosonde ascent resp. by means of comparing the estimated 3D refractivity field with input WRF refractivities.



**Figure 7.1:** The quality of the tomographic results is evaluated, in the real resp. in the synthetic data sets, by means of validating using a single radiosonde ascent resp. by means of comparing the estimated 3D refractivity field with input WRF refractivities. Therefore, the radiosondes are highlighted as validation possibility for the real data set. If available, any other technique yielding independent data sets that are not yet introduced into the tomographic system could also be used for validation, e.g. microwave radiometers or LiDAR.

The analysis of the improvement obtained by introducing InSAR SWD estimates within the tomographic system considers in how far the inclusion of InSAR SWD estimates regularizes the inverse model and improves the accuracy and the precision of the estimated wet refractivity parameters. When comparing the reconstruction qualities of the LSQ and the CS solution strategy to the tomographic model, special focus is set on the questions i) which solution approach is more accurate and more precise, ii) in how far one of the strategies is more flexible, i.e. less constraint-driven, and iii) if the CS solution can do with fewer observations than LSQ. Alternatively, question iii) investigates in how far CS is able to estimate the neutrospheric water vapor field at a higher spatial resolution than LSQ, or if CS can estimate the water vapor field more accurately and more precisely than LSQ, given a certain number of observations and a certain spatial resolution.

In addition, the effect of the observing geometry on the tomographic results is investigated by means of firstly answering the question in how far the number of GNSS sites and their horizontal distribution within the analyzed study area effect the accuracy and the precision of the tomographic results. This includes the questions i) if the site distributions should differ at different latitudes and ii) if the sites should be randomly distributed within the analyzed study area or rather situated along a regular grid. Secondly, the effect of the ray geometry and of the voxel discretization on the tomographic results is investigated. This is done by means of focusing on the question in how far the inclusion of rays of more satellites than given by the GPS orbits improves the reliability of the results and by investigating how large the tomographic voxels should be in order to yield reliable results within a changing orbit geometry.

The following indicates in which of the following sections the three main research questions stated at the beginning of this section are answered. At the end of each section, the answers to the respective research questions are summarized in a green box. While Section 7.1 refers to a real data set, Section 7.2 presents results obtained within different synthetic data sets.

#### Section 7.1: Real data set

- In how far does the inclusion of InSAR SWDs in addition to GNSS SWDs improve the tomographic reconstruction of the 3D refractivity field?
- Which of the two solution strategies LSQ and CS yields more accurate and more precise results?

#### Section 7.2: Synthetic data set comparable to real data set

- In how far does the inclusion of synthetic InSAR SWDs in addition to synthetic GNSS SWDs improve the tomographic reconstruction of the 3D neutrospheric water vapor field?
- Which of the two solution strategies LSQ and CS yields more accurate and more precise results?
- Which solution strategy is more flexible, i.e. less constraint-driven?

#### Section 7.2.2: Synthetic data set designed for analyzing the general effect of the observing geometry

- In how far do the number of synthetic GNSS sites and their horizontal distribution within a general study area have an effect on the accuracy and on the precision of the refractivity reconstruction by means of LSQ resp. CS?
  - Should the horizontal site distribution differ at different latitudes?
  - Should the horizontal site distribution within the analyzed study area be drawn at random from a uniform distribution or should the sites rather be situated along a regular grid?
- To what extent does a higher number of satellites than in the case of GPS improve the accuracy and the precision of the results obtained using the two solution strategies?
- In how far does the LSQ resp. the CS solution depend on the current position of the satellites within a certain orbit constellation?

#### Section 7.2.3: Synthetic data set designed for analyzing the effect of the orbits and of the voxel discretization

- In how far does the inclusion of the rays of more satellites than in the case of GPS improve the reliability of the results?
- How large should the tomographic voxels be in order to yield reliable results within a changing orbit geometry?

## 7.1 Real data set

This section presents a water vapor tomography solution obtained within the real data set available in Study Area 2 introduced in the Sections 6.1 and 6.2.5. This is done in order to investigate, for the real data set, in how far the inclusion of InSAR SWDs improves the tomographic reconstruction of the 3D neutrospheric refractivity field and in order to compare the LSQ and CS solution strategies. Section 7.1.2 tests the GNSS, InSAR, and radiosonde data used within the real data set for consistency. Only in case of consistent input data, the validation of the tomographic results obtained from GNSS and InSAR by means of radiosonde observations in Section 7.1.3 is possible. As neither the number nor the position of the GNSS sites can be influenced within the real data set, and as the validation of the tomographic results based on a single radiosonde site is challenging, the last of the three research questions itemized at the beginning of Chapter 7 is analyzed based on synthetic data only.

At the beginning of each section, an information box summarizes the observing geometry characteristics (real vs. artificial number and distribution of sites, ray directions, voxel discretization) as well as the input data (real vs. synthetic, GNSS only vs. GNSS and InSAR) of the data sets considered in that section. The information boxes shall serve as a means for better orientation of the reader, e.g. when thumbing through the thesis after a first read or when comparing the different analysis within this result section.

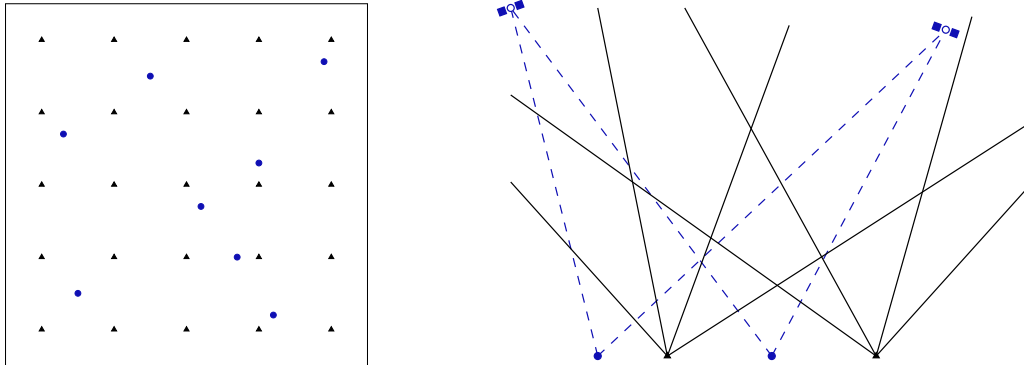
### 7.1.1 Tomographic settings in the real data set

The observing geometry and the input data of the tomographic setting used for reconstructing the water vapor distribution within Study Area 2 based on real SWD estimates are summarized in Table 7.1. As schematically illustrated in Figure 7.2, eight GNSS sites are available in the  $117 \times 122 \text{ km}^2$  large study area. A total of 36 InSAR sites is defined in the center of the  $6 \times 6$  ground voxels. Including the vertical direction, the analyzed atmospheric volume is subdivided into  $6 \times 6 \times 5$  voxels. For the GNSS sites, the ray directions are given by the GPS satellite positions around the SAR acquisition time, whereas in the case of the InSAR sites, artificial directions simulating a possible satellite geometry are defined, e.g. with azimuth angles  $A \in [0^\circ, 360^\circ[$  and elevation angles  $\epsilon \in [7^\circ, 90^\circ]$ .

**Table 7.1:** Observing geometry characteristics as well as input data for the real data set.

Real data set		
<b>Observing geometry and input data:</b>		
study area	about $117 \times 122 \text{ km}^2$ in the URG, containing SAR Track 22	
voxel discretization	$6 \times 6 \times 5$	
functional model $y = \Phi \cdot x$	inaccurate because $y$ is integrated along infinitesimally small steps of the ray path, while $\Phi$ and $x$ refer to a $6 \times 6 \times 5$ voxel discretization	
	<b>GNSS</b>	<b>InSAR</b>
number of sites	8	one per voxel, i.e. 36
site distribution	as given by the SAPOS <sup>®</sup> network	regular within voxel grid
ray directions	as given by the GPS satellites around the SAR acquisition time	artificial ray directions simulating a possible satellite geometry: $A \in [0^\circ, 360^\circ[$ , $\epsilon \in [7^\circ, 90^\circ]$
side rays	not included	
input data	GPS SWD estimates	absolute SWDs from InSAR

In the following, the reasoning yielding the  $6 \times 6 \times 5$  voxel discretization is explained. According to Table 3.1, the density of the available GNSS sites is comparably low w.r.t. the site densities encountered in other studies. The most comparable tomographic setting in previous research is described by Bi et al. [2006].



**Figure 7.2:** Schematic illustration of the real data set. The left plot schematically shows a top view on eight GNSS sites (blue dots) and on the regularly distributed InSAR sites (black triangles). The right plot shows a side view including the ray directions of the blue dashed GNSS rays that are given by the real GPS observing geometry and the ray directions of the black solid InSAR rays that are artificially defined.

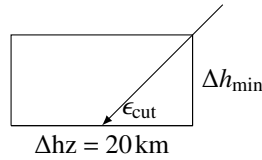
Bi et al. [2006] dispose of only four GNSS sites with maximum height differences of about 70m within their  $100\text{km}^2$  large study area. Horizontally, they discretize their study area into  $5 \times 6$  voxels of  $16 \times 20\text{km}^2$  size. They obtain best results at a constant thickness of the horizontal layers of 800m, when solving the tomographic system by means of LSQ adjustment with horizontal smoothing constraints as described in Section 5.3. In the vertical direction, instead of the exponential decay described in Section 5.3 and used in this section, Bi et al. [2006] add a priori information obtained from the average of three days of radiosonde observations.

In the setting described in this section, twice the number of GNSS sites than that in Bi et al. [2006] is available in an about 40% larger study area. I.e. the site density of the real data set described here is about 43% larger than in the tomographic approach described in Bi et al. [2006]. Therefore, the tomographic setting in Bi et al. [2006] is, on the one hand, worse than in the setting used in this section. On the other hand, Bi et al. [2006] dispose of a priori knowledge on the vertical refractivity decay from radiosondes, which is not available in this work. Although the horizontal voxel sizes of about  $20 \times 20\text{km}^2$  used for the real data set analyzed in this section are comparable to those in Bi et al. [2006], these voxel sizes are still much smaller than recommended by the rule of thumb of Champollion et al. [2004]. According to Champollion et al. [2004], the horizontal voxel sizes should be greater than or equal to the mean inter-site distance between the available GNSS sites, i.e. the voxel sizes should be greater than or equal to  $42 \times 42\text{km}^2$  in the data set analyzed in this section. Due to the small size of the analyzed study area and motivated by the goal of testing CS and its capability to do with fewer observations than LSQ for the tomographic solution, the rule of thumb of Champollion et al. [2004] is consciously broken.

In the vertical direction, five height layers are distinguished within the tomographic approach presented in this section. The lowest layer's thickness is set to about 1300m, and with increasing height, the height layer thicknesses increase from 1300m up to 2900m. The lowest layer's thickness is set to 1300m in order to ensure at least for rays with very low elevation angles that a ray arriving at the center of an about  $20 \times 20\text{km}^2$  large voxel is able to pass the horizontally neighboring voxel within the same height layer. As illustrated in Figure 7.3, this is only possible if the minimum thickness  $\Delta h$  of the height layers is related to the horizontal voxel size  $\Delta h_z = 20\text{km}$  and to the cutoff elevation angle  $\epsilon_{\text{cut}} = 7^\circ$  by means of

$$\Delta h_{\text{min}} = \frac{1}{2} \cdot \Delta h_z \cdot \tan \epsilon_{\text{cut}}. \quad (7.1)$$

The tomographic setting for the real data set does not include any side rays. In the real data set presented here, side rays cannot be approximated using radiosonde profiles, because the radiosonde profiles originating from the single radiosonde launch site available within Study Area 2 are used for validation. Moreover, according to Figure 6.4,



**Figure 7.3:** For rays with very low elevation angles close to  $\epsilon_{\text{cut}}$ , a ray arriving at the horizontal center of a voxel should be able to pass the neighboring voxel of the same height layer.

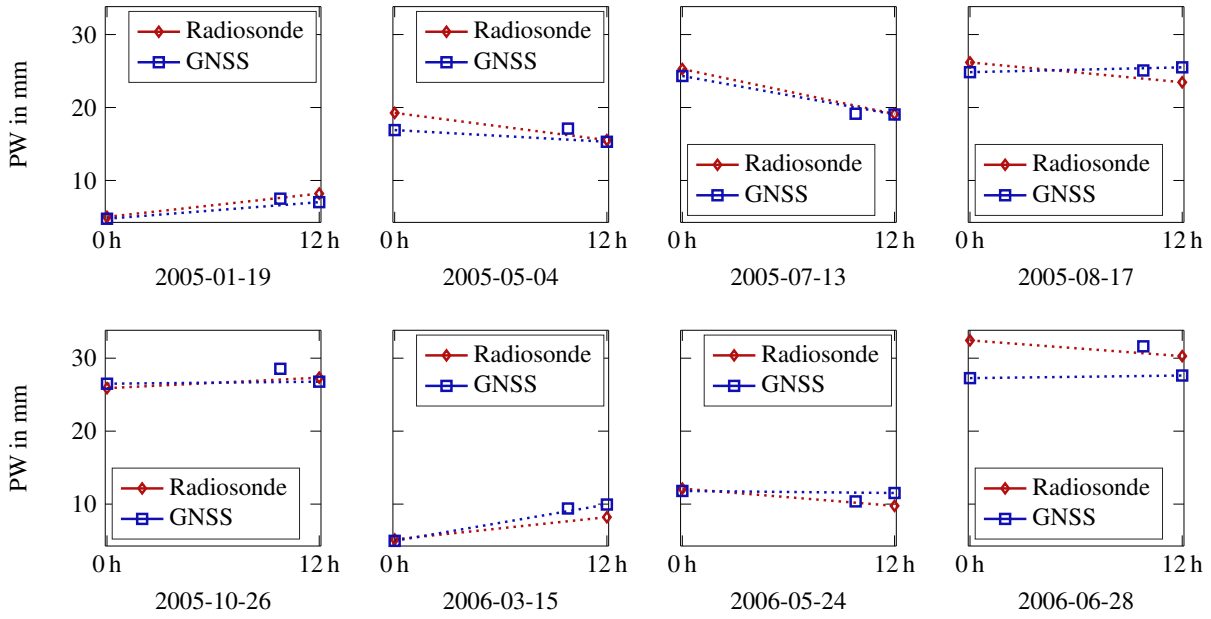
the available WRF data do not match well with the real weather patterns. Therefore, the possibility of computing the portion of the ray outside the study area using NWMs is rejected. Figure 6.4 also shows that the refractivity decay with height often does not follow an exponential law. As a consequence, side rays are not approximated by means of assuming an exponential refractivity decay with height. Finally, due to the limited size of the available SAR track, no two-step solution approximating side rays within the area of interest based on a pure top ray solution in a larger study area is implemented.

### 7.1.2 Consistency of GNSS, InSAR, and radiosondes

This section, which is based on Heublein et al. [2018], analyzes the quality of the GNSS and InSAR fusion yielding absolute water vapor maps and checks the GNSS and the radiosonde observations for consistency. These quality analyses need to be performed because i) the absolute ZWD maps from InSAR must match well with the GNSS ZWDs when introducing both GNSS and InSAR into the tomographic system and because ii) a validation using radiosondes is only possible if the radiosonde humidity information and that estimated from GNSS are consistent. Therefore, the InSAR ZWDs of all PS points situated within  $r_{\text{average}}$  around the available GNSS sites are averaged and compared with the respective GNSS ZWDs. If the mean difference of GNSS and InSAR ZWDs is less than 10 mm over all GNSS sites per acquisition date, i.e. less than twice the standard deviation of the input observations assumed within Section 5.3, the InSAR ZWDs are introduced into the tomographic system. In this work, this is the case for all acquisition dates except 2005-07-13 and 2006-06-28.

In addition, the precipitable water PW measured within the whole radiosonde profiles is compared with the PW derived from GNSS ZWDs. The radiosonde GNSS PW comparison is performed at 0UTC and at 12UTC, which correspond to the start times of the available radiosonde 10739 ascents over Stuttgart. As the distance between the Stuttgart radiosonde site and the Stuttgart GNSS site 0384 is about 6 km, the radiosonde ascent section should be covered by the GNSS geometry, even if the radiosonde does not ascend exactly vertically but is driven by winds. In addition, GNSS PW values have been computed for the SAR acquisition time at 9h48 UTC. This is done in order to get an idea of the humidity change between 0UTC and 12UTC.

As neutrospheric water vapor is highly variable in time and space, a validation of refractivities estimated around 9h48UTC by means of radiosonde observations at 0UTC and 12UTC is not optimal. However, a linear interpolation between the two radiosonde acquisition times is an acceptable option if i) the two sensors radiosonde and GNSS match well at both 0UTC and 12UTC, and if ii) a linear interpolation of the GNSS PW values at 0UTC and 12UTC is close to the GNSS PW value observed at 9h48UTC. In this context, a *good matching* and *close* PW values shall denote PW differences smaller than 2 mm between GNSS and the radiosonde at the three considered times of day. As shown in Figure 7.4, this is the case on 2005-01-19, 2005-07-13, 2005-10-26, 2006-03-15, and 2006-05-24 at the radiosonde site 10739 in Stuttgart. The accepted value of 2 mm PW difference is selected based on other studies comparing radiosonde and GNSS PW. The studies in Bock et al. [2005] and Niell et al. [2001] obtained mean PW differences between the two sensors of 1 mm to 2 mm, Bock et al. [2007] even 3 mm or more. The low temporal resolution as well as the unassured consistency of GNSS and radiosonde observations already indicate some weaknesses of a radiosonde validation. In addition, in reality, the radiosonde ascent takes some time, whereas the validation for this work assumes the radiosonde to take all the measures along the profile within a time instant.



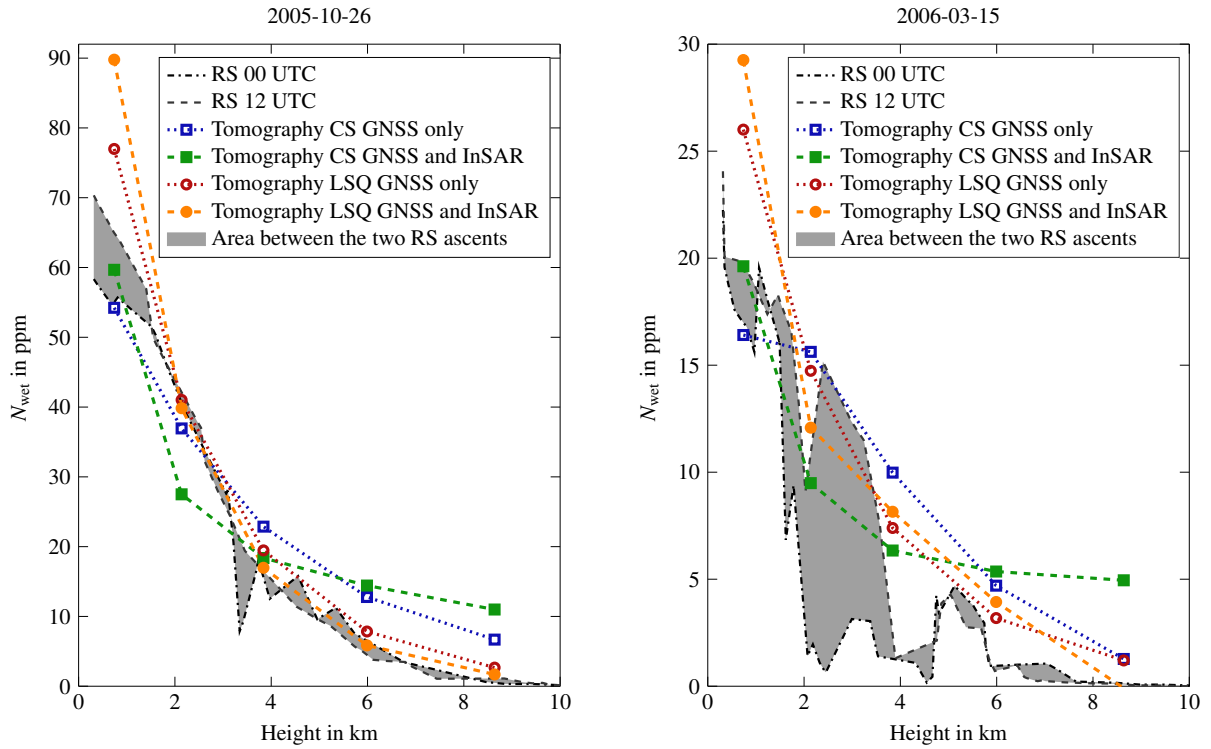
**Figure 7.4:** Precipitable water in mm from GNSS and from a radiosonde at the SAR acquisition dates of Track 22, according to Heublein et al. [2018]. The radiosonde PW values at 0UTC and 12UTC correspond to radiosonde ascents starting at these times above the Stuttgart radiosonde site 10739. In the case of GNSS, the PW values are deduced from GNSS ZWD estimates of 0UTC, 9h48UTC, and 12UTC. The dotted lines indicate linear interpolations between the sampling points at 0UTC and 12UTC. The shown PW range is equal for all subplots.

### 7.1.3 Validation of GNSS and InSAR based wet refractivities from CS and LSQ using radiosonde profiles

All in all, as in Heublein et al. [2018], to which this whole section refers, considering both the consistency of GNSS and InSAR as well as that of radiosonde and GNSS, the acquisition dates 2005-01-19, 2005-10-26, 2006-03-15, and 2006-05-24 remain for validation. Within these dates, the most resp. the least water vapor resides in the atmosphere on 2005-10-26 resp. on 2006-03-15. Therefore, these two dates representing different atmospheric states are selected for the radiosonde validation in this section. For these selected dates, Figure 7.5 shows the agreement between the wet refractivities reconstructed by means of LSQ or CS and the radiosonde profiles. The results shown in the figure are obtained based on GNSS rays corresponding to the real GPS orbit geometry around the SAR acquisition time and based on a set of artificial directions that simulate a possible satellite geometry, e.g. with azimuth angles  $A \in [0^\circ, 360^\circ[$  and elevation angles  $\epsilon \in [7^\circ, 90^\circ]$ . No prior knowledge from the surface meteorological site Stuttgart is included into the solution of the tomographic system. The height resp. the refractivity in Figure 7.5 are consciously plotted on the abscissae resp. on the ordinate, in order to make the exponential refractivity decay with height visibly similar to the Euler letters used in the height direction and introduced in Section 5.4.

The accuracy of the tomographic results is similar for LSQ and CS. For both LSQ and CS, the deviations of the tomographic solution from the exponential decay resp. from the linear combination of atoms based on Euler letters and Dirac letters in the height direction are not represented by the reconstructed refractivities. The LSQ solution overestimates the refractivity in the low atmospheric layers, while CS overestimates the refractivity in high altitudes on 2005-10-26 in the case of both GNSS only and GNSS and InSAR and on 2006-03-15 in the case of the GNSS only solution. However, as the radiosonde observations correspond, both temporally and locally, to other atmospheric snapshots than the GNSS and InSAR based tomographic results, no quantitative comparisons are drawn. Based on the real data set, no clear effect of including InSAR SWD observations as described in Section 6.2.5 is visible. Yet, the results depend on the directions defined for the synthetic InSAR rays. If the directions of the synthetic InSAR rays are varied, the solution obtained using both GNSS and InSAR changes.





**Figure 7.5:** Wet refractivity in ppm on 2005-10-26 and 2006-03-15. The tomographic refractivity estimates are based on GNSS SWDs only resp. on GNSS and InSAR SWDs. The tomography solution corresponds to 9h48 UTC. The temporal window of GNSS SWDs introduced into the tomographic system is set to  $\pm 15$  min.

In order to derive reliable conclusions from the real data set, a high number of different synthetic InSAR ray directions should be tested. Moreover, the validation should be performed for many SAR acquisition dates and at more than only a single radiosonde site. Table 7.2 summarizes the results obtained within the real data set.

When compared to the real data set presented in Heublein et al. [2018], the main difference consists in the much larger height layer thicknesses. In Heublein et al. [2018], the thickness of the lowest height layer is set to 500m, whereas here, it corresponds to 1300m. As a consequence, in this work, the voxel discretization error is much larger and the functional model from Equation 4.5 is less accurate than in Heublein et al. [2018]. While the observations still result from a continuous integration of the slant wet delay along the whole ray path, the design matrix discretizes the analyzed atmospheric volume into much coarser voxels than in Heublein et al. [2018].

**Table 7.2:** Research questions answered using the real data set.

<b>Real data set</b>	
<b>Research questions:</b>	<b>Answers based on the real data set:</b>
<ul style="list-style-type: none"> <li>• In how far does the inclusion of InSAR SWDs in addition to GNSS SWDs improve the tomographic reconstruction of the 3D refractivity field?</li> <li>• Which of the two solution strategies LSQ and CS yields more accurate and more precise results?</li> </ul>	<ul style="list-style-type: none"> <li>• In the analyzed real data set, no clear effect of the inclusion of InSAR SWDs in addition to the GNSS is visible.</li> <li>• In the analyzed real data set, both solution strategies yield comparable results.</li> </ul>

## 7.2 Synthetic data set

When compared to the real data set presented in Section 7.1, one advantage of synthetic data sets consists in the possibility to perform a 3D validation of the estimated wet refractivity field w.r.t. the WRF input data. In addition, the observing geometry can be flexibly designed within the synthetic data sets, i.e. the number and the distribution of the synthetic GNSS sites as well as the ray directions can be varied. Finally, a clear focus on the reconstruction qualities of LSQ and CS within the considered tomographic settings is possible, if no measurement noise is added to the observations. Different synthetic data sets are defined in this section in order to test different tomographic settings. For each tomographic setting, the corresponding synthetic data sets are obtained from WRF simulations of the neutrospheric wet refractivity.

### 7.2.1 Synthetic data set comparable to real data set

Table 7.3 characterizes the synthetic data set used in this section. In terms of observing geometry, this synthetic data set is comparable to the real data set of Section 7.1.3. The analyzed study area (Study Area 1) is about 34 % smaller than the study area used in the real data set (Study Area 2). While eight GNSS sites are available in the real data set, in the synthetic data set in this section, seven synthetic GNSS sites are defined at the positions at which real GNSS sites are available in Study Area 1. The study area is discretized into  $5 \times 5 \times 5$  voxels, and in the center of each of the 25 horizontal ground voxels, one synthetic InSAR site is defined. As in the real data set and as illustrated in Figure 7.2, the ray directions of the synthetic GNSS sites are again given by the GPS orbit geometry, whereas artificial ray directions simulating a possible satellite geometry with azimuth angles  $A \in [0^\circ, 360^\circ[$  and elevation angles  $\epsilon \in [7^\circ, 90^\circ]$  are used for the synthetic InSAR sites. The synthetic GNSS and the synthetic InSAR SWDs are deduced from WRF as described in Section 6.3.2.

**Table 7.3:** Observing geometry characteristics as well as input data of the synthetic data set on which the results of this section are based. In terms of observing geometry, the synthetic data set in this section is comparable to the real data set of Section 7.1.

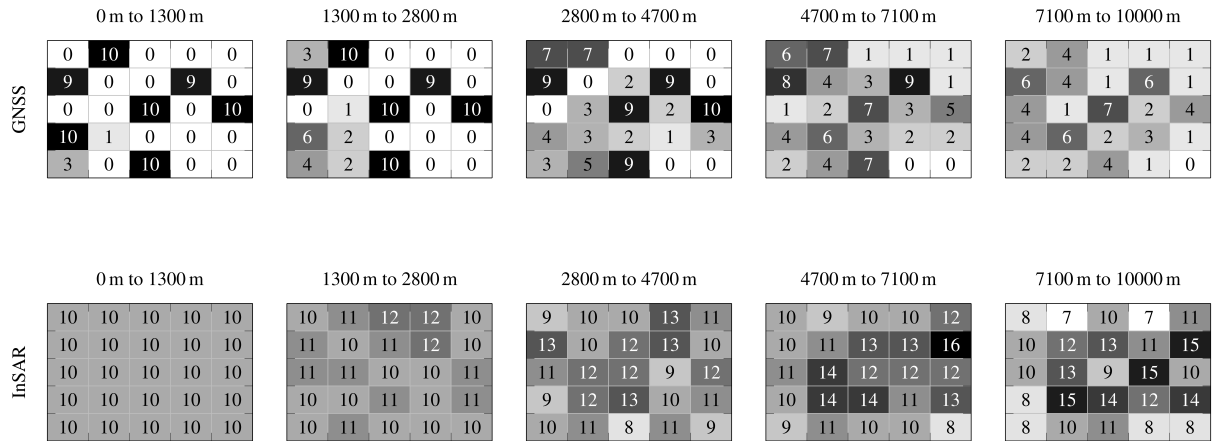
#### Synthetic data set comparable to real data set

##### Observing geometry and input data:

study area	95 × 99 km <sup>2</sup> in the URG	
voxel discretization	5 × 5 × 5	
functional model	accurate because $\mathbf{y}$ is results from WRF refractivities deduced within	
$\mathbf{y} = \Phi \cdot \mathbf{x}$	the same 5 × 5 × 5 discrete voxels to which $\Phi$ and $\mathbf{x}$ refer	
	<b>synthetic GNSS</b>	<b>synthetic InSAR</b>
number of sites	7	one per voxel, i.e. 25
site distribution	as given by the SAPOS <sup>®</sup> network	regular within voxel grid
ray directions	as given by the GPS satellites around the SAR acquisition time	artificial ray directions simulating a possible satellite geometry: $A \in [0^\circ, 360^\circ[$ , $\epsilon \in [7^\circ, 90^\circ]$
side rays	not included	
input data	synthetic GPS SWD estimates deduced from WRF	absolute SWDs from synthetic InSAR, deduced from WRF

When compared to the synthetic data set used in Heublein et al. [2018], the main difference consists in the about twice as thick height layers. Therefore, as stated in Section 7.1.3, the voxel discretization error is much larger than in Heublein et al. [2018]. Yet, as the observations in the synthetic data set do not result from a continuous integration of the slant wet delay along the whole ray path but are obtained along discrete ray path elements within the  $5 \times 5 \times 5$  voxels, the functional model from Equation 4.5 is accurate, even though the voxels are comparably thick. Therefore, the synthetic data set enables a clear focus on the reconstruction qualities of LSQ and CS, independently of inaccuracies in the functional model and independently of noise that would be encountered in the case of real SWD estimates.

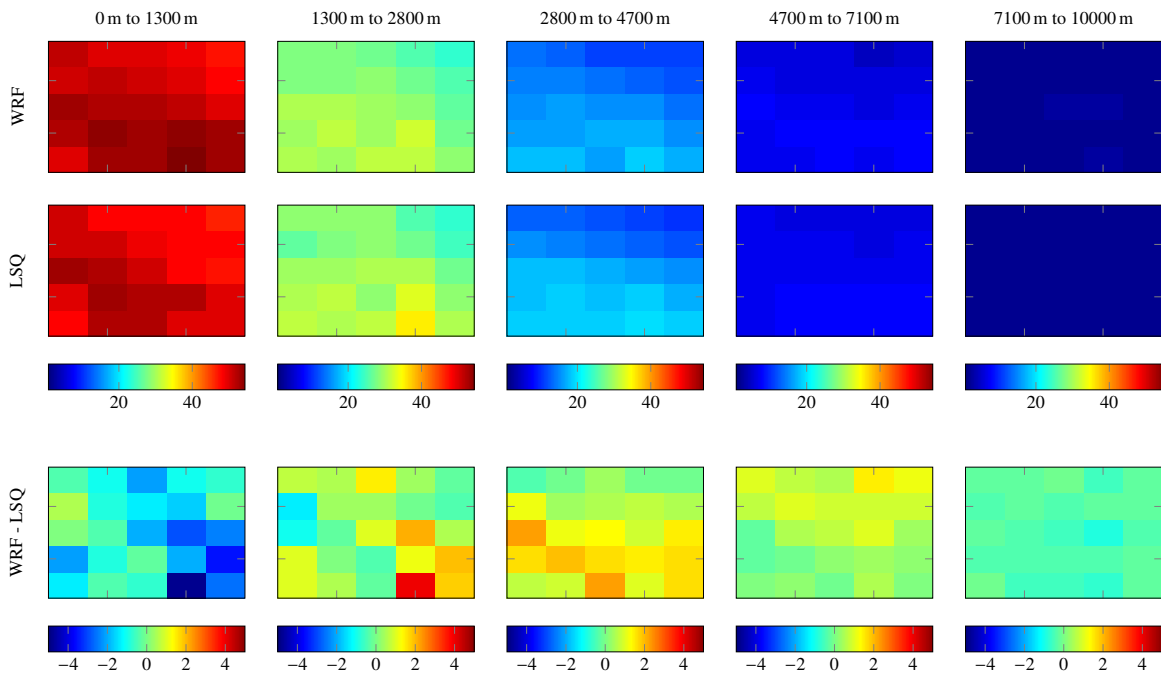
For the different acquisition dates of Study Area 1, a total of 48 to 68 GNSS observations are available. Considering one acquisition date in more detail, a total of 68 rays are available for the seven GNSS sites available within Study Area 1 on 2005-01-03. The upper plot in Figure 7.6 shows how many GNSS rays cross the tomographic voxels on that date. Due to the cone-shaped GNSS observing geometry, most of the voxels close to the surface are crossed by much fewer rays than voxels in the higher tomographic layers. However, if a low voxel is crossed, the number of rays passing through it is larger than in higher atmospheric layers. On 2005-01-03, the percentage of crossed voxels increases from 36% to 96% from the lowest to the highest layer. If synthetic InSAR observations are added, the number of crossed voxels increases, as shown in the lower plot of Figure 7.6.



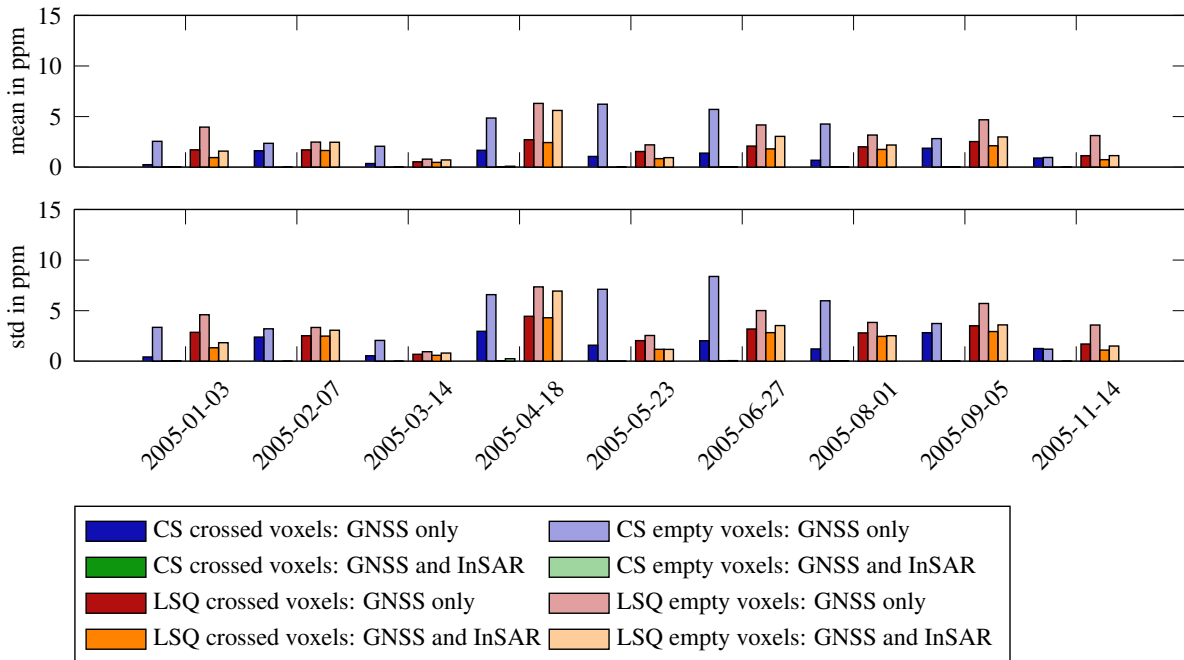
**Figure 7.6:** Number of rays crossing the tomographic voxels on 2005-01-03 within the synthetic data set. The upper plot shows the number of rays in the synthetic GNSS only observing geometry. The lower plot shows the number of additional rays from the synthetic InSAR observing geometry. The  $5 \times 5$  voxels per height layer correspond to five voxels in latitude and to five voxels in longitude. Longitude increases along the abscissae, latitude along the ordinate. Above the plots, the heights of the layers are given. The number of rays crossing a voxel is represented by the color of the voxel. Dark voxels correspond to voxels that are crossed by many rays, white voxels are not crossed by any ray. The higher a layer is situated within the tomographic grid in the upper plot, the more voxels per layer are crossed. Many of the voxels in the lowest layers are not crossed by any ray. However, if a low voxel is crossed, then the number of rays passing through it is larger than in higher atmospheric layers. In this work, ten synthetic rays are defined for each of the 25 synthetic InSAR sites.

In order to give an idea of the horizontal and vertical refractivity variations and the capability to tomographically reconstruct them using LSQ, a layer-wise comparison of the refractivities estimated by means of LSQ from both synthetic GNSS and synthetic InSAR SWDs and from WRF is given for 2005-05-23 in Figure 7.7. The refractivity differences between the tomographic reconstruction and the WRF data decrease with increasing height layers. This can be explained both by the decrease of the absolute value of  $N_{\text{wet}}$  with height and by the increase of rays per voxel observed in most voxels when reaching higher atmospheric layers.

The accuracy of the estimated refractivities w.r.t. the number of rays crossing the respective voxels is presented in Figure 7.8. As expected, for both synthetic GNSS only and synthetic GNSS and InSAR, the refractivities within crossed voxels are more accurately and more precisely reconstructed than the refractivities within empty voxels.



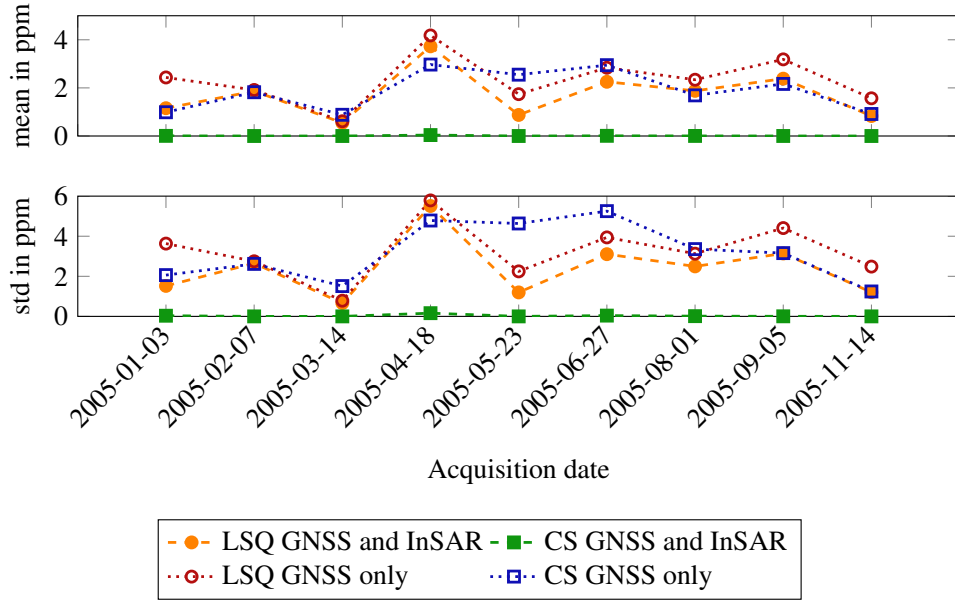
**Figure 7.7:** Plot of 2005-05-23 layer-wise WRF refractivities, estimated refractivities from LSQ, and the differences between the estimation and the WRF refractivities in ppm. The  $5 \times 5$  square-size voxels per height layer correspond to five voxels in latitude and to five voxels in longitude. Longitude increases along the abscissae, latitude along the ordinate. Above the plots, the heights of the layers are given. The estimates are deduced from both synthetic GNSS and synthetic InSAR SWDs.



**Figure 7.8:** Comparison of the reconstruction accuracy within crossed voxels and voxels that are not crossed by any GNSS rays. The absolute mean value and the standard deviation (std) of the differences between estimated and WRF refractivities are shown for the different acquisition dates.

When considering the differences between the tomographically reconstructed refractivities and those given by WRF independently of the number of rays crossing the respective voxels, Figure 7.9 shows that

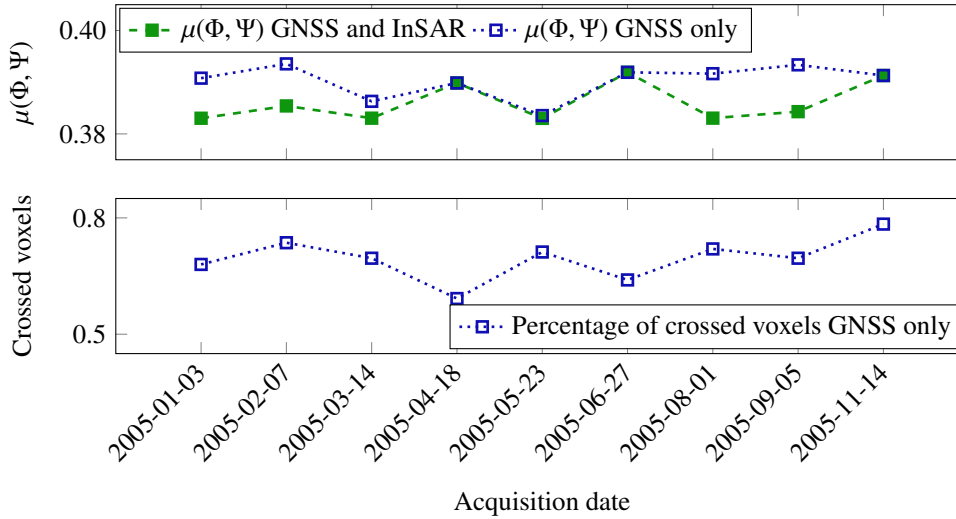
- on all acquisition dates, adding synthetic InSAR SWDs to the synthetic GNSS SWDs slightly resp. clearly improves the accuracy and the precision of the refractivities reconstructed using LSQ resp. CS (maximum improvement over all dates in i) accuracy of 2.1 ppm resp. 3.0 ppm and in ii) precision of 2.1 ppm resp. 5.2 ppm for LSQ resp. CS)
- when using both synthetic GNSS SWDs and synthetic InSAR SWDs, the use of CS clearly improves the reconstruction accuracy and precision on all analyzed acquisition dates (mean differences and standard deviations below 0.1 ppm), yet
- when using synthetic GNSS SWDs only, using CS instead of LSQ only slightly improves the accuracy resp. the precision of the refractivity estimates by up to 1.4 ppm resp. 1.5 ppm on about half the dates.



**Figure 7.9:** Mean and std of the differences between estimated refractivities and WRF refractivities over all voxels. The upper plot shows the mean of the absolute difference between estimated refractivities and WRF refractivities over all voxels, the lower plot shows the std of that difference. The dashed lines serve for better following the variation of the mean resp. of the std on the analyzed acquisition dates, but only the discrete values indicated by the markers at the acquisition dates should be evaluated.

When interpreting Figure 7.9, the seasonal variability of water vapor has to be taken into account. In the case of increasing humidity in the summer time, i.e. in the case of larger absolute values of  $N_{\text{wet}}$ , the reconstruction quality decreases. Consequently, larger values are obtained for the mean difference and the standard deviation of the difference between the WRF refractivities and the reconstructed refractivities. In addition, when evaluating the CS results, the coherences between the design matrix  $\Phi$  and the sparsifying basis  $\Psi$ , defined in Section 5.4, should be analyzed for the different settings. Figure 7.10 shows that the coherence  $\mu(\Phi, \Psi)$  remains equal or gets smaller when adding synthetic InSAR SWDs in addition to the GNSS SWDs.

The smaller the coherence, the better the  $L_1$ -norm reconstruction is expected to perform. However, adding synthetic InSAR SWDs does not systematically decrease the coherence. On 2005-04-18, the coherence is not effected by the inclusion of synthetic InSAR SWDs. Thus, in the analyzed setting, the very accurate results obtained when adding synthetic InSAR SWDs to the synthetic data set are not correlated with the coherence between the design matrix and the sparsifying basis. As the same dictionary is used at all acquisition dates, the coherence values at the investigated dates vary hardly. The coherence variations observed on the different acquisition dates only result from the variations in  $\Phi$ , i.e. from the changing ray geometry. Table 7.4 summarizes the results obtained based on the synthetic data set analyzed in this section.



**Figure 7.10:** Coherence  $\mu(\Phi, \Psi)$  between the design matrix  $\Phi$  and the sparsifying basis  $\Psi$ , and percentage of voxels crossed by rays in a synthetic GNSS only solution. The dash-dotted lines serve for better following the variation of the mean resp. of the std on the analyzed acquisition dates, but only the discrete values indicated by the markers at the acquisition dates should be evaluated.

## 7.2.2 Effect of the observing geometry on the tomographic results within a general study area

The work in [Bi et al., 2006] points out that the geometric configuration of the network has a significant impact on the accuracy of tomographic water vapor reconstructions using GNSS SWD estimates. Therefore, this section analyzes the general effect of the observing geometry on the accuracy and on the precision of the tomographic results. Table 7.5 summarizes the observing geometry characteristics as well as the input data of the synthetic data set used for this analysis of the effect of the observing geometry on the tomographic results.

### 7.2.2.1 Observing geometry settings

For Bi et al. [2006], the *geometric configuration of the network* mainly includes the voxels' size and a uniform distribution of the sites within the voxels. However, in a generalizing way, not only the components of the geometric configuration of the network mentioned by Bi et al. [2006], but the whole *observing geometry* has an effect on the tomographic solution. In this context, the *observing geometry* described in this section may be composed of e.g.

- the prevailing topography within the study area,
- the current GNSS satellite orbits,
- the horizontal extension of the study area,
- the upper boundary of the study area,
- the cutoff elevation angle used,
- the decision for rays entering the study area on its top only resp. for rays entering the study area both on its top and on its side,
- the horizontal and vertical resolution of the voxel grid, i.e. the voxel sizes in longitude, latitude, and height,
- the orientation of the voxels with respect to the North-South or the East-West directions,
- the number of sites situated within the study area,
- and the distribution of these sites within the study area.

Concentrating on selected items of the above components of the observing geometry, this section analyzes the importance of the number of synthetic GNSS sites and the effect of the horizontal distribution of these sites within

**Table 7.4:** Research questions answered using a synthetic data set which is, in terms of observing geometry, comparable to the real data set.

<b>Synthetic data set comparable to real data set</b>	
<b>Research questions:</b>	<b>Answers based on the analyzed synthetic data set:</b>
<ul style="list-style-type: none"> <li>• In how far does the inclusion of synthetic InSAR SWDs in addition to synthetic GNSS SWDs improve the tomographic reconstruction of the 3D neutrospheric water vapor field?</li>   <li>• Which of the two solution strategies LSQ and CS yields more accurate and more precise results?</li>   <li>• Which solution strategy is more flexible, i.e. less constraint-driven?</li> </ul>	<ul style="list-style-type: none"> <li>• In the analyzed synthetic data set, the inclusion of synthetic InSAR SWDs slightly resp. clearly improves the tomographic reconstruction of the 3D water vapor distribution when using LSQ resp. CS. In the case of LSQ, including synthetic InSAR SWDs decreases the mean difference by 0.0 ppm to 1.3 ppm and the standard deviation by 0.1 ppm to 2.1 ppm. In the case of CS, including synthetic InSAR SWDs decreases the mean difference by 2.5 ppm to 2.9 ppm to absolute values below 0.1 ppm and the standard deviation by 1.5 ppm to 5.3 ppm to absolute values below 0.2 ppm.</li>   <li>• In the analyzed synthetic data set, when using both synthetic GNSS and synthetic InSAR SWDs, CS yields more accurate and more precise results, attaining mean differences and standard deviations of the difference between estimated refractivities and WRF refractivities below 0.2 ppm. If synthetic GNSS SWDs only are introduced into the tomographic system, both solution strategies yield comparable results with mean accuracies over all voxels between 0.6 ppm to 4.2 ppm and with mean precisions over all voxels between 0.8 ppm and 5.8 ppm.</li>   <li>• In the analyzed setting, CS is more flexible. For a <math>5 \times 5 \times 5</math> voxel discretization and based on the available synthetic GNSS and synthetic InSAR SWDs, the dictionary composed of iDCT, Euler, and Dirac atoms is able to accurately represent the WRF refractivity field.</li> </ul>
<p>Based on both synthetic GNSS and synthetic InSAR SWDs, very accurate results are obtained for all investigated acquisition dates. Thus, these results should be transferable to any other acquisition date, if the tomographic settings (accurate functional model, total number of sites, voxel discretization, and variety of ray directions) are comparable.</p>	

the considered study area w.r.t. the quality of the refractivity reconstruction by means of LSQ resp. CS. The section aims at answering the questions to what extent the horizontal site distribution should differ at different latitudes, and in how far the site positions should be drawn at random from a uniform distribution or rather be situated along a regular grid. Furthermore, the section investigates to what extent a higher number of satellites than in the case of GPS improves the accuracy and the precision of the results obtained using the CS and LSQ solution strategies and in how far the LSQ resp. the CS solution depend on the current position of the satellites within a certain orbit constellation.

In order to answer the stated research questions, two  $100 \times 100 \text{ km}^2$  large study areas in high and mid latitudes (latitudes  $\varphi = 70^\circ$ , corresponding to the northern border of Norway, and  $\varphi = 49^\circ$ , e.g. in the URG region) are defined. Within these study areas, a large number of observing geometry settings is analyzed. As shown in the decision tree in Figure 7.12, the observing geometry settings result from the combination of i) four different discretizations of each of the ii) two study areas into voxels of constant refractivity, iii) uniform or non-uniform height layer thicknesses, iv) three different numbers of sites, v) two synthetic GNSS orbit geometries and vi)

**Table 7.5:** Observing geometry characteristics as well as input data of the synthetic data set used for the analysis of the effect of the observing geometry on the tomographic results in this section.

**Synthetic data set: observing geometry**

**Observing geometry and input data:**

study area	95 × 99 km <sup>2</sup> in the URG at mid-latitudes ( $\varphi = 49^\circ$ ) and equal-size study area at high latitudes ( $\varphi = 70^\circ$ ) corresponding to northern Norway
voxel discretization	3 × 3 × 5, 4 × 4 × 5, 6 × 6 × 5, or 9 × 9 × 5
functional model	both accurate and inaccurate functional model tested, i.e. both based on observations
$y = \Phi \cdot x$	computed within fine and within coarse voxel discretization
<b>synthetic GNSS for observing geometry analysis</b>	
number of sites	9, 16, or 25
site distribution	eight different site distributions as shown in Figure 7.11, all sites situated in a plane
ray directions	as given by GPS-like orbit geometries containing six resp. nine orbital planes with five resp. nine satellites each
side rays	included
input data	synthetic GPS SWD estimates deduced from WRF

eight different distributions of the considered number of sites within the respective study area. For each of the defined tomographic settings, a synthetic SWD data set is deduced from a synthetic 3D refractivity field from WRF referring vii) to one of the coarse voxel discretizations defined above resp. to a fine 36 × 36 × 20 voxel discretization, and the refractivity parameters are estimated by means of viii) LSQ and CS.

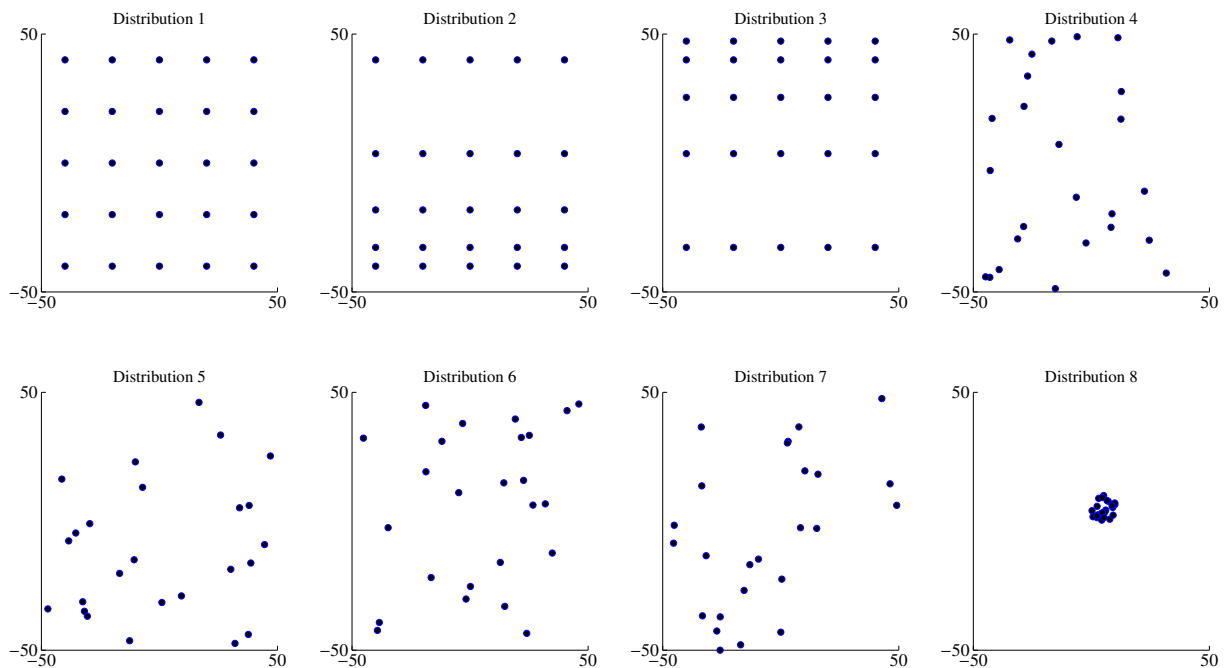
In the following, further details are given on the observing geometry parameters itemized at the beginning of the section and to some extent included within the defined tomographic settings. Firstly, the topography is reduced to zero in this theoretical study even though this is known to degrade the accuracy of the tomographic solution. This reduction to zero height is performed for the sake of transferability to any possible study area. The GNSS satellite orbits are approximated by two synthetic GPS-like orbit geometries composed of six resp. nine orbital planes inclined by 55° with respect to the equator as in the case of GPS. Per plane, a total number of five resp. nine satellites, uniformly distributed over the plane, is defined. That is, one of the two orbit constellations is quite similar to a real GPS orbit constellation with about 30 satellites distributed over six orbital planes (in the case of GPS, there would be four to six satellites per plane), whereas in the other orbit constellation, a total of 81 satellites, comparable to a combination of GPS, Global'naya Navigatsionnaya Sputnikova Sistema (GLONASS), and Galileo is defined. There are two main motivations for considering also the second orbit constellation disposing of more satellites. Firstly, the work aims at giving general rules for future GNSS tomographies, including both GPS and other GNSS, i.e. using the rays originating from more satellites on more planes than in the case of GPS. In addition, the analysis of the effect of the horizontal site distributions and of the voxel discretization shall be carried out devoid of any unwanted signal caused by the orbit geometry. Here, such unwanted signals are assumed to be reduced if many satellites are regularly distributed over the elevation and azimuth angles visible from the site.

The upper boundary of the study area was determined based on the synthetic input wet refractivity data set. As low elevations enable the rays to cross different voxels already in small heights, the cutoff elevation angle is set to  $\epsilon_{\text{cut}} = 7^\circ$ , even though real SWD data would, at such a low elevation, suffer more from site specific effects like multipath. Both rays entering the study area on its top and on its side are included in order to compensate the missing topography and in order to better resolve the vertical water vapor distribution. Independently of the selected number of sites, the horizontal resolution of the voxel grid was set to about 33 km (3 × 3 × 5 voxels), 25 km (4 × 4 × 5 voxels), 17 km (6 × 6 × 5 voxels), and 11 km (9 × 9 × 5 voxels). The thicknesses of the vertical

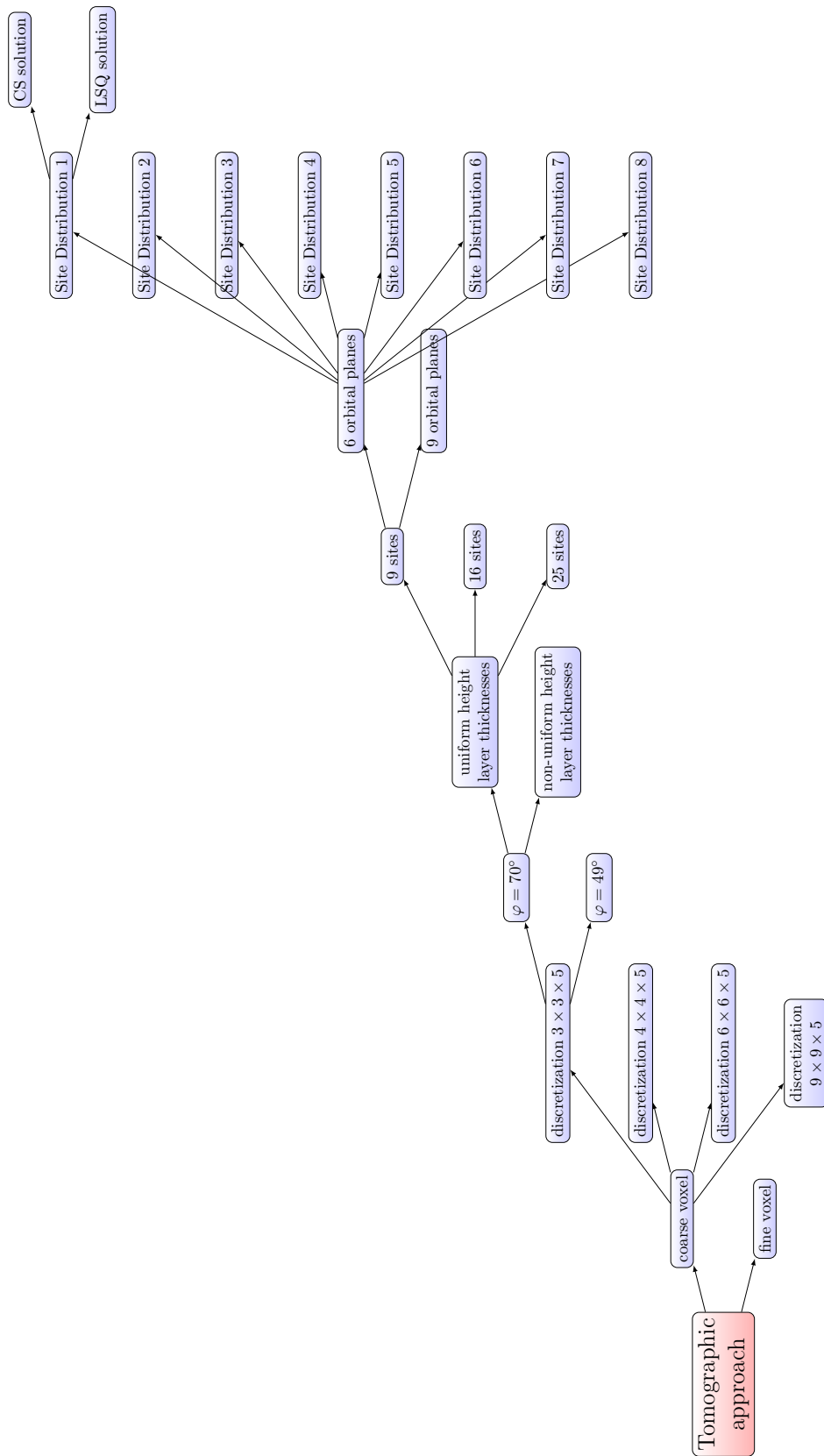


layers are set to five times 2000 m in the case of uniform height layer thicknesses resp. to 1000 m, 1500 m, 2000 m, 2500 m, and 3000 m in the case of non-uniform height layer thicknesses, increasing with increasing height above the surface. Finally, the rectangular voxels are oriented along the North-South resp. East-West direction. The inner voxel refractivity within each voxel is assumed to be constant. A total of nine, 16, or 25 sites are defined. In the case of the  $25 \times 25 \text{ km}^2$  voxel size, the rule of thumb of Champollion et al. [2004] recommending, for the case of LSQ, mean inter-site distances equal to the horizontal resolution, holds, if 16 sites are introduced into the system. Yet, in most of the other settings (fewer sites at the same number of voxels resp. the same numbers of sites but more voxels), the rule of thumb is not respected anymore. The maximum number of 25 sites is selected in order to obtain a similar site density as within the dense Japanese GPS Earth Observation Network (GEONET), see Nishimura et al. [2011].

As shown in Figure 7.11, besides a regular distribution, two site distributions with regular spacing between the sites in the East-West direction and decreasing resp. increasing spacing of the sites in the North-South direction are defined. These two distributions are generated in order to investigate in how far a closer spacing of the sites in the northern part of the study area improves the tomography accuracy in the respective study areas in mid and high latitudes. Furthermore, four site distributions drawn at random from a uniform distribution and spreading the sites within the whole study area, are defined. Finally, one site distribution drawn at random from a uniform distribution, but spreading the sites only within a small part at the center of the whole study area, is defined. Initially, the described site distributions shown in Figure 7.11 are defined in a local East North Up (ENU) coordinate system. Thereafter, they are transformed into global geodetic coordinates (longitude, latitude, height). This procedure ensures that similar site distributions are compared in the different latitude study areas.



**Figure 7.11:** Site distributions of 25 sites in a local ENU coordinate system. The abscissae and the ordinate show the sites' distances in km from the center of the considered study area. Besides a regular distribution, two site distributions with regular spacing between the sites in the East-West direction and decreasing resp. increasing spacing of the sites in the North-South direction are defined. These two distributions are generated in order to investigate in how far a closer spacing of the sites in the northern part of the study area improves the tomography accuracy in the respective study areas in mid and high latitudes. Furthermore, four site distributions drawn at random from a uniform distribution and spreading the sites within the whole study area are defined. Finally, one site distribution drawn at random from a uniform distribution, but spreading the sites only within a small part at the center of the whole study area, is defined.



**Figure 7.12:** Schematic representation of a decision tree for the tomographic approach options within the site distribution analysis. Within this schematic representation of the decision tree, children are only included at the first node of each level, even though at each level, each node could be continued in the same way as that shown for the first node.

The synthetic data sets for the described observing geometry settings are generated as follows. The design matrix  $\Phi$  is deduced by direct raytracing along the azimuth and elevation angles of the described synthetic orbits. As indicated by the first level of the decision tree from Figure 7.12, the design matrix is derived from a coarse  $p \times q \times k$  voxel discretization with  $p = q \in \{3, 4, 6, 9\}$  and  $k = 5$  from above. In contrast, the SWD observations  $\mathbf{y}$  of rays entering the study area from both the top and the side are computed either at a coarse  $p \times q \times k$  voxel discretization or at a fine  $36 \times 36 \times 20$  voxel discretization. According to Section 4.3, the functional model building the basis for the tomographic reconstruction is only exact if the parameters are estimated at the voxel discretization that has been used for deducing the observations and the design matrix. This could suggest to only deduce both the design matrix and the observations from the coarse voxel discretizations at which the refractivity estimation shall be carried out. However, if the observations used for the four defined voxel discretizations are derived from the four corresponding discretizations, then, the observations for each discretization differ, and the accuracy and the precision of the derived refractivity parameters cannot be compared within the different discretizations. As a consequence, the observations are also deduced from a fine  $36 \times 36 \times 20$  discretization, even though Section 4.3 warns that the functional model from Equation 4.3 will not be exact anymore, if the observations refer to another discretization than the design matrix and the estimated parameters.

Deducing the observations from the synthetic refractivity data set enables a direct comparison of the later estimated 3D water vapor field with the reference data available from WRF. When comparing the estimation quality of different voxel discretizations, all refractivity estimates are, independently of the voxel discretization used for the estimation, assigned to the fine  $36 \times 36 \times 20$  voxel level. The reference data from WRF are, for both the  $\varphi = 70^\circ$  and the  $\varphi = 49^\circ$  study area, derived from simulations of the WRF model for the  $\varphi = 49^\circ$  study area. I.e. the synthetic data sets do not represent the latitude dependent meteorological conditions of the two study areas. Instead, they only serve as a means for geometrically analyzing the performance of the different site distributions at different latitudes, independently of the prevailing weather. The study areas' topography is reduced to zero by means of setting the WRF DEM to a constant value. The mid-latitude WRF simulation can be transferred to the high latitude study area by setting the latitude of the WRF data to the respective study area's latitude and by stretching resp. compressing the longitudes associated with the WRF simulation as a function of the latitude of the considered study area.

### 7.2.2.2 Evaluation of the refractivity estimates

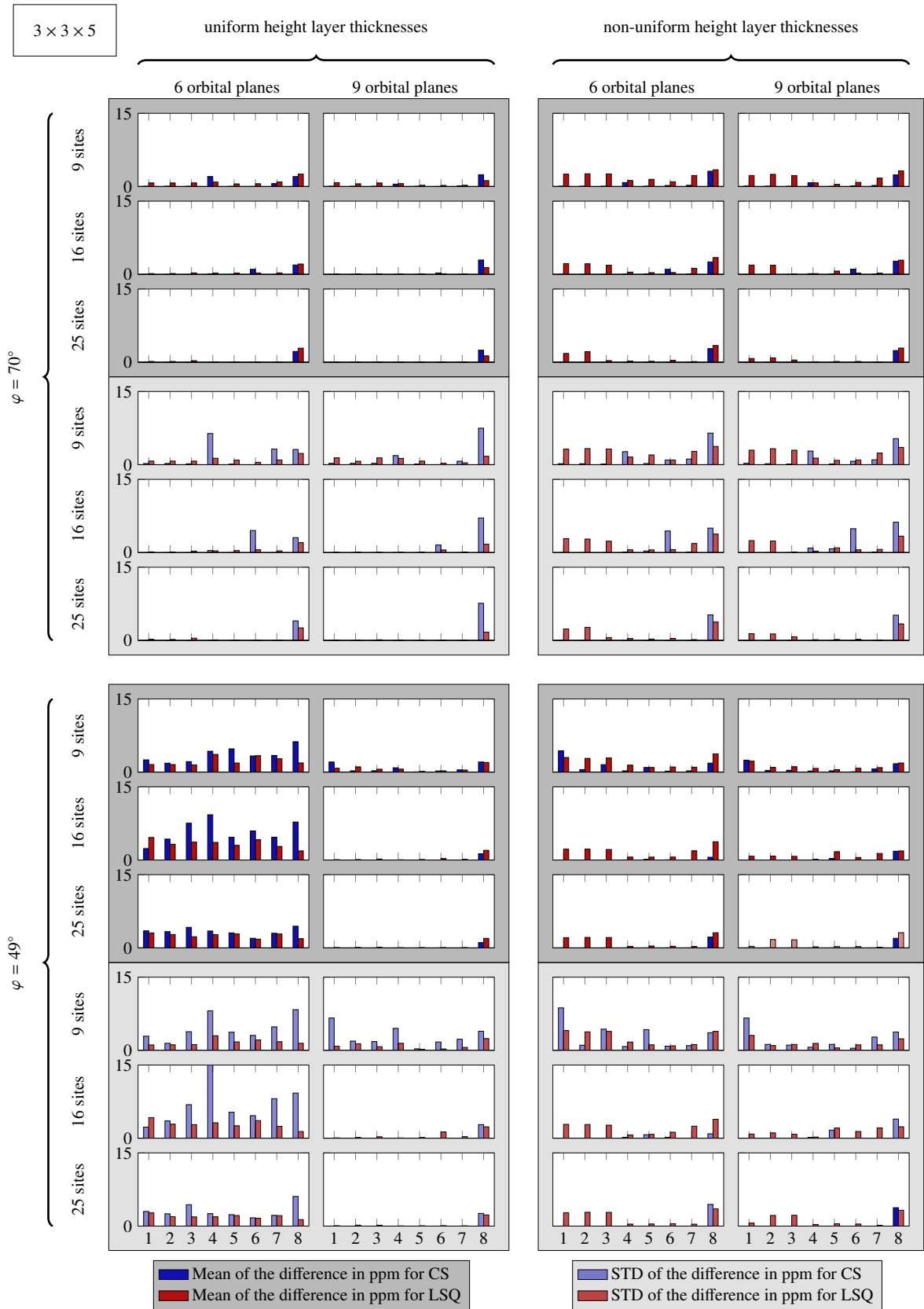
In order to evaluate the performance of the different observing geometry settings described in Section 7.2.2.1, the mean of the absolute difference and the standard deviation of the difference between the estimated refractivities and the input WRF refractivities are analyzed w.r.t. the parameters distinguishing the observing geometry settings. Figure 7.13 and Figure 7.14 show the mean of the absolute difference and the standard deviation of the difference between estimated refractivities and WRF input refractivities, obtained at a  $3 \times 3 \times 5$  resp. at a  $9 \times 9 \times 5$  voxel discretization. As both the observations  $\mathbf{y}$  and the design matrix  $\Phi$  refer to the same voxel discretization at which the wet refractivity field is estimated, the functional model is accurate. Figure 7.13 and Figure 7.14 distinguish the LSQ resp. CS results obtained within the two analyzed study areas in high and mid latitudes, with uniform and non-uniform height layer thicknesses, based on nine, 16, and 25 sites, and using six resp. nine orbital planes. As Figure 7.13 and Figure 7.14 refer to observations deduced based on different voxel discretizations ( $3 \times 3 \times 5$  resp.  $9 \times 9 \times 5$ ), the figures should not be compared with each other. Comparisons should only be drawn within the subplots of each figure.

When comparing the LSQ resp. CS solution strategies, CS shows a larger variability than LSQ in the accuracy and the precision of the results obtained within the different tomographic settings. In the case of CS, in some tomographic settings, much larger mean values and standard deviations of the differences between the estimated refractivities and the WRF refractivities occur than in the case of LSQ. Namely, when considering the  $9 \times 9 \times 5$  voxel discretization, maximum absolute mean differences of up to 9.8 ppm and standard deviations of up to 9.6 ppm are observed for CS, whereas the maximum absolute mean differences resp. standard deviations in the case of LSQ attain values of 3.7 ppm resp. of 4.3 ppm. In the  $3 \times 3 \times 5$  voxel discretization, maximum absolute mean differences

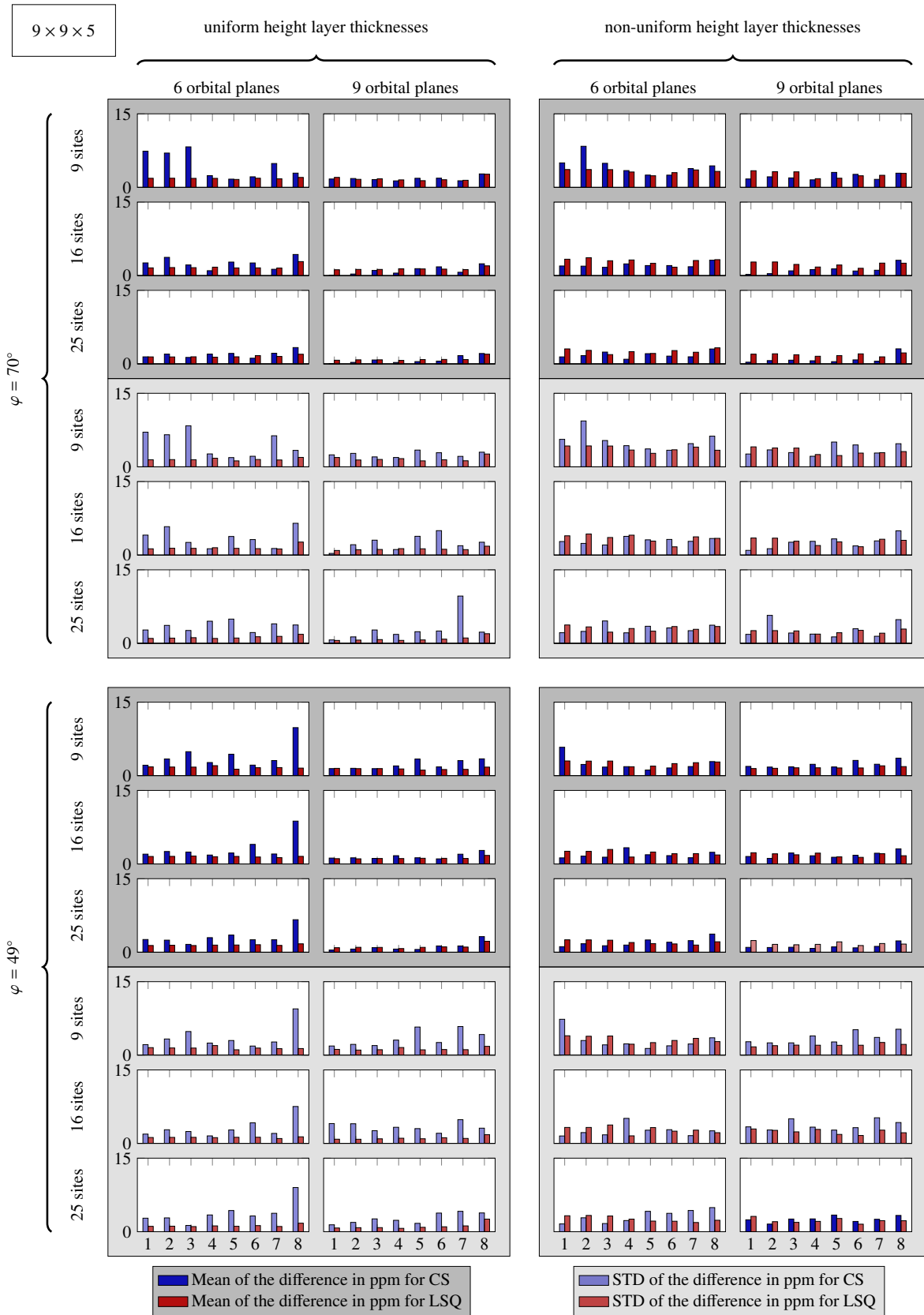
of up to 9.3 ppm and standard deviations of up to 16.3 ppm are observed for CS, whereas the maximum absolute mean differences resp. standard deviations in the case of LSQ attain values of 4.6 ppm resp. of 4.2 ppm. In many of the analyzed settings, the lowest accuracies and the lowest standard deviations are obtained for Site distribution 8, but there are also settings in which Site distribution 8 performs better than other site distributions, e.g. in the case of  $3 \times 3 \times 5$  voxels at a latitude of  $\varphi = 49^\circ$ , uniform height layer thicknesses, and six orbital planes. Independently of the voxel discretization, the wet refractivity estimates in the  $\varphi = 49^\circ$  study area do not show a higher accuracy or a higher precision than the estimates deduced from the  $\varphi = 70^\circ$  study area. No clear effect of using uniform or non-uniform height layer thicknesses is detected. While introducing the observations of more than nine GNSS sites improves the quality of the reconstructed refractivity field in some of the analyzed settings, there are also settings in which a higher number of GNSS sites does not yield a more accurate or more precise solution, e.g. in the case of  $3 \times 3 \times 5$  voxels, a latitude of  $\varphi = 49^\circ$ , uniform height layer thicknesses, and six orbital planes. Analogously, defining nine orbital planes with a total of 81 satellites instead of six orbital planes with a total of 30 satellites often increases the accuracy and the precision of the estimated wet refractivity field, even though the increased number of satellites does not improve the quality of the solution e.g. in the case of the Site Distributions 3 resp. 4 for LSQ resp. CS in the  $9 \times 9 \times 5$  voxel discretization at a latitude of  $\varphi = 49^\circ$ , a total of 16 sites, and non-uniform height layer thicknesses.

In order to investigate the differences between different voxel discretizations, the refractivity estimates of the respective coarse voxel discretizations could be compared at the fine  $36 \times 36 \times 20$  voxel discretization level. As a total of  $36 \cdot 36 \cdot 20 = 25920$  values is difficult to visualize, Figure 7.15, shows, averaged over all fine voxels, the absolute mean difference and the standard deviation of the mean difference between the refractivity estimates obtained at the coarse voxel discretizations and the WRF refractivities at a fine  $36 \times 36 \times 20$  voxel discretization. This is possible because the fine  $36 \times 36 \times 20$  voxel discretization represents a subdivision of the considered coarse voxel discretizations. The figure shows that the accuracy resp. the precision of the estimated refractivities at the different coarse voxel discretizations w.r.t. the  $36 \times 36 \times 20$  WRF refractivities is not systematically related to the roughness of the voxel discretization. While the  $9 \times 9 \times 5$  voxel discretization yields the smallest absolute mean differences (3.5 ppm resp. 3.4 ppm in the case of CS resp. LSQ) in the case of nine synthetic GNSS sites, six orbital planes, and Orbit Sample 1, its refractivity estimates are less accurate than those from the  $3 \times 3 \times 5$  voxel discretization e.g. in the case of nine synthetic GNSS sites, nine orbital planes, and any of the two orbit samples. In addition, Figure 7.15 illustrates that the accuracy and the precision of the refractivity estimates varies with the considered orbit sample. The figure shows that the observations of 25 synthetic GNSS sites do not necessarily yield more accurate and more precise refractivity estimates than the observations of nine synthetic GNSS sites. When considering the results obtained from six orbital planes, Orbit Sample 2, the CS resp. LSQ solution based on 25 synthetic GNSS sites is 2.9 ppm resp. 0.2 ppm less accurate than the solution based on nine synthetic GNSS sites. Finally, a benefit of using 81 satellites distributed over nine orbital planes instead of 30 satellites distributed over six planes is only visible in some of the analyzed settings. E.g. in the case of nine synthetic GNSS sites, a  $4 \times 4 \times 5$  voxel discretization, and Orbit Sample 2, the CS reconstruction accuracy decreases by 0.5 ppm when using nine orbital planes instead of six orbital planes. In addition, Figure 7.15 illustrates that the accuracy and the precision of the refractivity estimates varies with the considered orbit sample. The figure shows that the observations of 25 synthetic GNSS sites do not necessarily yield more accurate and more precise refractivity estimates than the observations of nine synthetic GNSS sites. When considering the results obtained from six orbital planes, Orbit Sample 2, the CS resp. LSQ solution based on 25 synthetic GNSS sites is 2.9 ppm resp. 0.2 ppm less accurate than the solution based on nine synthetic GNSS sites. Finally, a benefit of using 81 satellites distributed over nine orbital planes instead of 30 satellites distributed over six planes is only visible in some of the analyzed settings. E.g. in the case of nine synthetic GNSS sites, a  $4 \times 4 \times 5$  voxel discretization, and Orbit Sample 2, the CS reconstruction accuracy decreases by 0.5 ppm when using nine orbital planes instead of six orbital planes.

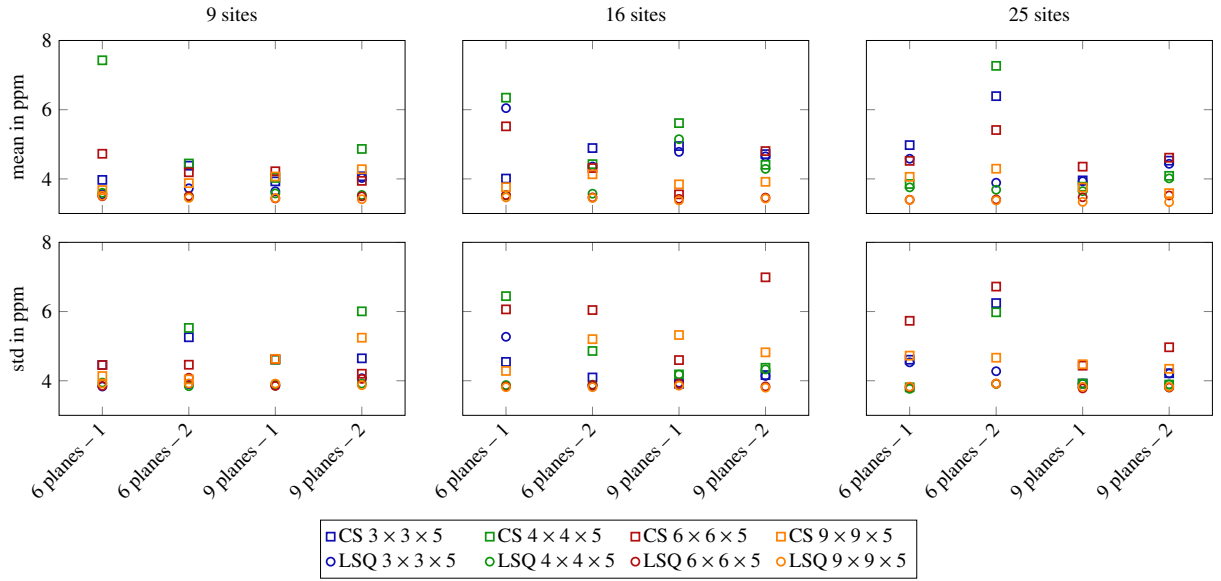
In order to further investigate the effect of the orbit geometry, independently of the voxel discretizations, a second sample of each of the two orbit geometries (GPS-like with six orbital planes and six satellites per planes resp. GPS-like with nine orbital planes and nine satellites per plane) is considered. For Site Distribution 1 and Site Distribution 8, Figure 7.16 resp. 7.17 exemplary visualize voxel-wise differences between the estimated refractivities and the WRF refractivities at different tomographic settings disposing of nine resp. 25 synthetic GNSS sites.



**Figure 7.13:** Mean of the absolute difference and std of the difference between estimated refractivities and WRF input refractivities. The observations, the validation data set from WRF, and the estimated parameters all refer to a  $3 \times 3 \times 5$  voxel discretization.



**Figure 7.14:** Mean of the absolute difference and std of the difference between estimated refractivities and WRF input refractivities. The observations, the validation data set from WRF, and the estimated parameters all refer to a  $9 \times 9 \times 5$  voxel discretization.



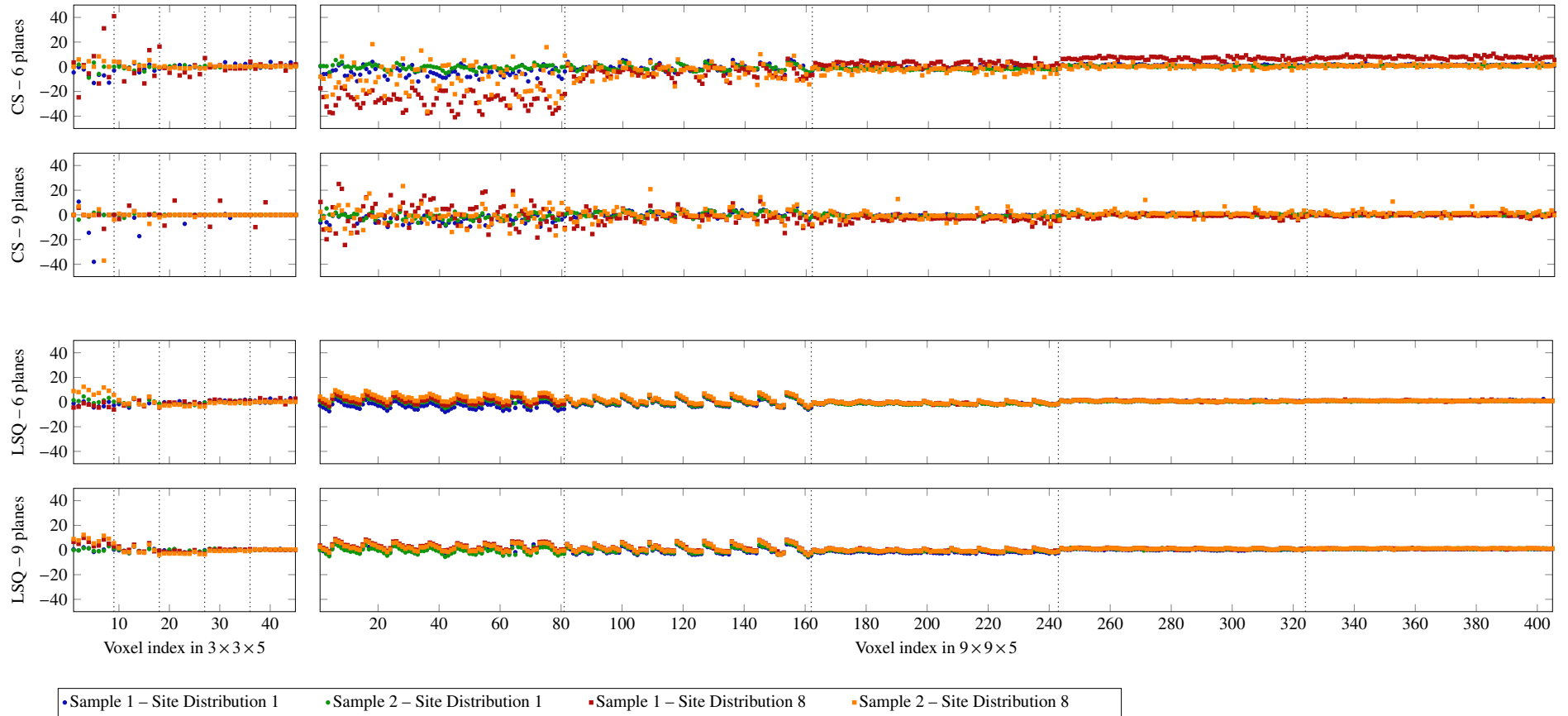
**Figure 7.15:** Absolute mean differences in ppm and standard deviations in ppm of the differences between estimated refractivities per coarse voxel discretization and the WRF refractivities at the fine  $36 \times 36 \times 20$  voxel discretization level for Site Distribution 1. The xticklabels indicate the number of orbital planes determining the ray directions as well as the index of the considered orbit sample.

The input SWDs are deduced from the two indicated coarse voxel discretizations ( $3 \times 3 \times 5$  and  $9 \times 9 \times 5$ ). The refractivity estimation is carried out at the same coarse voxel discretizations.

When considering the results in Figure 7.16 referring to nine synthetic GNSS sites, CS accurately estimates the refractivity of many voxels at the  $3 \times 3 \times 5$  voxel discretization level, but for some voxels, the estimates deviate by up to 40 ppm from the WRF refractivities. In the case of LSQ, the maximum differences between the estimated and the WRF refractivities are much smaller and attain only about 15 ppm. If ray directions originating from nine orbital planes and nine satellites per plane are used, the refractivity of less voxels is inaccurately estimated by CS than in the case of only six orbital planes with five satellites per plane. Yet, in the analyzed orbit samples, the maximum absolute differences do not significantly decrease by means of adding orbital planes and satellites. In the case of Site Distribution 8, the voxel refractivities can be more accurately estimated than in the case of Site Distribution 1. However, also in the case of Site Distribution 1, the estimates of some voxels' refractivities are completely thrown off in the case of CS. When applying a LSQ solution, the results show a smoother behavior and smaller maximum absolute differences w.r.t. the WRF refractivities.

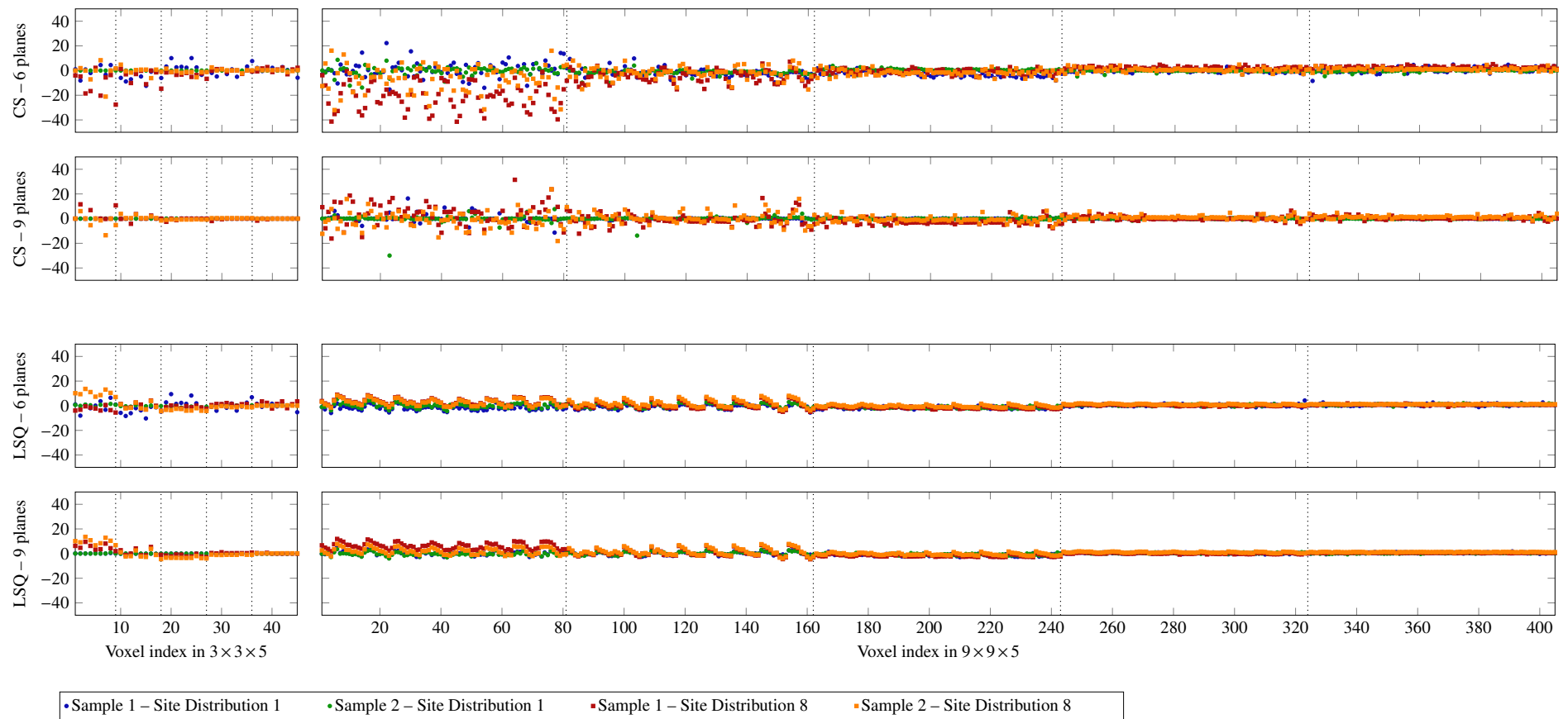
If the results obtained at the  $3 \times 3 \times 5$  voxel discretization are compared with those of the  $9 \times 9 \times 5$  voxel discretization, the CS estimates obtained based on Site Distribution 8 scatter widely, while the maximum differences in the case of Site Distribution 1 approximately range in the interval  $[-10 \text{ ppm}, 10 \text{ ppm}]$ . The LSQ estimates again show a much smoother behavior than the CS estimates. The LSQ absolute differences attain similar maximum values as those obtained by CS. A similar difference pattern is visible for all site distributions. This suggests that the LSQ solution at the  $9 \times 9 \times 5$  voxel discretization is dominated by the horizontal smoothing constraints.

At a first glance, a comparison of Figure 7.16 and Figure 7.17, representing the results deduced from nine synthetic GNSS sites and from 25 synthetic GNSS sites, shows a similar behavior of the reconstruction accuracies for both numbers of sites. However, at a closer look, the CS refractivity estimates at the  $3 \times 3 \times 5$  voxel discretization become accurate (maximum absolute differences below 0.1 ppm) for Site Distribution 1 resp. at least less inaccurate for Site Distribution 8 when introducing 25 sites instead of nine sites and when using, at a time, ray directions towards a total of 81 satellites placed on nine orbital planes instead of only 30 satellites on six orbital planes. In the case of the  $9 \times 9 \times 5$  voxel discretization and Site Distribution 1, the absolute differences between the CS refractivity estimates and the WRF refractivities decrease in most voxels when using a higher variety of ray directions. Yet, in some voxels, CS does not seem to be able to accurately estimate the water vapor content.



**Figure 7.16:** Differences in ppm per voxel between the estimated refractivities and the WRF refractivities at different tomographic settings disposing of nine GNSS sites. The input SWDs are deduced from the indicated coarse  $3 \times 3 \times 5$  resp.  $9 \times 9 \times 5$  voxel discretizations. The reconstruction accuracy of Site Distribution 1 and Site Distribution 8 are compared for two samples per orbit geometry, i.e. for two samples of a GPS-like orbit geometry with six orbital planes and five satellites per plane resp. for two samples of a GPS-like orbit geometry with nine orbital planes and nine satellites per plane. The dotted vertical lines indicate height layer borders. The voxel indexing is performed as described in Section 4.1.





**Figure 7.17:** Differences in ppm per voxel between the estimated refractivities and the WRF refractivities at different tomographic settings disposing of 25 GNSS sites. The input SWDs are deduced from the indicated coarse  $3 \times 3 \times 5$  resp.  $9 \times 9 \times 5$  voxel discretizations. The reconstruction accuracy of Site Distribution 1 and Site Distribution 8 are compared for two samples per orbit geometry, i.e. for two samples of a GPS-like orbit geometry with six orbital planes and five satellites per plane resp. for two samples of a GPS-like orbit geometry with nine orbital planes and nine satellites per plane. The dotted vertical lines indicate height layer borders. The voxel indexing is performed as described in Section 4.1.

As the voxels with inaccurate refractivity estimates vary from Orbit Sample 1 to Orbit Sample 2 even in the case of nine orbital planes, the CS solution strategy shows a high sensibility to the introduced ray directions. I.e. the assumption that unwanted orbit effects are reduced by introducing observations of many regularly distributed satellites can only be partly confirmed by the investigated settings. In many of the analyzed settings, a higher number of satellites improves the reconstruction quality. Yet, even in the case of an orbit geometry with many orbital planes and many satellites per plane, the refractivity estimates of different samples of that orbit geometry differ significantly. Table 7.6 summarizes the answers to the research questions investigated in this section.

**Table 7.6:** Research questions answered using a synthetic data set which is, in terms of observing geometry, comparable to the real data set.

**Synthetic data set: observing geometry**

- In how far do the number of synthetic GNSS sites and their horizontal distribution within a general study area have an effect on the accuracy and on the precision of the refractivity reconstruction by means of LSQ resp. CS?
- The horizontal distribution of the synthetic GNSS sites only shows an effect on the accuracy and on the precision of the refractivity reconstruction if the total number of synthetic GNSS sites within a general study area is clustered in a small part of that study area. In some of the analyzed tomographic settings, depending on the voxel discretization and the number of synthetic GNSS sites introduced into the tomographic model, such a clustered site distribution does not even perform worse than other site distributions. The presented research does not suggest to use different site distributions at different latitudes. The presented analyses suggest that site distributions drawn at random from a uniform distribution and regularly distributed sites perform equally well.
- To what extent does a higher number of satellites than in the case of GPS improve the accuracy and the precision of the results obtained using the two solution strategies?
- The accuracy improvement and the increase in precision obtained by means of an increased number of satellites on more orbital planes depends, for both LSQ and CS, on the voxel discretization and on the number of introduced synthetic GNSS sites. In coarse voxels, the improvement attained by introducing rays of more satellites is larger than in fine voxels.
- In how far does the LSQ resp. the CS solution depend on the current position of the satellites within a certain orbit constellation?
- The orbit geometry and the voxel discretization significantly influence the accuracy and the precision of the estimated refractivity field. This may be partly due to the fact that the synthetic GNSS sites introduced into the presented analysis are all situated in a horizontal plane. Therefore, future research should not focus on a general analysis of the horizontal site distribution's effect but rather concentrate on determining an ideal site distribution for a specific study area with given topography.

The presented results are obtained for two flat study areas in mid and high latitudes and should be transferable to any other study area with comparable tomographic settings (comparable site density, voxel sizes, and variety of ray directions). The conclusion that the orbit geometry and the voxel discretization significantly influence the accuracy and the precision of the estimated refractivity field should be valid for any study area.

### 7.2.3 Effect of the orbits and of the voxel discretization in a specific study area

Section 7.2 applied LSQ and CS to both a synthetic GNSS only and a synthetic GNSS and InSAR data set. When compared to Heublein et al. [2018], due to the coarser voxel discretization chosen in the height direction, CS showed to be capable of yielding very accurate results if synthetic InSAR SWDs at artificial ray directions were included into the tomographic system. Yet, in the case of the data set using synthetic GNSS SWDs only, neither LSQ nor CS were capable to accurately reconstruct the 3D refractivity field. Therefore, and as Section 7.2.2 demonstrated the effect of the voxel discretization on the quality of the tomographic results, the current section aims at determining a voxel discretization at which the refractivity field can be accurately reconstructed even in case of the small GNSS site density considered in Section 7.2.

Moreover, based on a small number of orbit geometry settings, Section 7.2.2 revealed the significant effect of the orbit geometry on the accuracy and on the precision of the tomographic results. The current section investigates a much larger number of orbit samples in order to analyze the information content of changing orbit geometries. The section aims at answering the questions

- to what extent the inclusion of observations of more GNSS than GPS (e.g. GNSS-like orbits with nine orbital planes and nine satellites per plane instead of GPS-like orbit constellations with six orbital planes with about five satellites per plane) improves the *repeatability* of the results and
- how coarse the voxels have to be discretized, if the tomographic estimates shall be *repeatable* within a changing orbit geometry.

In this context, the term *repeatable* shall denote that when repeating the estimation in changing orbit geometries, constant refractivity values shall be estimated for the different voxels. Similarly to the information boxes in the previous sections, Table 7.7 summarizes the characteristics of the synthetic data sets used in this section. Two different data sets are used in order to investigate the two stated research questions. In both data sets, the synthetic GNSS sites are distributed at the position of the real GNSS sites within Study Area 1 as in Section 7.2.

The smaller the voxels, the larger the uncertainty of the estimation is expected to be, because then, more parameters need to be determined based on a constant number of observations. In turn, as discussed in Section 4.3, the larger the voxels, the larger is the averaging caused by assuming constant inner voxel refractivities. I.e. the section aims at finding a trade-off between the roughness in spatial resolution and the solution's uncertainty. From a different point of view, this trade-off could also be considered as a compromise between the risk of over-smoothing the inner-voxel refractivities due to too large voxels and the risk of the over-smoothing that the horizontal constraint in the LSQ adjustment may generate in case of too small voxels.

While the term *uncertainty* often mainly refers to mean differences or standard deviations with respect to some validation data set, it shall, here, also include information theoretic uncertainty measures. The information theoretic view on the uncertainty of the tomographic results can be motivated by the fact that in the case of CS, depending on the tomographic settings, large outliers were observed in some voxels in Section 7.2.2. While outliers have a strong effect on the standard deviation and the mean difference analyzed in the previous sections, the information content resp. the entropy, that are in this section included into the uncertainty considerations, are not distorted by outliers. I.e. an information theoretic approach to the tomographic reconstruction of the 3D water vapor field is presented that analyzes the information contained in varying orbit geometries. After introducing information theoretic basics in Section 7.2.3.1, Section 7.2.3.2 designs three different settings for analyzing the information theoretic effect of varying orbit geometries on the uncertainty of the tomographic results at different voxel discretizations. Finally, Section 7.2.3.3 presents the results.

**Table 7.7:** Observing geometry characteristics as well as input data of the synthetic data set used for the information theoretic analysis of the effect of the orbits and the voxel discretization on the tomographic results.

**Synthetic data set: orbits and voxel size**

**Observing geometry and input data:**

study area	95 × 99 km <sup>2</sup> in the URG at mid-latitudes ( $\varphi = 49^\circ$ )	
voxel discretization	selected discretizations out of 2 to 8 voxels in longitude and latitude, and out of 2 to 16 voxels in height	
functional model $\mathbf{y} = \Phi \cdot \mathbf{x}$	inaccurate because $\mathbf{y}$ is summed up within 16 × 16 × 16 voxels, whereas $\Phi$ and $\mathbf{x}$ refer to coarser voxel discretizations	
	<b>synthetic GNSS for information theoretic approach</b>	
number of sites	7	
site distribution	as given by the SAPOS <sup>®</sup> network	
input data	synthetic GPS SWD estimates deduced from WRF	
side rays	included	
ray directions	case a) of Figure 7.18 as given by many samples of a GPS-like orbit geometry containing nine orbital planes with nine satellites each	case c) of Figure 7.18 as given by the GPS satellites around the SAR acquisition time

**7.2.3.1 Information theoretical basics**

According to Nearing et al. [2016], *information* is “the property of a signal that effects a change in our state of belief about some hypothesis.” Consequently, information can be considered as a measure of surprise. If an event is expected to happen, there is no surprise once it occurs. However, if an improbable event occurs, this is a big surprise, i.e. the data revealing this surprise contain a lot of information. In general, the information  $\tilde{I}(\chi)$  of some random variable  $\chi$  is inversely related to the probability of occurrence  $P(\chi)$  of an event  $\chi$ :

$$\tilde{I}(\chi) = \frac{1}{P(\chi)} \quad (7.2)$$

The probability of occurrence  $P(\chi)$  of a certain *event* is obtained by means of subdividing the data space, i.e. the range of possible values of the random variable, into a certain number  $n_{\text{bins}}$  of bins.

In order to avoid very large values of  $\tilde{I}(\chi)$  in case of low probabilities of occurrence  $P(\chi)$  in Equation 7.2, the information content  $I(\chi)$  is generally defined considering a logarithmic function of  $\tilde{I}(\chi)$ . As the information content is usually expressed in bits, the base of the logarithm is set to 2 and the information content is

$$I(\chi) = \log_2 \left( \frac{1}{P(\chi)} \right) \text{ bits} = -\log_2 P(\chi) \text{ bits.} \quad (7.3)$$

This also means that  $I(\chi)$  corresponds to the number of binary questions that had to be asked in a single random experiment to guess the outcome from randomly drawing from a known distribution. If the same experiment is repeated many times with random variables  $\nu$  drawn from a distribution  $\mathcal{X}$ , then the *entropy*  $E(\chi)$  is the average value of information:

$$E(\chi) = - \sum_{\nu \in \mathcal{X}} P(\nu) \cdot \log_2 P(\nu) \text{ bits} \quad (7.4)$$

As the standard deviation or the variance, the entropy can be considered as a measure of uncertainty of a random variable. However, in contrast to the standard deviation taking into account the values of the data, the entropy only depends on the probability distribution of the data.

The entropy is minimum resp. maximum, if either only one single value is attained or if all values have the same probability. If only one single value is attained, then the probabilities for each bin except one are zero, and one single bin within the data space has the probability

$$P_{\text{dirac}} = 1. \quad (7.5)$$

Consequently, the resulting minimum possible entropy corresponds to

$$E_{\text{min}}(\chi) = 0 \text{ bits}. \quad (7.6)$$

In contrast, if all values have the same probability, then the probability for each value is

$$P_{\text{uniform}} = \frac{1}{n_{\text{bins}}}. \quad (7.7)$$

The resulting maximum possible entropy is

$$E_{\text{max}}(\chi) = \log_2 n_{\text{bins}} \text{ bits}. \quad (7.8)$$

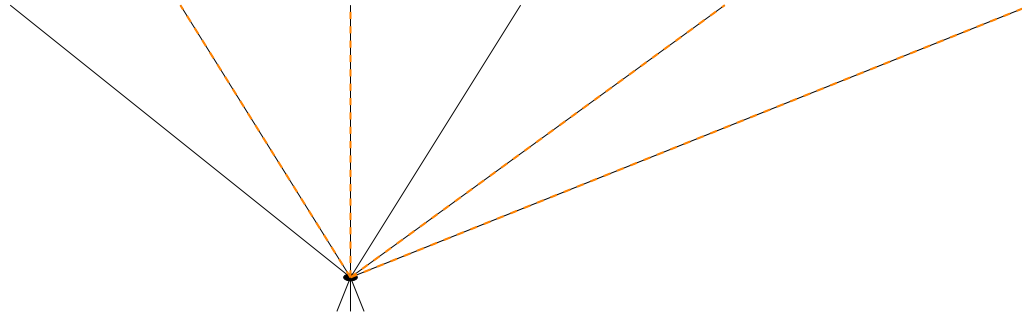
As a consequence, the maximum entropy is a function of the data binning.

### 7.2.3.2 Information theory in tomographic water vapor reconstruction – Methodology

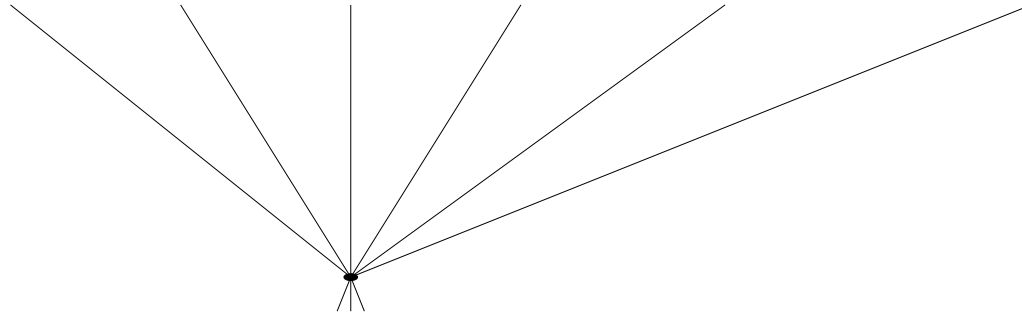
In a tomographic reconstruction of water vapor, an information theoretic uncertainty analysis can only be carried out if probability distributions of the estimated parameters are available. Such probability distributions would be automatically obtained if e.g. probability distributions for the SWD estimates were introduced into the tomographic system. However, the noise-free synthetic SWD estimates from WRF introduced within this analysis are single values without any probability distribution. Therefore, some kind of randomness has to be generated in this section in order to obtain probability distributions. Figure 7.18 proposes three different ways to create randomness in this work.

Figure 7.18 a) poses the question in how far the uncertainty of the estimated refractivities depends on the amount of data (i.e. rays) introduced into the adjustment. A total of nine GPS-like orbital planes with nine evenly distributed satellites per plane are defined, out of which one or two thirds are selected at random for reconstructing the atmospheric water vapor distribution. When selecting one third of the overall available rays, the rays of about 27 satellites remain, which may be a setting comparable to GPS only. In contrast, when selecting two thirds of the overall rays, a total of about 54 rays is used for the adjustment. This could be compared e.g. with a multi GNSS or a GPS and InSAR solution to tomography. Figure 7.18 b) is designed in order to investigate to what extent the uncertainty of the estimated refractivities varies in case of a varying atmosphere. This means that the question should be answered in how far the reconstruction algorithm yields more repeatable results for some atmospheric states than for others. Figure 7.18 c) illustrates the generation of randomness by means of varying the orbits. This setting analyzes to what extent the refractivity estimates depend on the orbits. The goal of this setting consists in discretizing the voxels in a way that the estimated parameters preferably do not depend on the orbits.

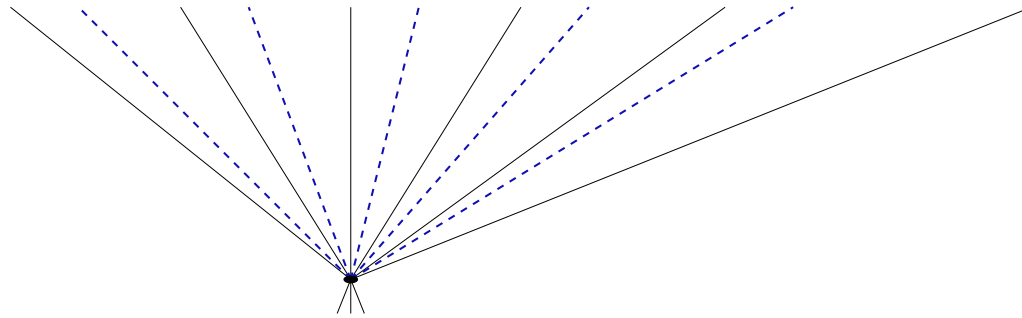
In this work, the questions posed in the cases a) and c) of Figure 7.18 are answered for the site distribution given by the real GNSS sites available within the URG study area. The dependency of the reconstruction quality on the prevailing atmospheric state from Figure 7.18 b) cannot be answered in an information theoretic way because too few WRF acquisitions are available for creating randomness in the atmospheric states. For the two other cases, a total of 48 samples are defined. In case a), 48 randomly selected sets of one third resp. two thirds of a fixed orbit's rays are defined. In case c), 48 orbit samples of the same orbit constellation are defined. The number of 48 samples originates from the idea to use half-hourly orbit samples in case c), i.e. to well cover all possible GPS orbit constellations. In this work, the questions posed in the cases a) and c) of Figure 7.18 are answered for the site distribution given by the real GNSS sites available within the URG study area. The study area is discretized into  $p$  voxels in longitude,  $q$  voxels in latitude, and  $k$  voxels in height, with  $p$ ,  $q$ , and  $k$  from  $\{2, 4, 8, 16\}$ .



a) Constant atmosphere, one single orbit out of which a certain amount of rays is randomly selected



b) Varying atmosphere, one single orbit, all rays of that orbit



c) Constant atmosphere, varying orbits, all rays of those orbits

**Figure 7.18:** Three possible ways of creating randomness enabling an information theoretic approach to water vapor tomography are presented. In case a), a single atmospheric state is observed by a single orbit constellation out of which a certain amount of rays is randomly selected. Case b) uses the variation of the atmosphere as a means for randomness, while only a single orbit is considered and all rays originating of that orbit and the available GNSS site distribution are used. In case c), randomness is generated by varying the orbits within a single atmospheric state.

For all selected orbit samples and the given site distribution, synthetic observations  $\mathbf{y}_{16 \times 16 \times 16}$  and a design matrix  $\Phi_{16 \times 16 \times 16}$  are deduced from WRF using a raytracing in a  $16 \times 16 \times 16$  voxel discretization. Design matrices for coarser voxel discretizations can be easily deduced from the  $\Phi_{16 \times 16 \times 16}$  design matrix by summing up the distances passed within all those  $16 \times 16 \times 16$  voxels situated within one larger  $p \times q \times k$  voxel. However, as indicated in Section 4.3, the functional model building the basis for the tomographic reconstruction is only exact if the parameters are estimated at the voxel discretization that has been used for deducing the observations and the design matrix. The voxel discretization at which the parameters are estimated clearly determines the entries of the design matrix, but both the discretization used for computing the observation vector and the discretization applied for deducing the validation data set from WRF  $N_{\text{wet}, p \times q \times k}$  resp.  $N_{\text{wet}, 16 \times 16 \times 16}$  have to be carefully selected. Based on a fixed design matrix  $\Phi_{p \times q \times k}$ , none of the following options is really good:

1. Use  $\mathbf{y}_{16 \times 16 \times 16}$  and validate using  $N_{\text{wet}, p \times q \times k}$ .
2. Use  $\mathbf{y}_{p \times q \times k}$  and validate using  $N_{\text{wet}, p \times q \times k}$ .
3. Use  $\mathbf{y}_{16 \times 16 \times 16}$  and validate using  $N_{\text{wet}, 16 \times 16 \times 16}$ .
4. Use  $\mathbf{y}_{p \times q \times k}$  and validate using  $N_{\text{wet}, 16 \times 16 \times 16}$ .

If the tomographic results at a specific voxel discretization  $p \times q \times k$  were validated using  $N_{\text{wet}, p \times q \times k}$ , the results obtained at varying voxel discretizations would be compared with varying validation data sets. Therefore, the first and the second option are discarded. When comparing the third and the fourth option, either the observation vector or the design matrix matches well with the validation data set. For this information theoretic analysis, Option 3 is selected instead of Option 4, because it corresponds better to the case of real data, in which the observations correspond to the integrated wet delay along the ray path from the satellite to the receiver, and are not obtained from a discretized atmosphere. I.e. the refractivity values  $\mathbf{x}_{p \times q \times k}$  are estimated based on  $\Phi_{p \times q \times k}$  and  $\mathbf{y}_{16 \times 16 \times 16}$ . The quality of these estimates is, on the one hand, determined by means of considering the difference

$$\Delta N_{\text{wet}} = N_{\text{wet}, 16 \times 16 \times 16} - \mathbf{x}_{16 \times 16 \times 16} \text{ from } p \times q \times k \quad (7.9)$$

and computing its standard deviation as well as the mean of its absolute value. The variable  $\mathbf{x}_{16 \times 16 \times 16}$  from  $p \times q \times k$  contains the estimated parameters, reshaped to  $16 \times 16 \times 16$  voxels by means of determining all  $16 \times 16 \times 16$  voxels situated in the respective  $p \times q \times k$  voxels and assuming all those  $16 \times 16 \times 16$  voxels to contain the refractivity estimated for the larger  $p \times q \times k$  voxel. On the other hand, the uncertainty of the refractivity estimates is described, for each voxel of each discretization, by the entropy of the refractivity estimates within the considered voxel. I.e. a comparison of the entropy in different parts of the study area is possible. For each voxel discretization, the voxel-wise entropy values are computed in the following steps:

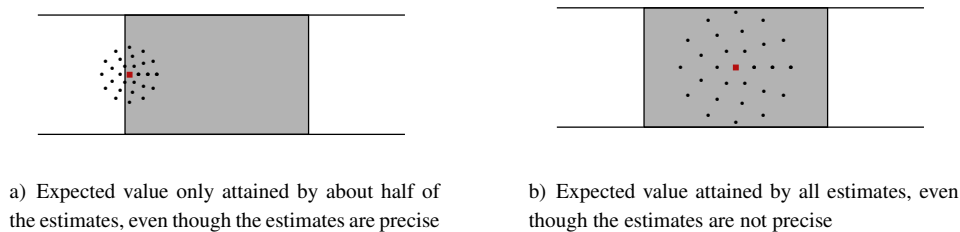
1. For each discretization and each of the 48 samples, estimate the refractivity parameters using the LSQ resp. CS adjustments described in Section 5.3.
2. Using a certain, pre-defined bin width, deduce a probability density function for each voxel's refractivity.
3. Based on the probability density function, compute an entropy value for each voxel's refractivity.

As proposed as precision for the parameter estimation in the LSQ adjustment in Section 5.3, the bin-width was set to 3.5 ppm. Taking into account a possible data space for the generation of the Probability Density Function (PDF) extending from 0 ppm to 80.5 ppm, a total of 16 bins is distinguished. Therefore, based on the Equations 7.6 and 7.8, the possible entropy values range from  $E_{\min} = 0$  to  $E_{\max} = \log_2(23) \approx 4.5$ . However, as the range between zero and one is more easy to imagine, the entropy values are normalized to the interval  $[0, 1]$  in the following. No smoothing filter or continuous smoothing function was applied to the probability density values, because both a smoothing filter and a smoothing function might over-estimate the probability density values of the histogram maximum, which could then result in entropy values larger than  $E_{\max}$ .

In order to avoid binning errors as illustrated in Figure 7.19, the bin centers are set, for each voxel, to the WRF refractivity value in that voxel. I.e. in the case of very precise refractivity estimates, and independently of the bin width, the refractivity values should not fall into two neighboring bins.

### 7.2.3.3 Information theory in tomographic water vapor reconstruction – Results

An information theoretic interpretation of the results is only possible, if the *goodness* of a certain voxel discretization can be expressed in terms of the entropy of that discretization. The following discusses the term *goodness* in the context of a tomographic reconstruction of water vapor based on synthetic GNSS SWDs. One goal could consist in estimating similar refractivities, independently of the ray samples selected from a fixed orbit in case a) resp. independently of the selected orbit in case c). At the same time, the discretization should not be trivial and should be computationally acceptable at a time, i.e. it should consist of, e.g. at least  $2 \times 2 \times 2$  voxels and a maximum of, e.g.  $16 \times 16 \times 16$  voxels. In case of very few, quite large voxels, the entropy within each voxel is expected to be small.



**Figure 7.19:** Challenge in binning: As shown in a), independently of the bin width, even very precise parameter estimates may be situated within different, neighboring bins, whereas less precise parameter estimates as those in b) may be situated within one single bin. The gray rectangles in the two figures shall indicate the bin in which the expected value is situated. The expected value is represented by a red rectangle. The black points correspond to the parameter estimates. For better visibility, the parameter estimates are arranged in the 2D surrounding of the expected value and not only on a 1D line.

In the case of large voxels, no big surprise is expected when selecting different rays from the same orbit resp. when varying the orbits. In the case of a high number of very small voxels, the entropy is either expected to be large (because the voxels' refractivities may be hard to estimate) or small (because the geometric constraints in LSQ may, independently of the orbit, over-smooth the solution).

If available, a discretization disposing of a local minimum in its voxels' entropy values could be considered as a good discretization. The term *local minimum* shall signify that the minimum is neither reached at the coarsest nor at the finest tested discretization level of the horizontal or the vertical direction. Alternatively, if no local minimum exists in the entropy values, the goal could consist in deciding for a precision that should be achieved, i.e. for some discretization interval resp. binning of the possible values in the data space. Then, if the information content were – in an ideal case – equal to zero for a certain discretization, only parameters within one bin would be estimated, which would mean that the estimation would be as precise as the bin width. However, in case of a very small but non-zero information content, this approach does not enable to answer the question by how much how many voxels vary from the value taken at most rays resp. at most orbits. An information theoretic analysis cannot detect the deviation of outliers from the expected value. Therefore, it is crucial to not only consider the entropy of a certain voxel discretization, but to analyze, if possible, at the same time, the standard deviation and the mean of the difference between the estimated refractivities and some refractivity validation data set. Finally, a third approach could consist in deciding on an entropy value or an entropy interval that should be reached. This would correspond to selecting a degree of repeatability that has to be reached, independently of the orbits. Yet, for larger bin numbers than  $n_{\text{bins}} = 2$ , the entropy values are difficult to interpret.

In an exemplary probability distribution in which the estimated refractivities within a voxel can be attributed to only three of the 3.5 ppm large bins, e.g. with probabilities [1/6; 2/3; 1/3], the entropy would attain a value of about 0.28 bit, when normalized to the interval [0, 1]. If the three bins with non-zero probability are assumed to be neighboring bins, and if the middle bin is assumed to contain the WRF refractivity value for that voxel, then this entropy would signify that in two thirds of the cases (i.e. in two thirds of the analyzed orbit samples), the voxel's refractivity would be estimated with an accuracy of 3.5 ppm. Moreover, the entropy value would signify that in all analyzed samples, the voxel's refractivity would be determined with an accuracy of 10.5 ppm, corresponding to three neighboring bins. However, as both assumptions made above are not generally fulfilled, the entropy values are difficult to interpret.

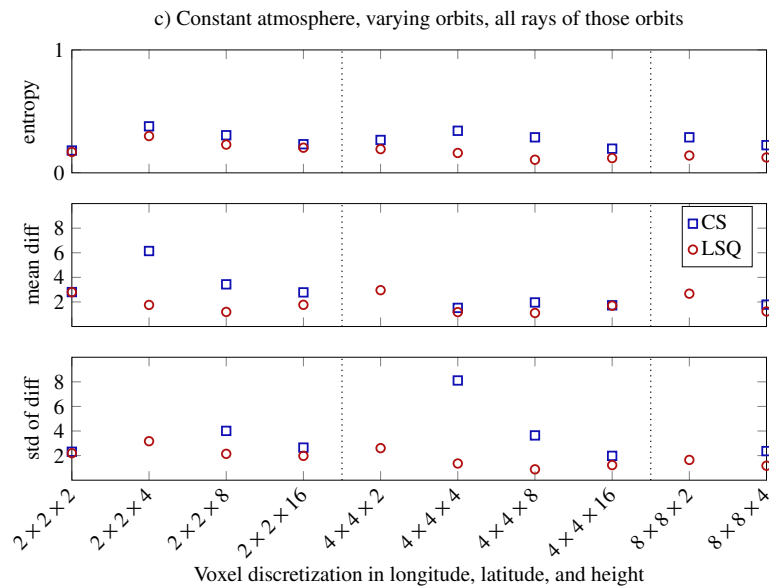
#### **Analysis of c) Constant atmosphere, varying orbits, all rays of those orbits:**

Therefore, when analyzing case c) of Figure 7.18, different criteria are considered at a time in order to evaluate the orbits' influence on the tomographic results at a considered voxel discretization. The mean entropy values per discretization are, on the one hand, compared within the different voxel discretizations in order to analyze in how far it is worth dividing in half an existing voxel. One point of view could be that, when aiming at an



accurate and precise solution, a voxel subdivision is only profitable if it does not yield an increase in the voxels' entropy values. However, as such a point of view may risk to always vote for the coarsest tested discretization level, on the other hand, a maximum accepted mean entropy value per discretization is set to 0.28 bit, as obtained in the above example, and maximum values for the absolute mean and for the standard deviation of the differences between estimated and WRF refractivities of 3.5 ppm are taken into account in order to decide for a *good* voxel discretization.

Figure 7.20 shows, for both LSQ and CS, the mean entropy per voxel discretization and the absolute mean and the standard deviation per discretization of the difference between estimated refractivities and WRF refractivities. The refractivities estimated by means of CS differ, on average over all voxels per discretization, more from the WRF refractivities than the refractivities estimated by means of LSQ. Similarly, the mean standard deviation per discretization is larger in the case of the refractivities estimated by means of CS than in the case of the refractivities obtained from LSQ. This can be explained by the higher sensitivity of the CS solution to the orbit geometry revealed in Section 7.2.2. Besides, CS may be more sensitive to inaccuracies in the functional model than LSQ.



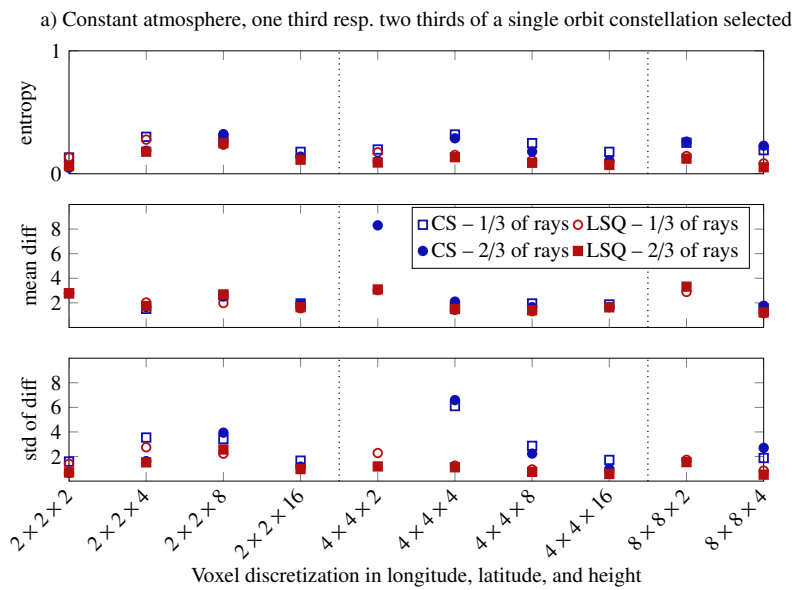
**Figure 7.20:** For both LSQ and CS, the figure shows uncertainty measures for the refractivity estimates obtained based on 48 half-hourly orbit samples of a real GPS orbit geometry, namely the mean entropy per voxel discretization in bits, the mean per voxel discretization of the absolute difference between estimated refractivities and WRF refractivities in ppm, and the standard deviation of the difference between estimated refractivities and WRF refractivities in ppm. The number of voxels of the respective discretizations are given on the axis of abscissae and first increase in the horizontal plane and then in the vertical direction. If no marker is visible for some voxel discretization in the case of CS, then the value is out of the plotted range.

The absolute mean difference resp. the standard deviation of the difference between estimated refractivities and WRF refractivities attains values below 3.5 ppm for all analyzed voxel discretizations except  $2 \times 2 \times 4$  resp. for the voxel discretizations  $2 \times 2 \times 2$ ,  $2 \times 2 \times 16$ ,  $4 \times 4 \times 16$ , and  $8 \times 8 \times 4$ . I.e. from the point of view of the 3.5 ppm threshold for the standard deviation and the mean difference, these voxel discretizations  $2 \times 2 \times 2$ ,  $2 \times 2 \times 16$ ,  $4 \times 4 \times 16$ , and  $8 \times 8 \times 4$  would be considered as good. Based on the entropy threshold, the same voxel discretizations and the additional voxel discretization  $4 \times 4 \times 2$ , would be judged as good. Yet, when taking into account that the entropy should not increase if a voxel is horizontally or vertically subdivided, no subdivision of the  $2 \times 2 \times 2$  voxel discretization could be accepted, because both the  $4 \times 4 \times 2$  and the  $2 \times 2 \times 4$  voxel discretization dispose of larger entropy values than the  $2 \times 2 \times 2$  voxel discretization. Considering further finer subdivisions, the LSQ mean entropy values per discretization decrease below the value of the  $2 \times 2 \times 2$  voxel discretization. However, this decrease may be attributed to a larger smoothing of the geometric constraints. Consequently, the analyzed goodness criteria do not clearly recommend one of the tested voxel discretizations for the investigated setting

composed of only seven synthetic GNSS sites and an inaccurate functional model deducing the synthetic SWD observations, independently of the voxel discretization at which the refractivity field is reconstructed, from a fine  $16 \times 16 \times 16$  voxel discretization.

**Analysis of a) Constant atmosphere, randomly selected rays of one single orbit:**

For case a) of Figure 7.18, Figure 7.21 shows the same uncertainty measures as Figure 7.20. Yet, while Figure 7.20 compares the uncertainties of the LSQ and CS refractivity estimates obtained at varying orbits, Figure 7.21 compares the uncertainties of the LSQ and CS refractivity estimates deduced from one third resp. two thirds of the rays of a GPS-like orbit geometry with nine orbital planes with nine satellites per plane. Adding a higher number of rays (i.e. two thirds of the rays instead of one third of the rays) of the same orbit constellation into the tomographic system decreases the mean entropy per discretization by up to 0.11 bits resp. 0.10 bits in CS resp. LSQ. This maximum decrease is observed at voxel discretization  $2 \times 2 \times 4$ , at which the absolute mean difference and the standard deviation decrease by 0.1 ppm resp. 0.3 ppm and by 1.9 ppm resp. 1.2 ppm when introducing twice the number of rays into the CS resp. LSQ solution.



**Figure 7.21:** For both LSQ and CS, the figure shows uncertainty measures for the refractivity estimates obtained based on 48 samples of one third resp. two thirds of the rays of one single GPS-like orbit constellation with nine orbital planes and nine satellites per plane, namely the mean entropy per voxel discretization in bits, the mean per voxel discretization of the absolute difference between estimated refractivities and WRF refractivities in ppm, and the standard deviation of the difference between estimated refractivities and WRF refractivities in ppm. The number of voxels of the respective discretizations are given on the axis of abscissae and first increase in the horizontal plane and than in the vertical direction. If no marker is visible for some voxel discretization in the case of CS, then the value is out of the plotted range.

**Discussion of the results of a) and c):**

The results shown for the cases a) and c) of Figure 7.18 both refer to mean uncertainty measures per discretization. I.e. they do not investigate in how far the analyzed uncertainty measures depend on e.g. the number of rays per voxel. The entropy values per voxel and the absolute mean and the standard deviation of the difference between estimated refractivities and WRF refractivities are compared at the  $16 \times 16 \times 16$  voxel level for all analyzed discretizations. Yet, no correlation between the refractivity estimates and the number of rays per voxel is determined in this section. I.e. the refractivity estimates at coarse voxels that are derived from observations referring to a fine  $16 \times 16 \times 16$  voxel discretization do not show systematically worse results in those parts of the study area in which, e.g. no synthetic GNSS site is available. Table 7.8 summarizes the answers to the research questions investigated in this section.

**Table 7.8:** Research questions answered using a synthetic data set which is designed for analyzing the effect of the orbits and of the voxel discretization.

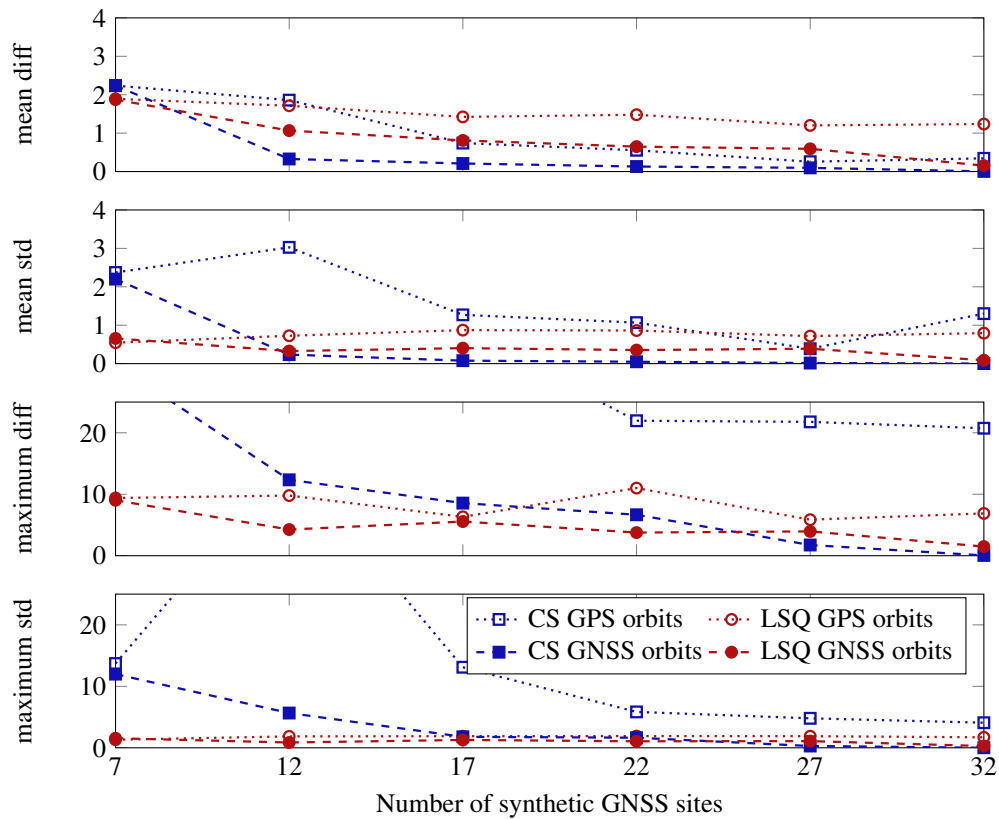
**Synthetic data set: orbits and voxel size**

- In how far does the inclusion of the rays of more satellites than in the case of GPS improve the repeatability of the results?
- How large should the tomographic voxels be in order to yield repeatable results within a changing orbit geometry, when dealing with a tomographic setting containing seven synthetic GNSS sites in the  $95 \times 99 \text{ km}^2$  large Study Area 1 and when deducing the synthetic SWD estimates from a fine  $16 \times 16 \times 16$  voxel discretization?
- Based on 48 samples of one single orbit constellation with nine orbital planes and nine satellites per plane, the inclusion of two thirds of the rays instead of one third of the rays does not show a systematic improvement.
- The three investigated goodness criteria do not clearly recommend one of the tested voxel discretizations for the analyzed setting composed of only seven synthetic GNSS sites and an inaccurate functional model.

## 7.2.4 Summary

Section 7.2.1 showed for two orbit samples that artificial ray directions drawn at random from  $\epsilon \in [7^\circ, 90^\circ]$  and from  $A \in [0^\circ, 360^\circ]$  enable an accurate reconstruction of the 3D refractivity field, if 25 synthetic InSAR sites are available in addition to the seven GNSS sites situated with Study Area 1. Also in the case of Section 7.2.2, when considering GPS-like ray directions resulting from 25 synthetic GNSS sites, all situated within a horizontal plane, and nine orbital planes with nine satellites per plane, very accurate results were obtained for some site distributions, if the voxel discretization was coarse enough. Yet, when introducing only the seven synthetic GNSS sites already used in Section 7.2, independently of the selected voxel discretization, Section 7.2.3 did not succeed in accurately estimating the 3D refractivity field with GPS-like ray directions resulting from either six or nine orbital planes with five resp. nine satellites per plane. The inaccurate estimates in Section 7.2.3 result from a small number of synthetic GNSS sites and from an inaccurate functional model using another voxel discretization for the generation of the synthetic observations than for the estimation of the refractivity parameters. Based on many orbit samples and an accurate functional model, this summarizing section aims at analyzing how many synthetic GNSS sites are necessary in addition to the seven sites given by the SAPOS<sup>®</sup> network in order to accurately reconstruct the 3D refractivity field.

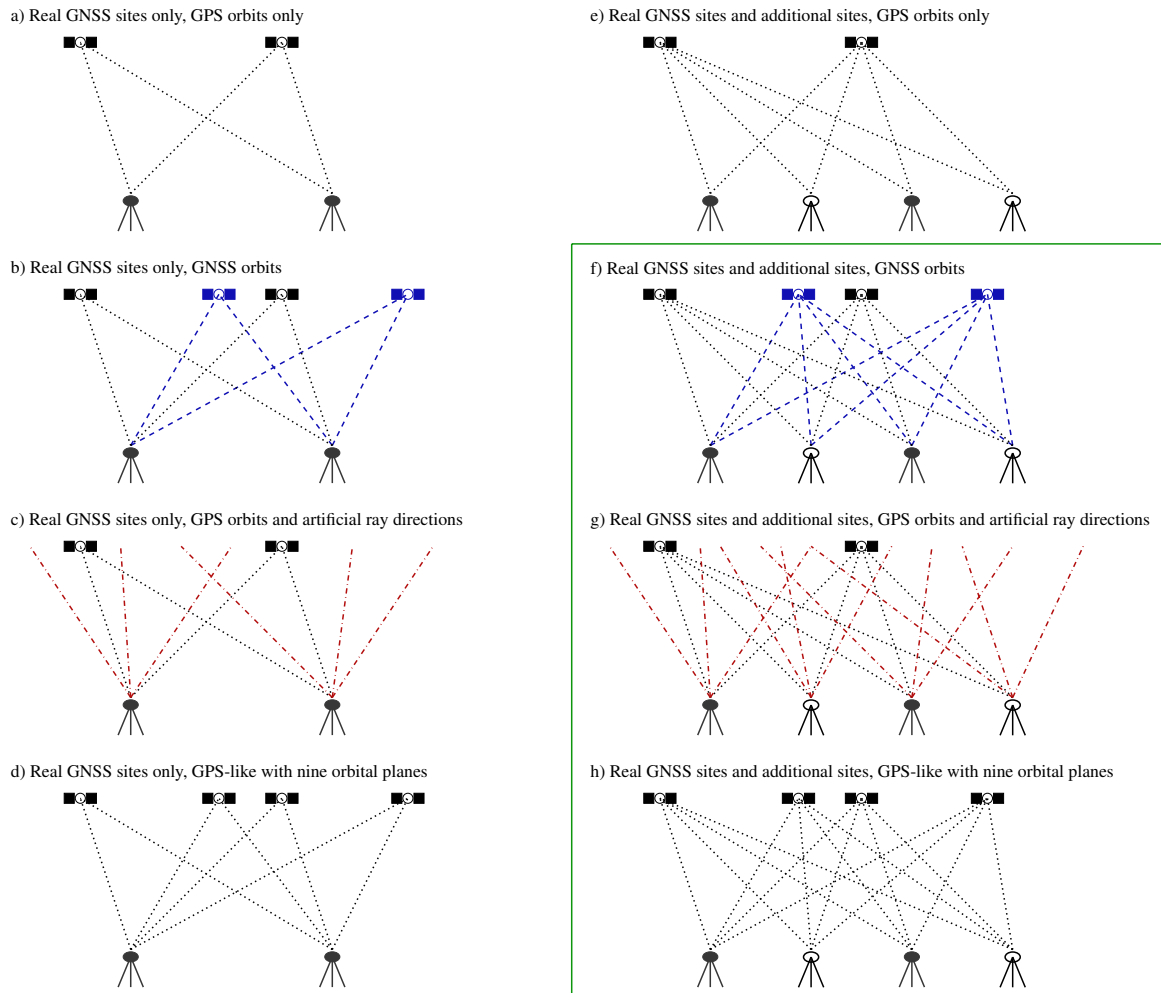
Figure 7.22 shows the accuracy and the precision of CS resp. LSQ refractivity estimates obtained based on seven, 12, 17, 22, 27, or 32 synthetic GNSS sites at a  $5 \times 5 \times 5$  voxel discretization of the  $99 \times 95 \text{ km}^2$  large URG Study Area 1. Per site, ten GPS-like resp. GNSS-like rays have been introduced at a time. I.e. the number of rays is equal, independently of selecting the ray direction from a GPS-like orbit geometry with six orbital planes and five satellites per plane or from a realistic GNSS-like orbit geometry. A GNSS-like orbit geometry is simulated according to the real GPS-Galileo-GLONASS orbit constellations, i.e. with i) six orbital planes inclined by  $55^\circ$  w.r.t. the equator and about five satellites per plane for GPS, plus ii) three orbital planes inclined by  $56^\circ$  w.r.t. the equator with nine satellites per plane for Galileo, plus iii) three orbital planes inclined by  $64.8^\circ$  w.r.t. the equator with seven satellites per plane for GLONASS. On the one hand, the mean and the maximum values of the absolute differences between estimated refractivities and WRF refractivities, averaged over all voxels and over all 48 tested orbit samples, are given. On the other hand, the mean and the maximum values of the standard deviation of the refractivity estimates within the 48 samples, encountered within any of the 125 voxels, is shown.



**Figure 7.22:** Comparison of the accuracy and the precision in ppm of the estimated refractivity field obtained at a  $5 \times 5 \times 5$  voxel discretization by means of LSQ resp. CS based on different numbers of synthetic GNSS sites and based on synthetic GPS resp. GNSS orbits. The first and second line show the mean value of the absolute difference resp. the standard deviation of the difference between estimated refractivities and WRF refractivities, averaged over all voxels and over all 48 tested orbit samples. In contrast, the third and fourth line show the maximum value of the mean difference resp. of the standard deviation encountered within any of the 48 samples and within any of the 125 voxels. The dashed and dotted lines serve for better following the variation of the mean resp. of the std at the different site numbers, but only the discrete values indicated by the markers should be evaluated.

While the LSQ estimates attain maximum absolute differences resp. the maximum standard deviations of 9.0 ppm resp. 1.6 ppm in the case of seven sites and GPS orbits only, the maximum absolute differences resp. the maximum standard deviations of the CS estimates w.r.t. the WRF refractivities attain values of 68.8 ppm resp. of 13.7 ppm. If the ray directions are selected from a GNSS-like orbit geometry instead of from a GPS-like orbit geometry, the maximum absolute differences resp. the maximum standard deviations of the CS estimates w.r.t. the WRF refractivities decrease by 37.7 ppm resp. by 1.7 ppm. When focusing on the absolute mean difference resp. on the mean standard deviation obtained based on seven sites over all samples and all voxels, LSQ also yields more accurate and more precise results than CS, even though the absolute mean difference for LSQ is only 0.3 ppm resp. 0.4 ppm smaller than for CS, and the mean standard deviation for LSQ is only 1.8 ppm resp. 1.5 ppm smaller than for CS, when using GPS only resp. GNSS ray directions. Yet, already as of a total of twelve synthetic GNSS sites, the CS absolute mean difference resp. the CS mean standard deviation attains values below the corresponding measure deduced from LSQ refractivity estimates. In the case of 22 resp. 25 synthetic GNSS sites, the maximum difference between the CS refractivity estimates and the WRF refractivities over all samples and all voxels is equal to 6.7 ppm resp. to 1.7 ppm, at absolute mean differences resp. mean standard deviations over all samples and all voxels lower or equal than 0.1 ppm. At the same number of sites and the same GNSS-like orbit geometry, the maximum difference between the CS refractivity estimates and the WRF refractivities over all samples and all voxels is equal to 3.8 ppm resp. to 4.0 ppm, at absolute mean differences resp. mean standard deviations over all samples and all voxels of about 0.6 ppm.

To sum up, and as shown in Figure 7.23, the 3D refractivity field can be reconstructed at high accuracies by means of CS if *enough* synthetic GNSS sites and a *variety* of ray directions are available. According to the results of the previous sections, the term *enough* shall correspond to the rule of thumb of Champollion et al. [2004] recommending a horizontal voxel size similar to the mean inter-site distance.



**Figure 7.23:** The figure shows different orbit geometries as well as a varying number of GNSS sites. For both LSQ and CS, the observing geometries highlighted by the green frame are advantageous. I.e. the presented research suggests to obey the rule of thumb of Champollion et al. [2004] recommending the horizontal voxel size similar to the mean inter-site distance. If a large variety of ray directions is available, CS showed to be able to yield very accurate results in synthetic data sets in which the functional model  $\mathbf{y} = \Phi \cdot \mathbf{x}$  is accurate.

A *variety* of ray directions shall not necessarily denote a high number of rays per site, but a variety of different directions of the considered rays. Based on ten rays per site taken out of a GPS orbit geometry, the reconstruction of the 3D refractivity field is less accurate than based on ten rays per site drawn e.g. from a GNSS orbit geometry. If less synthetic GNSS sites are available than recommended by the rule of thumb of Champollion et al. [2004], if the functional model is inaccurate, or if only GPS SWDs are available, this research recommends the use of a LSQ solution regularized with geometrical smoothing constraints. However, in case of mean inter-site distances corresponding to the horizontal voxel size, the availability of the rays of multiple GNSS, and an accurate functional model, the CS solution strategy is recommended.



## 8 Discussion and Outlook

This last section discusses the methodology applied resp. the results presented in this thesis and gives an outlook on future research.

### Compressive Sensing vs. Least Squares:

- Comparison of the reconstruction accuracy and precision

The presented research has shown that CS is a valuable method for tomographic water vapor reconstructions based on SWD estimates. In case of appropriate tomographic settings, CS can yield very precise and accurate results. This thesis shows that the quality of the tomographic reconstruction by means of CS mainly depends on i) the accuracy of the functional model, ii) the number of available GNSS sites, iii) the voxel discretization, and iv) the variety of ray directions introduced into the tomographic system. Therefore, the very accurate results obtained within synthetic data sets with an accurate functional model and e.g. artificial ray directions cannot readily be transferred to real data sets. Instead of mapping real InSAR SWDs into artificial directions, the large variety of ray directions required in the CS reconstruction could be, in the case of the real data set, obtained by means of introducing SWD estimates of multiple GNSS. Yet, as long as discrete refractivity parameters per voxel are estimated based on integrated SWD estimates, the functional model will never be accurate. As a consequence, future research first needs to focus on how to make the CS solution strategy less sensitive to inaccuracies in the functional model. Thereafter, noise should be added to the observations in order to analyze the robustness of the LSQ resp. CS solution strategies. A Monte-Carlo simulation could be applied in order to investigate in how far the LSQ resp. CS solution strategies are sensitive to e.g. Gaussian white noise.

- Risk of over-smoothing and suitability of the iDCT Euler Dirac dictionary

When comparing the LSQ and the CS solutions, on the one hand, LSQ risks to over-smooth the solution by applying geometric constraints that might not be able to represent the true atmospheric behavior. Yet, on the other hand, the geometric smoothing constraints prevent the LSQ solution from producing outliers as encountered in some tomographic settings in the case of CS. As many different Euler letters with varying steepness are introduced in the height direction and as a high number of linear combinations of the resulting iDCT Euler Dirac atoms can be built, the CS solution should be more adaptive and thus less constraint-driven than the LSQ solution. However, depending on the dictionary, there might be atmospheric behaviors that cannot be well represented by the proposed atoms. Although the iDCT Euler Dirac dictionary showed to yield, in appropriate tomographic settings, very accurate results at all dates analyzed within the synthetic data sets, this does not necessarily imply that this dictionary is also suitable for the tomographic reconstruction of the wet refractivity field based on real SWD estimates. In addition, the capability of the proposed dictionary to reconstruct local refractivity disturbances may be poor. If the refractivity is much higher in a small area of e.g. one or two voxels, the iDCT letters in longitude and latitude would be required to cancel out each other in the atmospheric sections around the disturbance. However, canceling out some atoms in some parts of the study area, but not in others, would require a large number of non-zero coefficients, which is not favored at all by the CS solution algorithm. While a larger variety of atoms may induce an accuracy improvement, the cost of these additional atoms in terms of sparsity of the solution and in terms of computing time should always be kept in mind. The CS solution aims at a sparse representation of the parameters in the transform domain. Therefore, selecting a high number of atoms is not possible. Consequently, similarly to LSQ, CS may risk to over-constrain the solution. In order to accurately represent local atmospheric disturbances while

keeping the solution sparse, compactly supported letters as wavelets should be added in future research to the iDCT letters in longitude and latitude.

While the dictionary composed of iDCT, Euler, and Dirac atoms is able to accurately represent the WRF refractivity field e.g. for a  $5 \times 5 \times 5$  voxel discretization, the dictionary may not be suitable for sparsely representing real atmospheric refractivity distributions. Therefore, future research should check the transferability of the iDCT Euler Dirac dictionary on CS solutions that are based on real SWD estimates. In addition, in order to better evaluate the capacity of CS resp. LSQ to reconstruct the 3D refractivity field, future research should aim at separating the effect of the dictionary e.g. from effects of the functional model's accuracy, from the density of the available sites within the chosen voxel discretization, and from the variety of ray directions introduced into the tomographic system.

In addition, a two-step  $L_1$ - $L_2$ -norm solution strategy should be investigated in the future. Section 7.2.4 showed e.g. in the case of twelve synthetic GNSS sites and GNSS-like ray directions, that the maximum differences between estimated refractivities and WRF refractivities may attain 8.1 ppm larger values in the case of CS than in the case of LSQ, even if the absolute mean difference between the estimated refractivities and the WRF refractivities over 48 orbit samples and over all  $5 \times 5 \times 5$  analyzed voxels is 0.7 ppm smaller. While CS accurately estimates the refractivity of many voxels, the refractivity of some voxels cannot be determined. Consequently, future research should investigate in how far the CS solution strategy can provide a quality measure for each voxel's refractivity estimate. Such a quality measure would enable a two-step  $L_1$ - $L_2$ -norm refractivity estimate using the CS refractivity estimates of all those voxels that could be accurately modeled as a priori information for an  $L_2$ -norm solution. For all those voxels that could not be accurately resolved by CS, this  $L_2$ -norm solution could then include a geometric smoothing constraint as described for the LSQ solution in Section 5.3.

- Voxel discretization vs. spatial resolution of the refractivity reconstruction

Both the horizontal and the vertical voxel discretization showed significant effects on the quality of the refractivity estimates. While the effect of the horizontal voxel discretization is analyzed in the Sections 7.2.2 to 7.2.4, the effect of the vertical voxel thicknesses becomes visible when comparing the results of Section 7.2 with the results presented in Heublein et al. [2018]. While Heublein et al. [2018] use comparable voxel sizes in longitude and latitude as Section 7.2 of this work, they define approximately twice the number of voxels in the height direction. As a consequence, Heublein et al. [2018] do not obtain as accurate and as precise refractivity estimates as presented in Section 7.2. Besides the rule of thumb of Champollion et al. [2004], GNSS tomography still lacks in concrete and transferable advice on how to discretize the considered study area based on the horizontal density of available GNSS sites. As the voxel discretization and the site density were proved to be essential for an accurate reconstruction of the 3D water vapor field, future research should further analyze the effect of both the horizontal and the vertical voxel discretization on the quality of the tomographic results.

Moreover, when considering the information theoretic analysis in Section 7.2.3, the capacity of CS to provide a solution within very coarse voxel discretizations like  $2 \times 2 \times 2$  should be questioned. The CS approach presented in Section 5.4 uses the sparsity of the coefficients in the iDCT Euler Dirac transform domain as a prior for regularization and requires the solution to be composed of about 5% to 15% of the total number of coefficients in that transform domain. When further requiring these 5% to 15% of the coefficients to correspond to e.g. at least three coefficients, then a total of at least 20 (15%) to 60 (5%) coefficients is necessary in the transform domain. If  $\mathcal{K} = 2$  is the number of height layers,  $L = \mathcal{P} \cdot \mathcal{Q} \cdot \mathcal{K} = 8$  is the number of voxels in the spatial domain in the  $\mathcal{P} \times \mathcal{Q} \times \mathcal{K} = 2 \times 2 \times 2$  voxel discretization, and  $M$  is the number of parameters in the transform domain, then the iDCT Euler Dirac dictionary of size  $L \times M$ , with  $M$  being the number of parameters in the transform domain, is composed of one  $L \times L = 8 \times 8$  Euler iDCT dictionary per introduced Euler steepness parameter  $\alpha_{CS}$  and one  $L \times \mathcal{K} = 8 \times 2$  iDCT Dirac dictionary. Thus at least three different steepness parameters need to be introduced in order to enable the CS approach described in Section 5.4 within a  $2 \times 2 \times 2$  voxel discretization.



The other way round, this discussion should analyze the maximum number of steepness parameters  $\alpha_{CS}$  for which 5% to 15% of the parameters  $s$  in the transform domain are fewer than the total number of voxels in a particular voxel discretization. One of the motivations for the CS approach was the possibility to reconstruct the 3D refractivity signal based on the determination of a smaller number of coefficients than in the case of LSQ, where one parameter per voxel needs to be determined. The proposed CS requires 5% to 15% of the coefficients in the transform domain to contain 99% of the signal power. If only these 5% to 15% non-zero coefficients are used to reconstruct the refractivity signal by means of CS, the number of steepness parameters  $n_{steepness}$  would need to satisfy

$$0.15 \cdot (n_{steepness} \cdot \mathcal{P} \cdot \mathcal{Q} \cdot \mathcal{K} + \mathcal{K}) < \mathcal{P} \cdot \mathcal{Q} \cdot \mathcal{K} \quad (8.1)$$

in order to enable CS to do with less parameters than LSQ. This can be simplified to

$$n_{steepness} < \frac{\mathcal{P} \cdot \mathcal{Q} + 0.15}{0.15 \cdot \mathcal{P} \cdot \mathcal{Q}}. \quad (8.2)$$

Thus, for  $\mathcal{P} \cdot \mathcal{Q} = 1$  a maximum of five steepness parameters should be used, and if the voxel number in longitude and latitude approached infinity, a maximum of six steepness parameters should be introduced within the dictionary, if the number 5% to 15% of the parameters in the transform domain shall be fewer than the number of parameters that an LSQ approach would estimate.

Defining only two voxels in the longitude and latitude direction in a  $95 \times 99 \text{ km}^2$  large study area as that in Section 7.2.3 might, on the one hand, seem very few. However, on the other hand, when comparing the tomographic setting of Section 7.2.3 to previous research as in Table 3.1, the total number of seven synthetic GNSS sites in the  $95 \times 99 \text{ km}^2$  large study area is very low. Many previous approaches to water vapor tomography disposed of more GNSS sites within even smaller study areas. Moreover, a lot of previous research applied horizontal smoothing constraints as the LSQ solution in this work. Yet, the horizontal voxel discretization should not be set equal to the spatial resolution refractivity estimates. As soon as a constraint tries to set a voxel's refractivity to the mean of the refractivities of the horizontally neighboring voxels, the horizontal resolution of the refractivity estimates risks to decrease. No reference to this mismatch between a refractivity estimate at a certain spatial resolution corresponding to the voxel discretization and the averaging nature of the horizontal smoothing constraints could be found in the literature.

- Tuning of trade-off parameters

This work points out that both LSQ and CS require a good tuning of the trade-off parameters. Both solution strategies make use of a two-step trade-off parameter selection ensuring both the stability of the solution and small observation residuals. In case of inappropriately selected trade-off parameters, both LSQ and CS are incapable to produce an accurate solution. Similarly, as stated in Heublein et al. [2018], the selection of appropriate values for the scale height  $H_{scale}$  resp. for the parameter  $\alpha_{CS}$  defining the steepness of the Euler letters is essential in the LSQ resp. CS solution. As the tuning of the trade-off parameters is crucial, the quality of the solution depends on the variety of trade-off parameters and on the variety of options for the  $H_{scale}$  and for  $\alpha_{CS}$  introduced within the LSQ and CS solution approaches. The above stated maximum number of six steepness parameters enabling CS to do with fewer parameters than LSQ may be considered as a limitation of the proposed CS approach with a iDCT Euler Dirac dictionary.

**Need for a variety of ray directions in CS:**

Section 7.2.4 emphasized that CS requires a variety of ray directions in order to very accurately reconstruct the 3D refractivity field. Yet, as the measurement matrices in water vapor tomography usually cannot be flexibly designed, taking the advice of e.g. Candès et al. [2006] and Rauht [2010] to involve randomness in the acquisition step is challenging. Therefore, this work suggests to introduce at least the rays of multiple GNSS instead of GPS rays only into the tomographic system. In addition, future research should investigate in how far the inclusion of GNSS RO observations improves the reconstruction of the 3D refractivity field by means of CS.

**Inclusion of InSAR SWDs in addition to the GNSS SWDs:**

While the inclusion of InSAR SWDs significantly improves the CS solution in the synthetic data set, no preference

can be given to any of the two analyzed solution strategies in the case of real data. The LSQ and the CS approaches to water vapor tomography are widely consistent for both GNSS only and for GNSS and InSAR. Yet, this may be due to the way of integrating the synthetic absolute GNSS and InSAR SWD estimates into the tomographic system. As discussed in Heublein et al. [2018], to which this whole paragraph refers, the GNSS and InSAR observing geometries differ, and therefore, the two measurement techniques observe different sections of the atmosphere. The horizontal resolution of InSAR atmospheric phase maps is much higher than the point-wise GNSS ZWD resolution. Reversely, each of the atmospheric phases at a persistent scatterer contains information on a much smaller atmospheric section than a GNSS ZWD averaging the atmospheric behavior within a large cone above the respective GNSS site. The mapping of such GNSS ZWDs into different ray directions as well as the mapping of InSAR ZWDs into different azimuth and elevation angles has to be further investigated. In this context, particular focus should be set on the choice of the mapping function and on a mapping of the InSAR ZWDs into appropriate slant directions. The mapping into uniformly distributed directions with azimuth angles  $A \in [0^\circ, 360^\circ[$  and with elevation angles  $\epsilon \in [7^\circ, 90^\circ]$  may cause unrealistic InSAR SWDs corrupting the tomographic solution, especially in the case of significant horizontal variations in the wet refractivity. Moreover, simple  $\sin \epsilon$  mapping functions should be avoided, especially in the case of low elevation rays within the voxels limited by elliptical upper and lower boundaries. Therefore, in future studies, more complex mapping functions will be used that depend e.g. on the day of year, on meteorological parameters, and on the height of the considered site.

As an alternative to the proposed approach, InSAR and GNSS could also be introduced as independent inputs into the tomographic system. However, the tomographic system would then have to be solved for more unknowns and the differences between the GNSS and the InSAR observing geometries would have to be understood exactly. As the complete SWD product resulting from the GNSS and InSAR combination is used in this work, the observing geometries of the two systems are combined. This proceeding does not necessarily imply a loss of information. Instead, as discussed in Heublein et al. [2018], this synergy highly densified the available ZWD network. The ZWD resp. PW product generated by combining InSAR and GNSS was validated and proved to show strong correlation with other data sets. Alshawaf et al. [2015c] compared the derived PW maps with PW maps measured by the optical sensor MERIS. The results of Alshawaf et al. [2015c] show strong spatial correlation between the two data sets, with values of uncertainty of less than 1.5 mm.

In future work, the potential of including observations of two SAR satellites with different viewing angles e.g. in descending and in ascending mode, shall be analyzed. As this work showed that a variety of ray directions is crucial for an accurate CS reconstruction of the 3D refractivity field, future research should also investigate the inclusion of InSAR observations acquired from different orbits as water vapor information into the tomographic system without applying any PS processing. PS is only possible if all SAR acquisitions have been acquired along the same orbit, which implies both a limitation in the ray directions into which the SAR delays can be mapped and a bad temporal resolution even with the additional SAR satellites launched in the previous years.

#### **Validation possibilities in the real data set:**

As stated in Section 7.1.3, the validation of the tomographic solution in the real data set is challenging because only a single radiosonde site is situated within the analyzed study area. Thus, no 3D validation is possible, but just a validation along a single profile. As discussed in Heublein et al. [2018], there might be solutions matching well along this profile, but having a bad reconstruction accuracy in the remaining parts of the study area. Furthermore, there are only radiosonde profiles available around 0 UTC and 12 UTC. As the temporal resolution of twelve hours is low and as the atmospheric water vapor is highly variable in time, no interpolation of the radiosonde data at 9h48 UTC is performed. Finally, although the radiosonde moves horizontally during its ascent, the whole profile is supposed to be situated vertically above the radiosonde site, which might cause inaccuracies in the validation.

If available, a 3D validation w.r.t. a numerical weather model could be a good alternative to the radiosonde validation. Yet, as shown in Figure 6.4, the WRF data and the radiosonde ascents available for this work do not necessarily describe similar water vapor distributions. Alternatively, as proposed in Heublein et al. [2018], if the tomographic model would take into account temporal variations in the refractivity, the GNSS and InSAR solution could be temporally propagated until the radiosonde ascent time. Thus, the radiosonde observations could be used

in order to validate the inclusion of InSAR data into the tomographic system even though the acquisition times of InSAR and of the radiosonde differ.

Comparing the refractivity estimates with independent GNSS SWD estimates only enables a validation of the integrated water vapor content through the whole atmosphere. GNSS SWD estimates could be compared to SWD values obtained based on the tomographically reconstructed 3D refractivity distribution. If these reconstructed SWDs are compared to the GNSS SWD estimates in many different ray directions, the 3D reconstruction quality could be validated. However, an independent set of GNSS SWD estimates usually is not available. In particular, once absolute SWDs from InSAR are included into the tomographic system as in the real data set in Section 7.1, GNSS SWDs of all available GNSS sites are automatically introduced within the tomographic system.

Besides validations via radiosondes, numerical weather models, or independent GNSS SWD estimates, radiometers could be used for validation, e.g. the Television Infra-Red Observing Satellite-N (TIROS-N) Operational Vertical Sounder (ATOVS) instruments as part of the payload of the NOAA and Meteorological Operational Satellite (MetOp) polar orbiting satellites, both containing a spaceborne microwave radiometer and high resolution infrared radiation sounders, which yield global water vapor products. ATOVS water vapor retrievals before March 2013 use first guess information from the German Meteorological Office (Deutscher Wetterdienst) (DWD) global numerical weather prediction model GME by means of solving the radiative transfer equation. This already indicates a limitation of the ATOVS water vapor profiles: They are not fully independent from the first guess input model, which, itself, is not independent of ATOVS because Advanced Microwave Sounding Unit (AMSU)-A data is assimilated in the GME model. Moreover, the low altitude water vapor distribution over land surfaces is affected by the quite variable surface radiation over land, and the horizontal resolution of  $90\text{ km} \times 90\text{ km}$  of the ATOVS water vapor profiles is very poor. Hence, ATOVS only yields rough information on the vertical distribution of neutrospheric water vapor but cannot be applied for validation or as prior knowledge on the horizontal water vapor distribution.

As none of the presented alternatives to the radiosonde validation is a good option, future research should try to use at least three spatially well distributed radiosonde sites for the validation of the tomographic refractivity estimates in the real data set and focus on a validation at the radiosonde ascent time. In a first step, such a validation at the radiosonde acquisition time can be performed based on a GNSS only solution including ray directions of multiple GNSS. In a second step, InSAR SWDs can be added. Yet, as discussed above, the fixed InSAR acquisition time represents a challenge when aiming at a radiosonde validation, and the mapping of the InSAR SWDs into artificial directions is only possible within an azimuthally symmetric atmosphere.

#### **Summary:**

This thesis presents CS as a powerful tool for accurately reconstructing the 3D water vapor field based on SWDs acquired along a variety of different signal directions. The developed tomography approach is based on geometrical and physical models that allow to combining and evaluating very different observation types and measurement techniques at a time. Therefore, the tomography approach represents a very flexible tool, especially under the light of new and heterogeneous satellite missions.



# A Appendices

## A.1 Niell mapping function

Niell [1996] define the Niell mapping functions as

$$\text{mf}(\epsilon) = \frac{1 + \frac{a_{\text{mf}}}{1 + \frac{b_{\text{mf}}}{1 + c_{\text{mf}}}}}{\sin \epsilon + \frac{a_{\text{mf}}}{\sin \epsilon + \frac{b_{\text{mf}}}{\sin \epsilon + c_{\text{mf}}}}} \quad (\text{A.1})$$

with parameters  $a_{\text{mf}}$ ,  $b_{\text{mf}}$ , and  $c_{\text{mf}}$  depending, in the case of the hydrostatic Niell mapping function, on the site latitude and height as well as on the Day Of Year (DOY), and in the case of the wet Niell mapping function, on the site latitude only. According to Niell [1996], the parameter  $a_{\text{mf}}$  for the hydrostatic mapping function at tabular latitude  $\varphi_i$  at time  $t$  from January 0.0 (in UT days) is calculated as

$$a_{\text{mf}}(\varphi_i, t_{\text{mf}}) = a_{\text{mf avg}}(\varphi_i) - a_{\text{mf amp}}(\varphi_i) \cdot \cos\left(2 \cdot \pi \cdot \frac{t_{\text{mf}} - t_{\text{mf},0}}{365.25}\right) \quad (\text{A.2})$$

where  $t_{\text{mf},0}$  is the adopted phase, DOY 28 and  $t_{\text{mf}}$  is the time from  $t_{\text{mf},0}$ , corresponding to January 0.0, in days. The value of  $a_{\text{mf}}(\varphi, t_{\text{mf}})$  is obtained by interpolating linearly between the nearest  $a_{\text{mf}}(\varphi_i, t_{\text{mf}})$  from Table A.1. The parameters  $b_{\text{mf}}$  and  $c_{\text{mf}}$  are determined in a similar way.

In case of the wet mapping function, the parameters are obtained from Table A.2 by linear interpolation in latitude.

## A.2 PPP processing in Bernese GPS Software 5.2

Table A.3 gives an overview over the main processing steps for Precise Point Positioning applied within the Bernese GPS Software 5.2.

**Table A.3:** Main processing steps for Precise Point Positioning in the Bernese GPS Software 5.2, according to Dach et al. [2007], Fuhrmann [2010], and Knöpfler [2015].

<b>Copy required files</b>	
GS01_COP	Copy and unpack GNSS observation files, orbits, satellite clock information, Earth rotation parameters (ERP), differential code biases, ionosphere information, station information files, approximate coordinate file, and ocean tide loading into active campaign
<b>Prepare the pole files, the orbits, and the satellite clock data</b>	
POLUPDH	Extract ERP information from an IERS formatted pole file (*.IEP) into a Bernese formatted pole file (*.ERP)
ORBMRGH PRETAB	Convert the orbit information available in the IGS SP3c format (Earth fixed frame, *.PRE) into tabular positions in the inertial frame (*.TAB) for subsequent numerical integration by ORBGEN

*Continued on next page*

Table A.3 – *Continued from previous page*

ORBGENH	Integrate the equations of motion using the positions given in the tabularx orbit file to produce a Bernese standard orbit file (*.STD) used in all processing programs needing orbit information. The orbit is represented by six osculating elements and nine dynamical parameters associated with radiation pressure. A summary file (PPPysssss.PRC) is generated, providing an overview of the orbit fit quality. It is contained in the processing summary. When relying on IGS products, the fit rms should be around 1 cm.
CCRNXCH	Converts clock RINEX files into a Bernese satellite clock file (extension CLK). The resulting file has the same name as the RINEX file, but resides in the campaign's ORB directory.
RNXCLK	Extracts satellite clock information from combined RINEX clock file
<b>Preprocess, convert, and synchronize observation data</b>	
RNXGRA	Produce a summary of the available RINEX observation data, giving a complete overview of observed satellites, involved stations and their performance. This file appears in the processing summary and may help to identify possible data tracking problems of observing sites.
RNXSMTAP	This script and the following form a unit. The purpose is to clean data on the RINEX level. It makes sense to parallelize this step. This first script prepares the parallelization, the actual processing is done in the next PID. The script deletes files that may be present from previous runs, and prepares a list of RINEX files for parallel processing.
RNXSMT_H	RNXSMT is called to clean the RINEX data and identify cycle slips. Processing is done in parallel, i.e., the program receives a list of RINEX files to be cleaned in one run. The output consists of smoothed RINEX files (*.SMT) containing smoothed instead of original code observations. Cycle slips and outliers are marked.
RXOBV3AP	Prepare parallel run of RXOBV3
RXOBV3_H	Run RXOBV3, which converts RINEX observation files into Bernese internal format
CRDMRDAT	
CODSPPAP	Prepare the parallel run of CODSPP
CODSPP_P	Run CODSPP. The receiver clock synchronization is the main task of CODSPP. The program also determines approximate station coordinates by a straightforward point positioning approach, using only smoothed code observations. This is useful in case the a priori coordinates (e.g., from the RINEX header) are of doubtful quality. Outliers in the observations can be detected in this step, also.
CODXTR	This script runs the output extraction program for CODSPP. There are several such extraction programs available for Bernese processing programs, and they allow an efficient, quick overview of the results for quality control. In this case, CODXTR produces a summary file informing on detected outliers, missing orbits, excluded satellites due to satellite problems, and missing clocks. The file is listed in the processing summary. A high rms error and/or a lot of outliers indicate a site with bad code tracking performance.
<b>Compute PPP solutions station by station (including data screening)</b>	
MAUPRPAP	Prepare the parallel run for program MAUPRP.
MAUPRP_P	Run MAUPRP, which automatically preprocesses phase observations (cycle slip detection and correction, outlier detection, update of the ambiguity list)
PPPEDTAP	Prepare the parallel run of GPSEST

*Continued on next page*

Table A.3 – Continued from previous page

PPPEDT_P	<p>The following processing programs are called to perform a station by station data cleaning process, and to compute the precise point positioning solution. Iteration over GPSEST to generate a residual file for data screening, based on the ionosphere-free linear combination (L3). Normalized residuals are written, as elevation dependent weighting of observations is applied. RESRMS to screen the residual file, SATMRK to mark identified outliers in the observation files. The actual observation data remains in the files, the corresponding records are flagged as bad.</p> <p>Thereafter</p> <p>GPSEST: basically the same run as above, this time based on the cleaned observations.</p> <p>ADDNEQ2 to generate PPP result files for each station in Bernese and external formats</p>
GPSXTR	Output extractor for program GPSEST, producing an overview of the PPP solution for the processing summary, and the data cleaning.
PPPRESAP	Prepare the parallel run of PPPRES
PPPRES_P	Run PPPRES: additional run of GPSEST saving real residuals
RESFMT	Conversion of residuals from binary to ASCII format
RES_SUM	Generates summary
CRDMERGE	Combination of the single coordinate files
ADDNEQ2	Combination of the single normal equation files; ADDNEQ2 is used to write a single SINEX and a combined normal equation file containing all stations based on the normal equation files generated in PID 302. The troposphere parameters are pre-eliminated in this step. The SINEX sections related to station information and the estimated coordinates are listed in the processing summary.
PPP_HLM	CCRNXC The station specific clock RINEX files are combined to obtain one RINEX clock file. The processing summary contains the resulting CCRNXC output providing information on the selected reference clock and the performance of a polynomial fit of the estimated clock corrections.
ADDNEQ2	ADDNEQ2 is used to write a single SINEX and a combined normal equation file containing all stations based on the normal equation files generated in PID 302. The troposphere parameters are pre-eliminated in this step. The SINEX sections related to station information and the estimated coordinates are listed in the processing summary.
<b>Create summary files and delete files</b>	
GS01_SUM	Creates processing summary
GS01_SAV	Saves result files in archive directory
GS01_DEL	Deletes contents of the sub directories within the campaign directory
BPE_CLN	Deletes log files

**Table A.1:** Parameters for the Niell hydrostatic mapping from Niell [1996]

coefficient	latitude				
	15°	30°	45°	60°	75°
	<i>average</i>				
a	1.2769934e-3	1.2683230e-3	1.2465397e-3	1.2196049e-3	1.2045996e-3
b	2.9153695e-3	2.9152299e-3	2.9288445e-3	2.9022565e-3	2.9024912e-3
c	62.610505e-3	62.837393e-3	63.721774e-3	63.824265e-3	64.258455e-3
	<i>amplitude</i>				
a	0.0	1.2709626e-5	2.6523662e-5	3.4000452e-5	4.1202191e-5
b	0.0	2.1414979e-5	3.0160779e-5	7.2562722e-5	11.723375e-5
c	0.0	9.0128400e-5	4.3497037e-5	84.795348e-5	170.37206e-5
	<i>height correction</i>				
		a <sub>ht</sub>	2.53e-5		
		b <sub>ht</sub>	5.49e-3		
		c <sub>ht</sub>	1.14e-3		

**Table A.2:** Parameters for the Niell wet mapping from Niell [1996]

coefficient	latitude				
	15°	30°	45°	60°	75°
a	5.8021897e-4	5.6794847e-4	5.8118019e-4	5.9727542e-4	6.1641693e-4
b	1.4275268e-3	1.5138625e-3	1.4572752e-3	1.5007428e-3	1.7599082e-3
c	4.3472961e-2	4.6729510e-2	4.3908931e-2	4.4626982e-2	5.4736038e-2



# Acronyms

1D	one dimensional.
2D	two dimensional.
3D	three dimensional.
AMSU	Advanced Microwave Sounding Unit.
ART	Algebraic Reconstruction Techniques.
ATOVS	TIROS-N Operational Vertical Sounder.
Bernese	Berner GPS Software.
COSMIC	Constellation Observing System for Meteorology, Ionosphere, and Climate.
CS	Compressive Sensing.
iDCT	inverse Discrete Cosine Transform.
DEM	Digital Elevation Models.
DOY	Day Of Year.
DWD	German Meteorological Office (Deutscher Wetterdienst).
ECEF	Earth Centered Earth Fixed.
EGG97	European Gravity Geoid 1997.
EGM96	Earth Gravitational Model 1996.
ENU	East North Up.
EOP	Earth Orientation Parameters.
Envisat	Environmental Satellite.
GEONET	GPS Earth Observation Network.
GLONASS	Global'naya Navigatsionnaya Sputnikova Sistema.
GNSS	Global Navigation Satellite Systems.
GNSS RO	GNSS radio occultation.
GPS	Global Positioning System.
GRS80	Geodetic Reference System 1980.
IERS	International Earth Rotation and Reference Systems Service.
IGS	International GNSS Service.
IR	infrared.
InSAR	Interferometric SAR.
LiDAR	Light Detection And Ranging.
LSQ	Least Squares.
LUBW	Landesanstalt für Umwelt, Messungen und Naturschutz Baden-Württemberg.

MERIS	ME <sup>DI</sup> um Resolution Imaging Spectrometer.
MetOp	MEteorological Operational Satellite.
MODIS	MO <sup>DE</sup> rate Resolution Imaging Spectrometer.
NOAA	American National Oceanic and Atmospheric Administration.
NP	Non-deterministic Polynomial-time.
NWP	Numerical Weather Prediction.
NWM	Numerical Weather Models.
PDF	Probability Density Function.
PPP	Precise Point Positioning.
PS	Persistent Scatterer.
PSI	Persistent Scatterer Interferometry.
RADAR	RA <sup>DI</sup> o Detection And Ranging.
RINEX	Receiver Independent Exchange Format.
RIP	Restricted Isometry Property.
SAPOS <sup>®</sup>	German satellite positioning service.
SAR	Synthetic Aperture Radar.
SLAR	Side-Looking Airborne Radar.
SLC	Single Look Complex.
TEC	Total Electron Content.
TIROS-N	Television Infra-Red Observing Satellite-N.
URG	Upper Rhine Graben.
VLBI	Very Long Baseline Interferometry.
voxels	volumetric pixels.
WGS84	World Geodetic System 1984.
WRF	Weather Research and Forecasting Model.
WVR	Water Vapor Radiometers.

# List of symbols

Symbol	Unit	Description
$\alpha$	–	Raytracing: coefficient for computing intersection with constant latitudes
$\alpha$		Kronecker multiplication: row of matrix $A$
$\alpha_{CS}$	–	Parameter describing the steepness of the Euler base functions
$\beta$	–	Raytracing: coefficient for computing intersection with constant latitudes
$\beta$		Kronecker multiplication: row of matrix $B$
$\gamma$	–	Raytracing: coefficient for computing intersection with constant latitudes
$\delta$	–	Compressive Sensing: restricted isometry constant
$\epsilon$		Elevation angle
$\epsilon_{cut}$		Cutoff elevation angle
$\epsilon_{\rho_{pseudo}}$		Measurement noise in pseudo-range
$\epsilon_{\Phi}$		Measurement noise in carrier-phase
$\theta$		Compressive Sensing: sensing matrix
$\theta_{inc}$		Radar: incidence angle
$\kappa$	–	Degree of sparsity of a signal
$\lambda$		Longitude
$\lambda_{\Phi}$	m	Carrier-phase wavelength
$\lambda_{radar}$	m	Radar: wavelength
$\mu$	–	Compressive Sensing: mutual coherence between $\Phi$ and $\Psi$
$\nu$	m	Leveling segment
$\nu$		Compressive Sensing: any p-sparse vector
$\varphi$		Latitude
$\varphi_{n_{\Phi}}$	m	Row of design matrix
$\rho_{geom}$	m	Geometric range between the GNSS satellite and the receiver
$\rho_{pseudo}$	m	Pseudo range observations
$\sigma$		Singular value
$u$		Random variable drawn from distribution
$\chi$		Random variable
$\psi_{n_{\Psi}}$		Compressive Sensing: Column of sparsifying basis
$\omega$	–	Index of the dimension of a signal
$\Gamma$	–	Parameter weighing different terms of a cost function
$\Delta E$	m	Horizontal gradient in Easting
$\Delta N$	m	Horizontal gradient in Northing
$\Delta t$	s	Receiver clock offset
$\Delta ZWD$	m	Short-scale ZWD difference between two SAR acquisition times
$\Pi$	–	Conversion factor between IWV and ZWD
$\Phi$		Carrier phase observations
$\Phi$	m	Design matrix
$\Phi$	m	Single entry of the design matrix
$\Phi_{neu}$		InSAR phase difference

Symbol	Unit	Description
$\mathcal{X}$		Probability distribution
$\Psi$		Compressive Sensing: sparsifying basis
$a$	–	Voxel index in latitude
$\mathbf{a}$		Kronecker multiplication: column of matrix $\mathbf{A}$
$a_{i_A j_A}$		Kronecker multiplication: entry of matrix $\mathbf{A}$
$a_{\text{ell}}$	m	Semi-major axis of an ellipsoid
$a_{\text{mf}}$		Parameter in Niell mapping function
$b$	–	Voxel index in longitude
$\mathbf{b}$		Kronecker multiplication: column of matrix $\mathbf{B}$
$b_{i_B j_B}$		Kronecker multiplication: entry of matrix $\mathbf{B}$
$b_{\text{mf}}$		Parameter in Niell mapping function
$c$	m/s	Speed of light
$c_{\text{mf}}$		Parameter in Niell mapping function
$dl$	m	Differential along slant path
$e$	mb	Partial pressure of water vapor
$e_0$	mb	Partial pressure of water vapor at the surface
$e_{\text{ell}}$	–	First eccentricity of an ellipsoid
$f_1$	1/s	GPS frequency at wavelength 19.05 cm
$f_2$	1/s	GPS frequency at wavelength 24.45 cm
$g$	m/s <sup>2</sup>	Gravity
$h$	m	Height
$h_{\text{ell}}$	m	Ellipsoidal height
$i$	–	Ray index
$i_A$	–	Row number of matrix $\mathbf{A}$
$i_B$	–	Row number of matrix $\mathbf{B}$
$j$	–	Overall voxel index
$j_A$	–	Column number of matrix $\mathbf{A}$
$j_B$	–	Column number of matrix $\mathbf{B}$
$k$	–	Voxel index in height
$\ell$	–	Parameter index
$l_{\rho_{\text{pseudo}}}$	m	Ionosphere-free linear combination of dual-frequency pseudo-range observations
$l_{\Phi}$	m	Ionosphere-free linear combination of dual-frequency carrier-phase observations
$m$	–	Voxel index in latitude
$\text{mf}$	–	Mapping function
$m_{\text{dry air}}$	kg	Mass of dry air
$m_{\text{total air}}$	kg	Mass of total air
$m_{\text{water vapor}}$	kg	Mass of water vapor
$n$	–	Refractive index
$n_{\text{bins}}$	–	Number of bins
$n_{\Phi}$	–	Number of row of design matrix
$n_{\Psi}$	–	Compressive Sensing: number of column of sparsifying basis
$n_{\text{steepness}}$	–	Number of steepness parameters
$p$	mb	Pressure
$\mathbf{p}$	–	Real number describing an $L_{n_b, p}$ -norm
$p_0$	mb	Pressure at the surface
$q$	–	Voxel index in longitude
$q_v$	g/kg	Specific humidity
$r$	–	Raytracing: free parameter describing the ray path

<b>Symbol</b>	<b>Unit</b>	<b>Description</b>
$r_1$	–	Parameter index in the definition of the base functions
$r_2$	–	Parameter index in the definition of the base functions
$r_{\text{average}}$	m	Radius of averaging cone
$s$	–	Raytracing: free parameter describing the plane limiting a voxel in longitude
$\mathbf{s}$		Sparse parameters in transform domain
std		Standard deviation
$t$	–	Raytracing: free parameter describing the plane limiting a voxel in longitude
$t_{\text{mf}}$	days	Time from January 0.0 for the computation of the coefficients for the hydrostatic Niell mapping function
$t_{\text{mf},0}$		Adopted Phase (DOY 28) for the computation of the coefficients for the hydrostatic Niell mapping function
$v$	m/s	Velocity
$\mathbf{v}$		Compressive Sensing: basis vector 1
$\tilde{\mathbf{v}}$		Compressive Sensing: basis vector 2
$w$		Least Squares: Cutoff eigenvalue
$\mathbf{w}$		$L_1$ -norm vs. $L_2$ -norm: second parameter vector
$w_v$	g/kg	Water vapor mixing ratio
$x$		Component of parameter vector in spatial domain
$\mathbf{x}$		Parameter vector in spatial domain
$x_{\text{rec}}$	m	Receiver position in $X$ direction
$\mathbf{y}$		Observations
$y_i$		Observation of ray $i$
$y_{\text{rec}}$	m	Receiver position in $Y$ direction
$z$		Zenith angle
$z_{\text{rec}}$	m	Receiver position in $Z$ direction
$A$		Azimuth angle
$\mathbf{A}$		Kronecker multiplication: matrix
$B$		Correction term for Saastamoinen model
$\mathbf{B}$		Kronecker multiplication: matrix
$C$	$\text{m}^2/\text{s}^2$	Geopotential number
$\mathcal{C}$		iDCT letters
$D$		Correction term for Saastamoinen model
$\mathcal{D}$		Dirac letters
$E$	bits	Entropy
$\mathcal{E}$		Euler letters
$H$	m	Orthometric height
$H_{\text{geopotential}}$	m	Geopotential height
$H_{\text{scale}}$	m	Scale height
$I$	bits	Information content
$\mathbf{I}$	–	Identity matrix
IWV	$\text{kg}/\text{m}^2$	Integrated Water Vapor
$\mathcal{K}$	–	Number of voxels in height
$L$	–	Number of parameters
$L_0$		$L_0$ -norm minimization
$L_{0.5}$		$L_{0.5}$ -norm minimization
$L_1$		$L_1$ -norm minimization
$L_2$		$L_2$ -norm minimization
$L_4$		$L_4$ -norm minimization

<b>Symbol</b>	<b>Unit</b>	<b>Description</b>
$L_\infty$		$L_\infty$ -norm minimization
$L_p$		$L_p$ -norm minimization
$M$	–	Number of parameters in the transform domain
$N$	–	Number of observations
$N$		Ambiguity of the carrier-phase ionosphere-free linear combination
$N_{\text{hydrostatic}}$	mm/km	Hydrostatic refractivity
$N_{\text{total}}$	mm/km	Total refractivity
$N_{\text{wet}}$	mm/km	Wet refractivity
$N_\varphi$	m	Radius of curvature of a reference ellipsoid
$P$		Point on the Earth's surface
$\mathbf{P}$	–	Weight matrix
$P$	–	Probability
$\mathcal{P}$	–	Number of voxels in longitude
$P_0$		Point on the geoid
PH	mb	Perturbation geopotential
PHB	mb	Base state geopotential
PW	mm	Precipitable Water Vapor
$Q$	–	Conversion factor between IWV and ZWD
$Q$	–	Number of voxels in latitude
$R$	–	Number of voxels in a considered direction
$R_{\text{cone}}$	m	Raytracing: radius of the cone of constant latitudes
$R_{\text{Earth}}$	m	Radius of the Earth
RH	%	Relative humidity
$S$		Solution space
STD	m	Slant Total Delay
SWD	m	Slant Wet Delay
$T$	K	Temperature
$T_0$	K	Temperature at the surface
$T_d$	K	Dew point temperature
$T_m$	K	Neutrospheric mean temperature
$U$	m	Geoid undulation
$W$	$\text{m}^2/\text{s}^2$	Geopotential
$X$	m	ECEF coordinate in first direction
$\mathbf{X}$	m	ECEF coordinates of a point $(X, Y, Z)$
$\mathbf{X}_0$	m	Raytracing: ECEF coordinates of ray origin $(X_0, Y_0, Z_0)$
$X_{\text{SV}}$	m	Satellite position in $Z$ direction
$Y$	m	ECEF coordinate in second direction
$Y_{\text{SV}}$	m	Satellite position in $Y$ direction
$Z$	m	ECEF coordinate in third direction
$Z_{\text{SV}}$	m	Satellite position in $Z$ direction
ZDD	m	Zenith Dry Delay
ZHD	m	Zenith Hydrostatic Delay
ZTD	m	Zenith Total Delay
ZWD	m	Zenith Wet Delay

## List of constants

Symbol	Value	Unit	Description
$\epsilon'$	0.622		Ratio between gas constants of dry air and of water vapor
$g$	9.80665	m/s <sup>2</sup>	Mean Earth gravity acceleration
$k_1$	77.6	K/hPa	Meteorological constant
$k_2$	72	K/hPa	Meteorological constant
$k_2'$		K/hPa	Meteorological constant
$k_3$	$3.75 \cdot 10^5$	K <sup>2</sup> /hPa	Meteorological constant
$L_T$	6.5	°/km	Temperature lapse rate
$M_{\text{water vapor}}$	18.0153	g/mol	Molar mass of water vapor
$M_{\text{dry air}}$	28.9647	g/mol	Molar mass of dry air
$R$	8.31447	J/(mol K)	Universal gas constant





# Bibliography

- Adavi, Z. and Mashhadi-Hossainali, M. (2015). 4D-tomographic reconstruction of water vapor using the hybrid regularization technique with application to the North West of Iran. *Advances in Space Research*, 55(7):1845–1854.
- Aguilera, E., Nannini, M., and Reigber, A. (2013). A data-adaptive compressed sensing approach to polarimetric SAR tomography of forested areas. *IEEE Geoscience and Remote Sensing Letters*, 10(3):543–547.
- Alonso, M. T., López-Dekker, P., and Mallorquí, J. J. (2010). A novel strategy for radar imaging based on compressive sensing. *IEEE Transactions on Geoscience and Remote Sensing*, 48(12):4285–4295.
- Alshawaf, F. (2013). *Constructing water vapor maps by fusing InSAR, GNSS and WRF data*. PhD thesis, Karlsruhe Institute of Technology.
- Alshawaf, F., Fersch, B., Hinz, S., Kunstmann, H., Mayer, M., and Meyer, F. (2015a). Water vapor mapping by fusing InSAR and GNSS remote sensing data and atmospheric simulations. *Hydrology and Earth System Sciences*, 19(12):4747–4764.
- Alshawaf, F., Fuhrmann, T., Knopfler, A., Luo, X., Mayer, M., Hinz, S., and Heck, B. (2015b). Accurate Estimation of Atmospheric Water Vapor Using GNSS Observations and Surface Meteorological Data. *IEEE Transactions on Geoscience and Remote Sensing*, 53(7):3764–3771.
- Alshawaf, F., Hinz, S., Mayer, M., and Meyer, F. (2015c). Constructing accurate maps of atmospheric water vapor by combining interferometric synthetic aperture radar and GNSS observations. *Journal of Geophysical Research: Atmospheres*, 120:1391–1403.
- Annadurai, S. (2007). *Fundamentals of digital image processing*. Pearson Education India.
- Baraniuk, R. (2007). Compressive sensing. *IEEE signal processing magazine*, 24(4).
- Baraniuk, R., Davenport, M. A., Duarte, M. F., and Hegde, C. (2011). An introduction to compressive sensing. *Connexions e-textbook*.
- Barclay, L. (2003). *Propagation of radiowaves*, volume 2. IET.
- Bender, M., Dick, G., Ge, M., Deng, Z., Wickert, J., Kahle, H.-G., Raabe, A., and Tetzlaff, G. (2011). Development of a GNSS water vapour tomography system using algebraic reconstruction techniques. *Advances in Space Research*, 47(10):1704 – 1720. {GNSS} Remote Sensing-2.
- Benevides, P., Nico, G., Catalão, J., and Miranda, P. (2016). Bridging InSAR and GPS Tomography: A New Differential Geometrical Constraint. *IEEE Transactions on Geoscience and Remote Sensing*, 54(2):697–702.
- Berg, H. (1948). *Allgemeine Meteorologie: Einführung in die Physik der Atmosphäre*. Dümmler, Bonn.
- Bevis, M., Businger, S., Herring, T. A., Rocken, C., Anthes, R. A., and Ware, R. H. (1992). GPS meteorology: Remote sensing of atmospheric water vapor using the Global Positioning System. *Journal of Geophysical Research: Atmospheres*, 97(D14):15787–15801.
- Bi, Y., Mao, J., and Li, C. (2006). Preliminary results of 4-D water vapor tomography in the troposphere using GPS. *Advances in Atmospheric Sciences*, 23(4):551–560.
- Bock, O., Bouin, M.-N., Walpersdorf, A., Lafore, J. P., Janicot, S., Guichard, F., and Agusti-Panareda, A. (2007). Comparison of ground-based GPS precipitable water vapour to independent observations and NWP model reanalyses over Africa. *Quarterly Journal of the Royal Meteorological Society*, 133(629):2011–2027.
- Bock, O., Keil, C., Richard, E., Flamant, C., and Bouin, M.-n. (2005). Validation of precipitable water from ECMWF model analyses with GPS and radiosonde data during the MAP SOP. *Quarterly Journal of the Royal Meteorological Society*, 131(612):3013–3036.
- Böhm, J. (2004). *Troposphärische Laufzeitverzögerungen in der VLBI*. Inst. für Geodäsie u. Geophysik.
- Böhm, J. and Schuh, H. (2013). *Atmospheric effects in space geodesy*, volume 5. Springer.
- Boussakta, S. and Holt, A. G. J. (1995). New separable transform. *IEE Proceedings – Vision, Image and Signal Processing*, 142(1):27–30.
- Bowman, J. (1990). The greenhouse effect. *Land use policy*, 7(2):101–108.
- Braun, J. J. (2004). *Remote Sensing of Atmospheric Water Vapor with the Global Positioning System*. PhD thesis, University of Colorado.

- Budillon, A., Evangelista, A., and Schirinzi, G. (2011). Three-dimensional SAR focusing from multipass signals using compressive sampling. *IEEE Transactions on Geoscience and Remote Sensing*, 49(1):488–499.
- Burde, J.-P., Wilhelm, T., and Kuhn, J. (2015). Höhenmessungen über Luftdruckmessungen. *Physik in unserer Zeit*, 46(6):307–308.
- Caballero, R. (2014). *Physics of the Atmosphere*. 2053-2563. IOP Publishing.
- Cai, T. T., Wang, L., and Xu, G. (2010). New bounds for restricted isometry constants. *Information Theory, IEEE Transactions on*, 56(9):4388–4394.
- Candès, E. J. et al. (2006). Compressive sampling. In *Proceedings of the international congress of mathematicians*, volume 3, pages 1433–1452. Madrid, Spain.
- Candès, E. J. and Wakin, M. B. (2008). An introduction to compressive sampling. *Signal Processing Magazine, IEEE*, 25(2):21–30.
- Champollion, C., Flamant, C., Bock, O., Masson, F., Turner, D., and Weckwerth, T. (2009). Mesoscale GPS tomography applied to the 12 June 2002 convective initiation event of IHOP\_2002. *Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography*, 135(640):645–662.
- Champollion, C., Masson, F., Bouin, M.-N., Walpersdorf, A., Doerflinger, E., Bock, O., and Van Baelen, J. (2004). GPS water vapour tomography: preliminary results from the ESCOMPTE field experiment. *Atmospheric research*, 74(1):253–274.
- Chen, B. and Liu, Z. (2014). Voxel-optimized regional water vapor tomography and comparison with radiosonde and numerical weather model. *Journal of geodesy*, 88(7):691–703.
- Chen, B. and Liu, Z. (2016). Assessing the Performance of Troposphere Tomographic Modeling Using Multi-Source Water Vapor Data During Hong Kong’s Rainy Season from May to October 2013. *Atmospheric Measurement Techniques Discussions*, 2016:1–23.
- Chen, G. and Herring, T. (1997). Effects of atmospheric azimuthal asymmetry on the analysis of space geodetic data. *Journal of Geophysical Research: Solid Earth*, 102(B9):20489–20502.
- Curry, J. A. and Webster, P. J. (1999). *Thermodynamics of atmospheres and oceans*. International geophysics series ; 65. Acad. Press, San Diego, Calif. [u.a.]. Includes bibliographical references and index; : No price : Formerly CIP.
- Dabberdt, W., Shellhorn, R., Cole, H., Paukkunen, A., Hörhammer, J., and Antikainen, V. (2002). Radiosondes.
- Dach, R., Fridez, P., Hugentobler, U., and Meindl, M. (2007). Bernese GPS software version 5.0. *Astronomical Institute, University of Berne*.
- Davis, J., Herring, T., Shapiro, I., Rogers, A., and Elgered, G. (1985). Geodesy by radio interferometry: Effects of atmospheric modeling errors on estimates of baseline length. *Radio science*, 20(6):1593–1607.
- Davis, J. L., Elgered, G., Niell, A. E., and Kuehn, C. E. (1993). Ground-based measurement of gradients in the “wet” radio refractivity of air. *Radio Science*, 28(06):1003–1018.
- Ding, N., Zhang, S., Wu, S., Wang, X., and Zhang, K. (2018). Adaptive node parameterization for dynamic determination of boundaries and nodes of GNSS tomographic models. *Journal of Geophysical Research: Atmospheres*, 123(4):1990–2003.
- Ding, N., Zhang, S., and Zhang, Q. (2017). New parameterized model for GPS water vapor tomography. *Annales Geophysicae*, 35(2):311–323.
- Donoho, D. L. (2006). Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306.
- Duarte, M. F. and Baraniuk, R. G. (2012). Kronecker compressive sensing. *IEEE Transactions on Image Processing*, 21(2):494–504.
- Elosegui, P., Ruis, A., Davis, J., Ruffini, G., Keihm, S., Bürki, B., and Kruse, L. (1998). An experiment for estimation of the spatial and temporal variations of water vapor using GPS data. *Physics and Chemistry of the Earth*, 23(1):125 – 130.
- Fischer, J. and Bennartz, R. (1997). Retrieval of total water vapour content from MERIS measurements. *Algorithm Theoretical Basis Document PO-NT-MEL-GS-005, ESA-ESTEC*, 4.
- Flores, A., De Arellano, J.-G., Gradinarsky, L. P., and Rius, A. (2001). Tomography of the lower troposphere using a small dense network of GPS receivers. *IEEE Transactions on Geoscience and Remote Sensing*, 39(2):439–447.

- Flores, A., Ruffini, G., and Rius, A. (2000). 4D tropospheric tomography using GPS slant wet delays. In *Annales Geophysicae*, volume 18. Springer.
- Forsell, B. (2008). *Radionavigation systems*. GNSS Technology and Applications Series. Artech House, Boston.
- Fuhrmann, T. (2010). *Zur GNSS-basierten Bestimmung des atmosphärischen Wasserdampfgehalts mittels precise point positioning*. KIT Scientific Publishing.
- Fuhrmann, T., Heck, B., Knöpfler, A., Masson, F., Mayer, M., Ulrich, P., Westerhaus, M., and Zippelt, K. (2013). Recent surface displacements in the Upper Rhine Graben – Preliminary results from geodetic networks. *Tectonophysics*, 602(Supplement C):300 – 315. TOPO-EUROPE III.
- Gradinarsky, L. and Jarlemark, P. (2002). GPS tomography using the permanent network in Goteborg: simulations. In *Position Location and Navigation Symposium, 2002 IEEE*, pages 128–133. IEEE.
- Gradinarsky, L. P. and Jarlemark, P. (2004). Ground-based GPS tomography of water vapor: Analysis of simulated and real data. *Journal of the Meteorological Society of Japan*, 82(1B):551–560.
- Grohnfeldt, C., Zhu, X. X., and Bamler, R. (2013). Jointly sparse fusion of hyperspectral and multispectral imagery. In *Geoscience and Remote Sensing Symposium (IGARSS), 2013 IEEE International*, pages 4090–4093. IEEE.
- Hadamard, J. (1902). Sur les problèmes aux dérivées partielles et leur signification physique. *Princeton University Bulletin*, 13(49-52):28.
- Hajj, G. A., Ibañez-Meier, R., Kursinski, E. R., and Romans, L. J. (1994). Imaging the ionosphere with the global positioning system. *International Journal of Imaging Systems and Technology*, 5(2):174–187.
- Hansen, M. H. and Yu, B. (2001). Model selection and the principle of minimum description length. *Journal of the American Statistical Association*, 96(454):746–774.
- Heck, B. (2003). *Rechenverfahren und Auswertemodelle der Landesvermessung: Klassische und moderne Methoden; 3*. Wichmann.
- Henderson, H. V. and Searle, S. R. (1981). The vec-permutation matrix, the vec operator and Kronecker products: A review. *Linear and multilinear algebra*, 9(4):271–288.
- Heublein, M., Alshawaf, F., Erdnöß, B., Zhu, X. X., and Hinz, S. (2018). Compressive sensing reconstruction of 3D wet refractivity based on GNSS and InSAR observations. *Journal of Geodesy*.
- Hirahara, K. (2000). Local GPS tropospheric tomography. *Earth, Planets and Space*, 52(11):935–939.
- Hofmann-Wellenhof, B., Lichtenegger, H., and Wasle, E. (2008). *GNSS – Global Navigation Satellite Systems : GPS, GLONASS, Galileo and more*. Springer, Wien.
- Hurter, F. and Maier, O. (2013). Tropospheric profiles of wet refractivity and humidity from the combination of remote sensing data sets and measurements on the ground. *Atmospheric Measurement Techniques*, 6(11):3083–3098.
- Huschke, R. E. et al. (1959). Glossary of meteorology.
- Jehle, M., Rüegg, M., Small, D., Meier, E., and Nüesch, D. (2005). Estimation of ionospheric TEC and Faraday rotation for L-band SAR. In *Remote Sensing of Clouds and the Atmosphere X*, volume 5979, page 59790U. International Society for Optics and Photonics.
- Jiang, C., Zhang, H., Shen, H., and Zhang, L. (2014). Two-step sparse coding for the pan-sharpening of remote sensing images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(5):1792–1805.
- Kelley, M. C. (2009). *The Earth's ionosphere: plasma physics and electrodynamics*, volume 96. Academic press.
- Knöpfler, A. (2015). *Korrektur stationsabhängiger Fehler bei GNSS*. KIT Scientific Publishing.
- Kouba, J. and Héroux, P. (2001). Precise Point Positioning Using IGS Orbit and Clock Products. *GPS Solutions*, 5(2):12–28.
- Kraus, H. (2007). *Die Atmosphäre der Erde: Eine Einführung in die Meteorologie*. Springer-Verlag.
- Lawrence, M. G. (2005). The relationship between relative humidity and the dewpoint temperature in moist air: A simple conversion and applications. *Bulletin of the American Meteorological Society*, 86(2):225–234.
- League, A. R. R. (1997). *The ARRL antenna book*, volume 18. The League.
- Lewis, H. (2007). Geodesy calculations in ROPP. GRAS SAF Report 02, Met Office, UK.

- Li, S. and Yang, B. (2011). A new pan-sharpening method using a compressed sensing technique. *IEEE Transactions on Geoscience and Remote Sensing*, 49(2):738–746.
- Mallat, S. (1999). *A wavelet tour of signal processing*. Academic press.
- Menke, W. (2012). *Geophysical data analysis: discrete inverse theory: MATLAB edition*, volume 45. Academic press.
- Mockler, S. (1995). Water vapor in the climate system, special report, american geophysical union (agu). Technical report.
- Nearing, G. S., Tian, Y., Gupta, H. V., Clark, M. P., Harrison, K. W., and Weijs, S. V. (2016). A philosophical basis for hydrological uncertainty. *Hydrological Sciences Journal*, 61(9):1666–1678.
- Niell, A. (1996). Global mapping functions for the atmosphere delay at radio wavelengths. *Journal of Geophysical Research: Solid Earth*, 101(B2):3227–3246.
- Niell, A., Coster, A., Solheim, F., Mendes, V., Toor, P., Langley, R., and Upham, C. (2001). Comparison of measurements of atmospheric wet delay by radiosonde, water vapor radiometer, GPS, and VLBI. *Journal of Atmospheric and Oceanic Technology*, 18(6):830–850.
- Nilsson, T. and Elgered, G. (2007). Water vapour tomography using GPS phase observations: Results from the ESCOMPTE experiment. *Tellus A: Dynamic Meteorology and Oceanography*, 59(5):674–682.
- Nilsson, T., Gradinarsky, L., and Elgered, G. (2005). Estimating the 3D structure of the atmospheric water vapor using GPS phase observations. In *Proc. RVK*, pages 521–526.
- Nishimura, T., Munekane, H., and Yarai, H. (2011). The 2011 off the Pacific coast of Tohoku Earthquake and its aftershocks observed by GEONET. *Earth, Planets and Space*, 63(7):22.
- Notarpietro, R., Cucca, M., Gabella, M., Venuti, G., and Perona, G. (2011). Tomographic reconstruction of wet and total refractivity fields from GNSS receiver networks. *Advances in Space Research*, 47(5):898–912.
- Notarpietro, R., Gabella, M., and Perona, G. (2008). Tomographic reconstruction of neutral atmospheres using slant and horizontal wet delays achievable through the processing of signal observed from small GPS networks. *Ital. J. Remote Sens*, 40(2):63–74.
- NTIA (2015). Manual of Regulations and Procedures for Federal Radio Frequency Management. U.S. Department of Commerce, National Telecommunications and Information Administration.
- Park, S. K. (1999). Nonlinearity and predictability of convective rainfall associated with water vapor perturbations in a numerically simulated storm. *Journal of Geophysical Research*, 104:31.
- Park, S. K. and Droegemeier, K. K. (2000). Sensitivity analysis of a 3D convective storm: Implications for variational data assimilation and forecast error. *Monthly weather review*, 128(1):140–159.
- Parker, A. L. (2017). *Systematic Assessment of Atmospheric Uncertainties for InSAR Data at Volcanic Arcs Using Large-Scale Atmospheric Models: Application to the Cascade Volcanoes*, pages 59–90. Springer International Publishing, Cham.
- Perler, D. (2011). *Water vapor tomography using global navigation satellite systems*. PhD thesis, Diss., Eidgenössische Technische Hochschule ETH Zürich, Nr. 20012, 2011.
- Perler, D., Geiger, A., and Hurter, F. (2011). 4D GPS water vapor tomography: new parameterized approaches. *Journal of Geodesy*, 85(8):539–550.
- Petit, G. and Luzum, B. (2010). IERS conventions (2010). Technical report, Bureau international des poids et mesures.
- Potter, L. C., Ertin, E., Parker, J. T., and Cetin, M. (2010). Sparsity and compressed sensing in radar imaging. *Proceedings of the IEEE*, 98(6):1006–1020.
- Pruente, L. (2010). Application of compressed sensing to SAR/GMTI-data. In *Synthetic Aperture Radar (EUSAR), 2010 8th European Conference on*, pages 1–4. VDE.
- Rauhut, H. (2010). Compressive sensing and structured random matrices. *Theoretical foundations and numerical methods for sparse recovery*, 9:1–92.
- Rauhut, H., Schnass, K., and Vandergheynst, P. (2008). Compressed sensing and redundant dictionaries. *IEEE Transactions on Information Theory*, 54(5):2210–2219.
- Riverson, Y. and Stern, A. (2009). Compressed imaging with a separable sensing operator. *IEEE Signal Processing Letters*, 16(6):449–452.
- Rizos, C., Janssen, V., Roberts, C., and Grinter, T. (2012). PPP versus DGNSS. *Geomatics world*, 20(6):18–20.

- Rocken, C., Ware, R., Van Hove, T., Solheim, F., Alber, C., Johnson, J., Bevis, M., and Businger, S. (1993). Sensing atmospheric water vapor with the global positioning system. *Geophysical Research Letters*, 20:2631–2634.
- Rocken, C. A. R. W. C. and Braun, J. (2000). Obtaining single path phase delays from GPS double differences.
- Rodgers, C. D. (2000). *Inverse methods for atmospheric sounding: theory and practice*, volume 2. World Scientific Publishing.
- Rohm, W. (2012). The precision of humidity in {GNSS} tomography. *Atmospheric Research*, 107(0):69 – 75.
- Rohm, W. (2013). The ground GNSS tomography–unconstrained approach. *Advances in Space Research*, 51(3):501–513.
- Rothacher, M. (2002). Estimation of station heights with GPS. In *Vertical reference systems*, pages 81–90. Springer.
- Rüeger, J. M. (2002). *Refractive indices of light, infrared and radio waves in the atmosphere*. School of Surveying and Spatial Information Systems, University of New South Wales.
- Saastamoinen, J. (1972). Contributions to the theory of atmospheric refraction. *Bulletin Géodésique (1946-1975)*, 105(1):279–298.
- Seeber, G., editor (2003). *Satellite geodesy*. de Gruyter, Berlin.
- Seidel, D. J. (2002). Water vapor: Distribution and trends. *Encyclopedia of Global Environmental Change*, John Wiley & Sons, Ltd, Chichester.
- Shapiro, S. S. and Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52(3/4):591–611.
- Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Barker, D. M., Duda, M. G., Huang, X.-Y., Wang, W., and Powers, J. G. (2008). A description of the advanced research WRF version 3. Technical report, DTIC Document.
- Smith, E. K. and Weintraub, S. (1953). The constants in the equation for atmospheric refractive index at radio frequencies. *Proceedings of the IRE*, 41(8):1035–1037.
- Song, S., Zhu, W., Ding, J., and Peng, J. (2006). 3D water-vapor tomography with Shanghai GPS network to improve forecasted moisture field. *Chinese Science Bulletin*, 51(5):607–614.
- Stull, R. (2016). *Practical Meteorology: an algebra based survey of atmospheric science*. BC Campus.
- Tang, W., Liao, M., and Yuan, P. (2016). Atmospheric correction in time-series SAR interferometry for land surface deformation mapping – A case study of Taiyuan, China. *Advances in Space Research*, 58(3):310 – 325.
- Tarantola, A. (2005). *Inverse problem theory and methods for model parameter estimation*. Society for Industrial and Applied Mathematics.
- Teunissen, P. and Montenbruck, O. (2017). *Springer handbook of global navigation satellite systems*. Springer.
- Torge, W. (2001). *Geodesy*. de Gruyter, Berlin [u.a.], 3., completely rev. and extended ed. edition.
- Troller, M. (2004). *GPS based determination of the integrated and spatially distributed water vapor in the troposphere*, volume 67.
- Troller, M., Geiger, A., Brockmann, E., Bettems, J.-M., Bürki, B., and Kahle, H.-G. (2006). Tomographic determination of the spatial distribution of water vapor using GPS observations. *Advances in Space Research*, 37(12):2211–2217.
- Tuller, S. E. (1973). Seasonal and annual precipitation efficiency in Canada. *Atmosphere*, 11(2):52–66.
- Wallace, J. M. and Hobbs, P. V. (2006). *Atmospheric science : an introductory survey*. International geophysics series; 92. Academic Press, Amsterdam.
- Wang, X., Zhang, K., Wu, S., Fan, S., and Cheng, Y. (2016). Water vapor-weighted mean temperature and its impact on the determination of precipitable water vapor and its linear trend. *Journal of Geophysical Research: Atmospheres*, 121(2):833–852.
- Whiteman, D., Melfi, S., and Ferrare, R. (1992). Raman lidar system for the measurement of water vapor and aerosols in the Earth’s atmosphere. *Applied Optics*, 31(16):3068–3082.
- Wiegand, G. (2016). *Gasmestechnik in Theorie und Praxis: Messgeräte, Sensoren, Anwendungen*. Springer-Verlag.
- Wiggins, R. A. (1972). The general linear inverse problem: Implication of surface waves and free oscillations for Earth structure. *Reviews of Geophysics*, 10(1):251–285.

- Xia, P., Cai, C., and Liu, Z. (2013). GNSS troposphere tomography based on two-step reconstructions using GPS observations and COSMIC profiles. In *Annales Geophysicae*, volume 31.
- Xin, L. and Reuter, G. (1996). Numerical simulation of the effects of mesoscale convergence on convective rain showers. *Monthly weather review*, 124(12):2828–2842.
- Xu, G. and Xu, Y. (2007). *GPS: theory, algorithms and applications*. Springer.
- Yao, Y. and Zhao, Q. (2016). Maximally Using GPS Observation for Water Vapor Tomography. *IEEE Transactions on Geoscience and Remote Sensing*, 54(12):7185.
- Yao, Y. and Zhao, Q. (2017). A novel, optimized approach of voxel division for water vapor tomography. *Meteorology and Atmospheric Physics*, 129(1):57–70.
- Yao, Y., Zhao, Q., and Zhang, B. (2016). A method to improve the utilization of GNSS observation for water vapor tomography. *Ann. Geophys*, 34:143–152.
- Yilmaz, N. (2008). Comparison of different height systems. *Geo-spatial Information Science*, 11(3):209–214.
- Zebker, H. A., Rosen, P. A., and Hensley, S. (1997). Atmospheric effects in interferometric synthetic aperture radar surface deformation and topographic maps. *Journal of Geophysical Research: Solid Earth*, 102(B4):7547–7563.
- Zhao, Q., Yao, Y., and Yao, W. (2018). Troposphere Water Vapour Tomography: A Horizontal Parameterised Approach. *Remote Sensing*, 10(8):1241.
- Zhu, X. X. and Bamler, R. (2010). Tomographic SAR inversion by  $L_1$ -norm regularization – The compressive sensing approach. *IEEE Transactions on Geoscience and Remote Sensing*, 48(10):3839–3846.
- Zhu, X. X. and Bamler, R. (2013). A sparse image fusion algorithm with application to pan-sharpening. *Geoscience and Remote Sensing, IEEE Transactions on*, 51(5):2827–2836.
- Zhu, X. X. and Bamler, R. (2014). Superresolving SAR tomography for multidimensional imaging of urban areas: Compressive sensing-based TomoSAR inversion. *IEEE Signal Processing Magazine*, 31(4):51–58.
- Zhu, X. X., Grohnfeldt, C., and Bamler, R. (2016). Exploiting joint sparsity for pansharpening: The J-SparseFI algorithm. *IEEE Transactions on Geoscience and Remote Sensing*, 54(5):2664–2681.
- Zumberge, J., Heflin, M., Jefferson, D., Watkins, M., and Webb, F. H. (1997). Precise point positioning for the efficient and robust analysis of GPS data from large networks. *Journal of Geophysical Research: Solid Earth*, 102(B3):5005–5017.

## **Acknowledgments**

Many thanks go to all data providers and financial supporters, without whom this work would not have been possible. I would like to thank Franz Ulmer for providing WRF data, Kurt Seitz for providing geoid undulations, and Fadwa Alshawaf for providing absolute InSAR SWDs. Many thanks go to the Graduate School of Climate and Environment as well as the Landesgraduiertenförderung Baden-Württemberg for financially supporting this thesis. In addition, special thanks go to the Karlsruhe House of Young Scientists that supported both a participation at the Global Young Scientists Summit in Singapore in 2016 and a three months research stay at the Istituto per le Applicazioni del Calcolo (IAC) dal Consiglio Nazionale delle Ricerche (CNR) di Bari, Italy. Moreover, many thanks to Deutsche Telekom Stiftung for its great support going far beyond financial aspects and for sure leaving a mark on each of its scholars.