Constructing water vapor maps by fusing InSAR, GNSS, and WRF data

Zur Erlangung des akademischen Grades eines

DOKTOR-INGENIEURS

von der Fakultät für

Bauingenieur-, Geo- und Umweltwissenschaften

des Karlsruher Instituts für Technologie

genehmigte

DISSERTATION

von

M.Sc. Fadwa Alshawaf

Tag der mündlichen Prüfung: 19.11.2013

Hauptreferent: Prof. Dr.-Ing. Stefan Hinz
Korreferent: Prof. Dr.-Ing. habil. Dr. h.c. Bernhard Heck
Korreferent: Prof. Dr.-Ing Franz J. Meyer

Karlsruhe 2013

Abstract

An accurate measurement of the content of water vapor in the atmosphere is a key requirement for the weather forecasting and climate research. Since water vapor affects the microwave signal propagating in the atmosphere by a high temporally and spatially variable delay, precise determination of water vapor content is also essential for remote sensing applications of global navigation satellite systems (GNSS) and interferometric synthetic aperture radar (InSAR). The specific objective of this work is to derive maps of the wet delay, caused by atmospheric water vapor, by analyzing the atmospheric delay "error" in GNSS and InSAR observations. GNSS provide absolute measurements of the wet delay at spatially sparse locations, while Persistent Scatterer InSAR (PSI) is applied to build temporally-differenced maps of the wet delay at a high spatial resolution. We present a method for estimating maps of the wet delay eliminated by building interferograms or later when reducing the topographic phase and orbital ramp based on the wet delay measurements at the sparse network of GNSS sites. Least squares inversion is applied to the wet delay-difference maps derived from PSI to obtain maps of the wet delay residuals at each SAR acquisition time, under the constraint of zero temporal mean. We combine the complementary maps from GNSS and PSI to build absolute maps of the wet delay at a high spatial resolution. Comparing the maps derived using the presented approach with integrated water vapor (IWV) maps from MERIS (MEdium Resolution Imaging Spectrometer) demonstrates strong spatial correlation up to 92%. The difference maps have a normal distribution with root mean square values below 1 mm (IWV), for 5 different examples.

The second objective of this work is to achieve a rigorous data fusion of the derived absolute wet delay and IWV maps simulated by the Weather Research and Forecasting Modeling System (WRF). The fusion strategy should properly address the problems of computational burden for massive data, incompatible support, gaps and noise. Therefore, we exploit the methodology of spatial statistical data fusion (SSDF). SSDF extends the approach of fixed-rank kriging (FRK), which is used for data prediction from single sets, for assimilating multiple data sets. The FRK covariance model highly reduces the computational complexity for obtaining the predictions and uncertainties, solves the change of data support (see 4) and it does not require the assumptions of stationarity and isotropy. The maps inferred based on the SSDF have better correlation with MERIS maps than those predicted from single data sets and the artifacts that might exist in single sets are suppressed. The results show that the difference between the maps computed by applying the SSDF method and MERIS has a root mean square value of less that 1 mm (IWV), while the maps predicted from single sets show larger values.

Kurzfassung

Eine präzise Bestimmung von atmosphärischem Wasserdampfgehalt ist von größer Bedeutung für Meteorologie und Klimaforschung. Elektromagnetische Wellen von Globalen Satellitennavigationssystemen (GNSS) oder RADAR Interferometrie (InSAR) werden auf dem Weg vom Sender zum Empfänger beeinträchtigt. Die Einflüsse der Erdatmosphäre können in iono- und neutrosphärische Einflüsse eingeteilt werden. Die Neutrosphäre ist für Mikrowellen mit Frequenzen weniger als 30 GHz nicht dispersiv und deren Einflüsse können nicht durch lineare Kombination von Mehrfrequenzmessung eliminiert werden. Die neutrosphärischen Einflüsse lassen sich in eine trockene Komponente, verursacht durch trockene Gase, und eine feuchte Komponente, verursacht durch Wasserdampf, unterteilen. Obwohl der Einfluss der feuchten Komponente weniger als 10% beträgt, enthält dieser jedoch wertvolle Informationen über den Wasserdampfgehalt in der Neutrosphäre. Diese Fehlerquelle wird in dieser Arbeit als nützliche Information benutzt um den atmosphärischen Wasserdampfgehalt zu bestimmen.

Die Nutzung von InSAR zur Wasserdampfbestimmung hat ein größes Potential, da aus InSAR 2D Wasserdampffelder in höher räumlicher Auflösung bestimmt werden können. Da die InSAR Messungen relative zu einem Referenzbild definiert werden, sind die modellierten Wasserdampfgehälter somit auch relative Werte. Im Gegensatz zu InSAR, werden absolute Werte des Wasserdampfgehaltes mit der Precise Point Positioning (PPP) Strategie aus GNSS Phasenbeobachtungen bestimmt. Im Hauptteil dieser Arbeit wird eine Strategie für die Kombination von InSAR und GNSS entwickelt um die gesamten räumlich hochaufgelösten integrierten Wasserdampfwerte (IWV) bestimmen zu können. Die Ergebnisse werden anschließend mit den integrierten Wasserdampfwerten aus MERIS (MEdium resolution Imaging Spectrometer) validiert. Die Validierung der Daten gegen MERIS zeigt eine hohe räumliche Korrelation, welche 92% erreicht. Die RMS-Werte der Differenzen zwischen den Wasserdampfgehaltkarten, die basierend auf der entwickelten Strategie ermittelt wurden, und deren von MERIS sind unter 1 mm IWV.

Ein weiteres Ziel der Arbeit ist es ein Verfahren zu entwickeln, das eine rigorose Fusion von den Wasserdampfgehaltkarten, welche den Wasserdampfgehalt aus der Kombination von InSAR und GNSS ableiten, und Karten von dem Model Weather Research and Forecasting (WRF) ermöglicht. Das Fusionsverfahren soll verschiedene Faktoren betrachten: Rechenzeit für größe Datensätze, change of support problems, Datenlücken, und Rauschen in den Daten. In dieser Arbeit wird das räumliche-statistische Datenfusionsverfahren (spatial statistical data fusion, SSDF) benutzt. Der Verfahren ist für die Datensätze von WRF und

InSAR-GNSS angepasst. Die Ergebnisse der Fusion werden anschließend mit den integrierten Wasserdampfwerten aus MERIS validiert. Die Ergebnisse der Fusion produzieren, im gegensatz zu den einzelnen Datensätzen, genauere und durchgängige IWV Karten.

Contents

List of Figures

x

List of Tables

List of Abbreviations

BS5	Bernese GNSS software, version 5.0
CC	Correlation coefficient
COSP	Change of support problem
DEM	Digital elevation model
DGNSS	Differential GNSS
E-M	Expectation-maximization
FRK	Fixed-rank kriging
GNSS	Global navigation satellite systems
GPS	Global positioning system
IDW	Inverse distance weighting
IGS	International GNSS Service
InSAR	Interferometric synthetic aperture radar
IR	Infrared
IWV	Integrated water vapor
MERIS	MEdium resolution Imaging Spectrometer
MF	Mapping function
MM5	Mesoscale Model
MM	Method-of-moments
MODIS	MODerate resolution Imaging Spectroradiometer
MP	Multipath
MSPE	Mean-squared prediction error
NWP	Numerical weather prediction
OK	Ordinary kriging
PCV	Phase center variations
PPP	Precise point positioning
PSI	Persistent scatterer InSAR
RMS	Root mean square
SAR	Synthetic aperture radar

SRE	Spatial random effects
SME	Spatial mixed model
SRTM	Shuttle Radar Topography Mission
StaMPS	Stanford Method for Persistent Scatterers
SSDF	Spatial statistical data fusion
SSNP	Site specific neutrospheric parameter
STD	Slant total delay
SWD	Slant wet delay
TEC	Total electron content
URG	Upper Rhine Graben
UTC	Coordinated universal time
WRF	Weather research and forecasting
ZDD	Zenith dry delay
ZTD	Zenith total delay
ZWD	Zenith wet delay

List of Symbols

A	Satellite azimuth angle
$\mathbf{C}$	Covariance function
$\mathbf{c}$	Covariance of the observations and the true process
c	Speed of light
$D(\mathbf{h})$	Structure function
e	Partial pressure of water vapor
f	Carrier frequency
g	Earth's gravitational acceleration
$\mathbf{K}$	Covariance matrix for fixed-rank kriging
$\mathbf{k}$	Kriging weighting vector
k_1, k_3, k_3	Physical constants
L	Temperature lapse rate
M_d	Molar mass of dry air
M_w	Molar mass of water vapor
MF_{Niell}	Niell mapping function
n	Neutrospheric refractive index
n_e	Electron density
n_I	Ionospheric refractive index
N	Neutrospheric refractivity
N_{dry}	Refractivity of dry gases
N_r^k	Integer number ambiguity
$N_{r,IF}^k$	Inosphere-free ambiguity
N_{wet}	Refractivity of water vapor
P	Air pressure
P_d^{gnss}	Pressure of dry air at GNSS site
P_d^{MET}	Pressure of dry air at meteorological station
P_s	Pressure of dry air

Q	Empirical constant
R	Universal gas constant
rh	Relative humidity
R_w	Specific gas constant of water vapor
$\mathbf{S}$	Matrix of weights for fixed-rank kriging
$S^{(l)}$	Basis function
$\mathbf{S}_p$	Weighting matrix of the predictions
STD^{gnss}	Slant total delay from GNSS
SWD^{gnss}	Slant wet delay from GNSS
$\mathbf{T}$	Matrix of covariates for linear trend
T_d	Dew point temperature
T_m	Weighted mean temperature of the atmosphere
T_{meteo}	Air temperature at the meteorology station
T_s	Surface temperature
W	Liquid water content
$\hat{Z}$	Estimated signal
$\tilde{\mathbf{Z}}$	Detrended observations
ZDD^{meteo}	Zenith dry delay obtained using meteorological data
ZDD^{model}	Zenith dry delay from Saastamoinen model
ZTD_{iso}^{gnss}	Azimuthally-isotropic zenith total delay from GNSS
ZTD^{model}	Zenith total delay from Saastamoinen model
ZWD^{gnss}	Zenith wet delay from GNSS
ZWD_{iso}^{gnss}	Azimuthally-isotropic zenith wet delay from GNSS
ZWD^{model}	Zenith wet delay from Saastamoinen model
z_{gnss}	Ellipsoidal altitude of the GNSS antenna
z_{meteo}	Ellipsoidal altitude of the Meteorology station
$Y(B_i)$	Block-level true process
$Y(\mathbf{s})$	Point-level true process
$\hat{\boldsymbol{\alpha}}$	Regression coefficients
$\gamma(\mathbf{h})$	Semivariogram
Δ^e	Easting neutrospheric gradients
Δ^{model}	model value of the neutrospheric delay

Δ^n	Northing neutrospheric gradients
Δ^r	Anisotropic wet delay extracted from the phase residuals in
Δ^z	Site-spacific neutrospheric paramter
ΔL_{iono}^{gnss}	Ionospheric delay from GNSS
ΔL_{neu}^{gnss}	Neutrospheric delay from GNSS
ΔL_{iono}	Ionospheric delay
$\Delta \phi^{psi}$	PSI unwrapped noise-reduced phase
$\Delta \psi^{psi}$	PSI residual interferometric phase
ΔSWD^{psi}	PSI slant wet delay-difference
ΔSTD^{psi}	PSI slant neutrospheric delay-difference
ΔZWD^{psi}	PSI zenith wet delay-difference
δt^k	Satellite clock error
δt_r	Receiver clock error
$\boldsymbol{\epsilon}$	Observation noise
ϵ_r^{gnss}	GNSS phase noise
$\boldsymbol{\zeta}$	Fine-scale signal of the geophysical process
$\boldsymbol{\eta}$	Hidden state vector for the FRK forward model
θ_{el}	Elevation angle
θ_{inc}	Incidence angle
λ	Wavelength
Π	Empirical constant
ρ_r^k	Geometrical distance from the kth satellite to the receiver r
ρ_w	Density of water vapor
$\boldsymbol{\Sigma}$	Covariance matrix
σ^2	Kriging mean-squared prediction error
$\sigma_\epsilon^2 \mathbf{V}_\epsilon$	Noise covariance matrix
$\sigma_\zeta^2 \mathbf{V}_\zeta$	Covariance matrix of the fine-scale signal
$\boldsymbol{\nu}$	Medium- to small-scale signal of the geophysical process
ϕ_{disp}	Displacement phase
ϕ_{flat}	Earth's reference phase
$\phi_r^{gnss,k}$	Carrier phase from GNSS
$\phi_{r,IF}^{gnss,k}$	Ionosphere-free GNSS phase

ϕ_{int}	Interferometric phase
ϕ_{iono}	Ionospheric phase
ϕ_{neu}	Neutrospheric phase
ϕ_{noise}	Phase noise
ϕ_{orbit}	Orbital phase
ϕ_{topo}	Topographic phase
ϕ	Carrier phase

1 Introduction

Over the past twenty years, repeat-pass spaceborne Interferometric Synthetic Aperture Radar (InSAR) has been widely used as a geodetic technique to generate maps of the Earth's topography and to measure surface deformation. In this work, we exploit InSAR as a meteorological tool to derive maps of water vapor content in the atmosphere. Water vapor resides in the electrically-neutral atmosphere (neutrosphere) even under clear skies, and its content shows high variations over time and space. It is the most active greenhouse gas in the atmosphere and it influences the Earth's radiation budget, energy transfer, cloud formation and precipitation. Therefore, it is a key element in climate studies and weather forecasting. Also, water vapor got high attention in the field of active microwave remote sensing systems, such as InSAR and Global Navigation Satellite Systems (GNSS). Microwave signals transmitted from remote sensing sensors are delayed by atmospheric water vapor, which increases the time they need to reach the ground. Although it contributes to less than 10% of the total neutrospheric delay, this delay is considered as a significant source of limitation in InSAR and GNSS applications. That is because water vapor content is highly variable in time and space which makes the corresponding delay not easily modeled. The time delay caused by water vapor, called wet delay, can be related to the integrated water vapor (IWV) content along the signal path; hence, we consider IWV and wet delay as two sides of the same coin.

InSAR has been used to generate high-resolution maps of the Earth's surface topography [Zebker and Goldstein, 1986]. One of the projects established for generating digital elevation models (DEM) of the Earth's surface is the Shuttle Radar Topography Mission (SRTM), which provides DEM maps up to latitudes of 60° at a spatial resolution of 30-90 m. The TanDEM-X (TerraSAR-X Add-on for Digital Elevation Measurements) is a radar mission started in 2010 with the primary goal of deriving global DEM maps at a resolution of 12 m and an absolute height accuracy of 10 m [Eineder et al., 2012]. InSAR has also been widely used to study geodynamic processes, where the main advantage of InSAR is the ability to provide maps of surface deformation at a high spatial resolution, particularly in regions where in situ measurements are impossible. This includes using InSAR observations, for example, to detect crustal deformation from single interferograms [Massonnet et al., 1993; Wright et al., 2003; Zebker et al., 1994] or by time series analysis [Ferretti et al., 2001; Hetland et al., 2012; Hooper et al., 2007], to measure volcanic inflation [Hooper et al., 2004; Lu, 2007] and to observe surface subsidence [Hoffmann et al., 2003; Meyer et al., 2007; Osmanoglu et al., 2011; Plattner et al., 2010].

One of the limiting error sources in the measurement of, for example, crustal deformation using repeat-pass InSAR is the effect of spatial and temporal variations of the propagation delay due to turbulent mixing of atmospheric water vapor [Hanssen, 2001; Williams et al., 1998; Zebker et al., 1997]. Several studies have been carried out to develop robust methods for mitigating the effect of neutrospheric distortions in InSAR interferograms either by stacking or calibration based on external data. Based on the fact that the neutrospheric wet delay is uncorrelated for time periods longer than one day [Hanssen, 2001; Williams et al., 1998; Zebker et al., 1997], the stacking approach mitigates the delay by averaging an adequate number of independent interferograms [Peltzer et al., 2001; Sandwell and Price, 1998; Schmidt et al., 2005; Wright et al., 2001]. The neutrospheric artifacts can also be mitigated by temporal filtering or smoothing in time series analysis, which have, for instance, been proposed by [Adam et al., 2005; Ferretti et al., 2001; González et al., 2010; Hooper et al., 2007; Kampes, 2005]. Reducing neutrospheric noise via calibration involves the elimination of a part or all of the delay using data from independent sources. Delacourt et al. [1998] used a model originally developed for correcting the neutrospheric delay in GPS (Global Positioning Systems) for correcting InSAR interferograms. Data from near-IR sensors such as MODerate resolution Imaging Spectrometer (MODIS) and MEdium Resolution Imaging Spectrometer (MERIS) have been successfully used to correct water vapor distortions in InSAR interferograms as presented in [Li et al., 2005] and [Li et al., 2006b], respectively. Also, data from GNSS networks have been used by [Doin et al., 2009; Li et al., 2006a; Onn and Zebker, 2006] to model and correct the elevation-dependent wet delay (vertically-stratified wet delay) in study areas with strong topography. Numerical atmospheric models provide maps of IWV and meteorological data, which have also been exploited for correcting the neutrospheric phase distortions [Foster et al., 2006; Gong et al., 2010; Wadge et al., 2002]. The correction of the wet delay using the introduced methods has different limitations. Calibration of the neutrospheric delay using data from MODIS and MERIS is limited under the presence of cloud coverage, while the sparse spatial distribution of the sites is the main limitation of the GNSS data. In addition, based on data from the GNSS, neutrospheric effects with a wavelength longer than the sites spacing and signals correlated with surface elevation can be reduced. The correction of short-scale (say, less than 15 km) neutrospheric distortions, however, requires highly-dense GNSS sites within the interferogram. The use of numerical atmospheric models is still limited by the coarse horizontal resolution, unpredictable weather and sensitivity of the model to the boundary conditions.

In the aforementioned studies, the atmospheric effects have been addressed as an error term that should be corrected. Since InSAR and GNSS are affected by the atmosphere in a similar way [Onn and Zebker, 2006], this "error" signal can be investigated as a valuable source of information to determine the atmospheric water vapor content. Since the 1990's, GNSS observations have been considered as an efficient tool for atmospheric sounding. Different studies were carried out to prove the potential of the neutrospheric parameters estimated from GNSS observations [Bevis et al., 1992; Rocken et al., 1995]. Since then, numerous studies have been carried out for estimating and mapping water vapor in the atmosphere using GNSS observations [Bai and Feng, 2003; Bender et al., 2008; Jade and Vijayan, 2008; Karabatić et al.,

2011; Luo et al., 2008]. Using InSAR observations as a meteorological tool is a relatively new research field. The focus of previous research was put on comparing the neutrospheric phase maps in InSAR with data from numerical weather models, GNSS and MERIS to anticipate the ability of these data to mitigate phase distortions in InSAR images. Pichelli et al. [2010] have compared IWV maps from MM5 model (Mesoscale Model, `http://www.mmm.ucar.edu/mm5`) with the corresponding maps from InSAR. The results presented by the authors show good agreement between the IWV difference maps from InSAR and MM5. Therefore, they suggested the assimilation of water vapor fields derived from InSAR with data from MM5 to improve the model resolution and the structure of atmospheric patterns. Mateus et al. [2010] compared IWV content estimated from GNSS observations with IWV difference maps extracted from InSAR interferograms. IWV estimates (temporal differences) from GNSS observations show similar values and similar spatial variations to those derived from InSAR. Also, the results presented by Meyer et al. [2008] show a good agreement between IWV maps derived based on persistent scatterer InSAR (PSI) and the corresponding maps from MERIS.

In this work, we use data from PSI, GNSS, MERIS and numerical weather prediction (NWP) models to quantify and analyze water vapor content and properties. These systems show complementary temporal or spatial properties. **Figure 1.1** shows IWV content in the atmosphere, at SAR overpass time on 09.05.2005 (format: month.day.year), produced by four systems: (a) GNSS, (b) PSI, (c) MERIS and (d) Weather Research and Forecasting Modeling System (WRF). The method of Precise Point Positioning (PPP) is applied to the GNSS measurements to produce time series of the total (absolute) atmospheric content of IWV at a temporal resolution of 60 minutes. These measurements are available at horizontally distant sites (**Figure 1.1** (a)), such that continuous 2D fields of IWV can be generated using inference techniques. However, short-scale spatial variations of water vapor cannot be observed based only on spatially-sparse GNSS measurements, even by using highly sophisticated interpolation techniques. PSI can be used to derive fields of IWV content at a high spatial density where atmospheric patterns within sub-kilometer range are detected. Since PSI observations are obtained with respect to a reference scene, they do not provide the total (absolute) content of water vapor, rather they measure temporal and spatial differences of IWV as shown **Figure 1.1** (b). Moreover, InSAR data are lost in areas of low coherence such as forests, vegetation regions and over water surfaces. MERIS is a near-IR sensor such that the observations of IWV under cloud cover are extremely underestimated and thus not representative for IWV content in the atmosphere. In (c) is shown a map of IWV measured by MERIS under clear weather. WRF is a NWP model that provides simulations of atmospheric parameters. The IWV maps generated by WRF have relatively coarse spatial resolution (here 3×3 km^2), so they do not model small-scale structures in the atmosphere. The model output data are limited by the selection of initial conditions and the model configurations for boundary layers. Therefore, IWV maps received from WRF may deviate from the actual state of the atmosphere. In **Figure 1.1** (d), we illustrate an example for an IWV map generated by the model, which shows a poor spatial correlation with the map measured by MERIS in some regions and good correlation in others.

Combination of GNSS and InSAR IWV observations taking the benefit of their complementary properties is considered as a promising strategy to build 2D fields of absolute IWV content. These maps can be investigated for (i) to better understand and model temporal and spatial variations of water vapor, (ii) to test the influence of local observations in improving the estimation of initial conditions for the WRF model and the ability to model short-scale variations due to turbulent mixing of water vapor and (iii) to better adjust the configurations of the boundary layer. If the assimilation of these maps into the model can improve the quality of the output data and the ability to model short-scale variations of water vapor, then the maps generated by the WRF model can successfully be used in a reverse way to correct for the neutrospheric delay in InSAR and GNSS measurements.

(a) GNSS   (b) PSI

(c) MERIS   (d) WRF
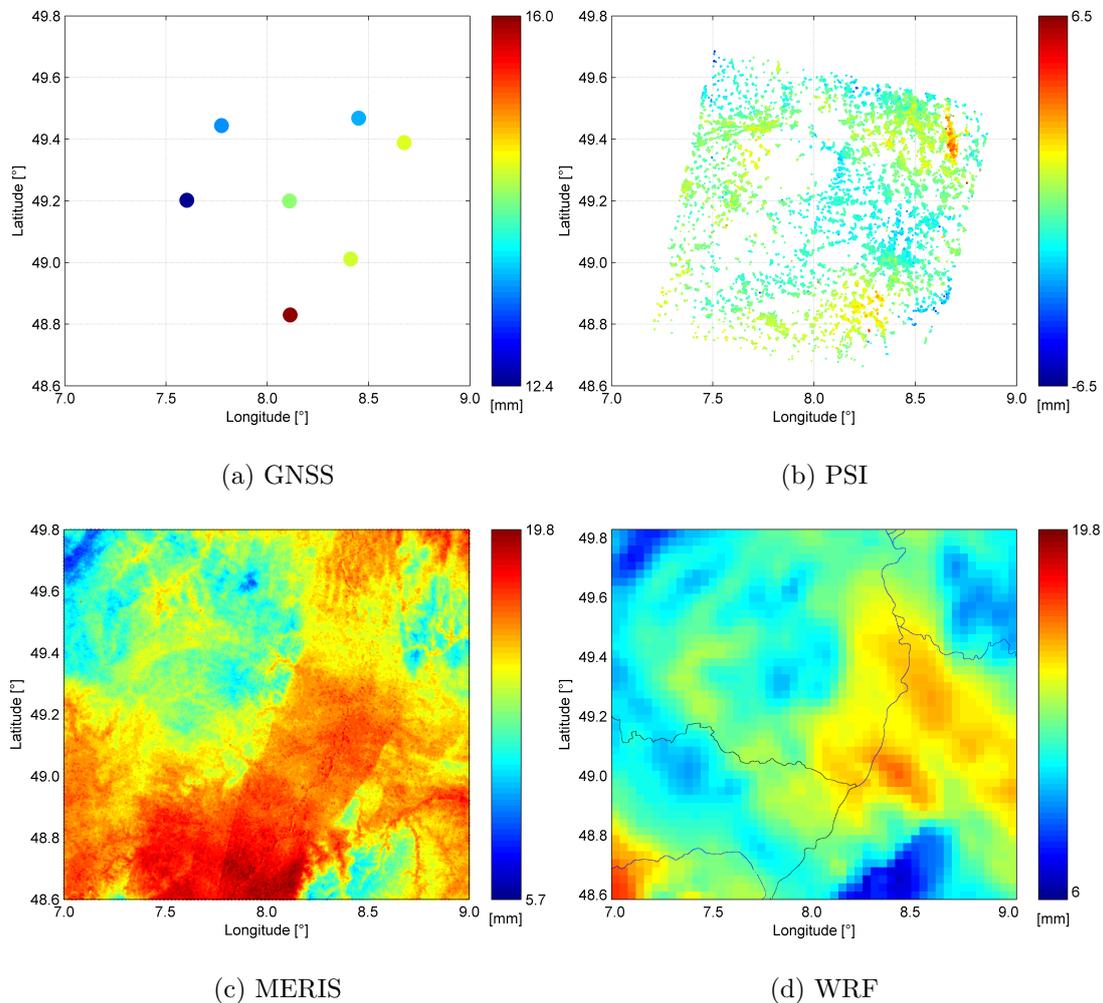
Figure 1.1: IWV derived from GNSS, PSI and MERIS observations and WRF simulations at the SAR overpass time on 09.05.2005. The resolution of MERIS and WRF data are, respectively 0.3×0.3 km^2 and 3×3 km^2. PSI provides point-level data with a density of 11 points/km^2. The minimum spacing between the GNSS sites approximates 20 km.

1.1 Objectives

In this work, the focus is put on the time delay caused by atmospheric water vapor in InSAR and GNSS observations as a data source for water vapor mapping. Persistent Scatterer InSAR (PSI) methods are used to derive wet delay-difference maps, where a large component is eliminated by interferometric data processing. The primary objective of the work is to achieve a proper combination of the wet delay derived from PSI and GNSS observations to reconstruct the eliminated wet delay signal and build maps of absolute wet delay, which can be converted into IWV content. Because wet delay is correlated in space, we use statistical interpolation methods to produce wet delay over regular grids. We then apply spatial statistical data fusion of the wet delay maps derived from the combined PSI-GNSS solution and IWV (converted into wet delay) maps from WRF.

As will be explained in Chapter 5, we extract maps of the neutrospheric phase from InSAR by processing a stack of coregistered images using the PSI technique. Separating the displacement phase component from InSAR interferograms may pose a challenge depending on the surface motion in the test area. In our test region, however, only very small long-term tectonic motions were observed, such that the corresponding phase in the interferogram can reasonably be assumed negligible. The neutrospheric phase maps at each SAR acquisition are obtained by least squares (LSQ) inversion. These maps contain only a component of the neutrospheric phase, while a large part of the phase vanished due to interferogram formation. At the GNSS sites distributed within the SAR image, the total neutrospheric delay is estimated during GNSS data processing. In the presence of meteorological observations of air pressure, temperature and relative humidity, the delay due to water vapor (wet delay) can be accurately determined. Based also on meteorological data, a value of the empirical constant used to convert the wet delay into IWV content is calculated. The main advantage of GNSS wet delay estimates over those from PSI is that they determine the total (absolute) wet delay, while PSI provides only the residuals of the wet delay. Based on GNSS wet delay observations we model the wet delay components eliminated in PSI processing. We combine the modeled component with the wet delay-partial maps derived from PSI to derive 2D fields of the absolute wet delay.

Several authors have suggested the mitigation of neutrospheric effects in InSAR by using GPS observations [Emardson et al., 2003; Li et al., 2006a; Williams et al., 1998]. Nevertheless, the authors have mentioned that the success of this approach to model high spatial variations of water vapor is limited by the spatial density of the GNSS sites. We use GNSS measurements to reconstruct the subtracted wet delay signal. This includes estimating the wet delay component that is correlated with topography and a signal that has a wavelength larger than the sites spacing (2D linear trend). This means that the GNSS sites do not have to be dense, but they should be well distributed within the SAR scene to properly cover the entire topography range.

IWV fields generated by combining GNSS and PSI observations produce absolute, accurate and spatially dense measurements, but the measurements may be absent in forest and vege-

tated regions. WRF on the other side, provides continuous less accurate fields of IWV at a coarser horizontal resolution. Therefore, the other objective of this work is to fuse IWV from the combined PSI-GNSS measurements and WRF simulations to infer IWV values at continuous grids of a predefined resolution and to measure the quality of the predictions. We take the benefit of the model to obtain values of the wet delay in areas where no remote sensing measurements are available.

1.2 Contributions

This thesis signifies the importance of GNSS and InSAR or PSI systems as successful meteorological tools for measuring IWV content in the atmosphere. Moreover, within this research we investigate spatial statistical data fusion of remote sensing observations of IWV from GNSS and PSI and data from the WRF model. In particular, (i) we use meteorological observations to derive time series of the wet delay from the total neutrospheric delay estimated during GNSS data processing. (ii) We develop a new method that uses wet delay observations from the GNSS sites distributed within the SAR image to reconstruct the delay components vanished when building interferograms. This includes modeling an elevation-dependent component and a linear trend. Since the GNSS sites are sparsely distributed within the SAR image, we first estimate an elevation-dependent wet delay component and then the residuals are used to model the long-wavelength trend. We combine the elevation-dependent wet delay, the long-wavelength wet delay and the wet delay-partial map extracted from the interferograms to build absolute maps of the wet delay. (iii) We apply spatial statistical data fusion to assimilate the wet delay maps derived in (ii) with water vapor maps from WRF first, to even out the deficits of each data set and second, to test the influence of local measurements on the quality of the maps generated by the model.

1.3 Outline

Chapter 2 presents a background overview of atmospheric physics. We focus on water vapor and the instrumentation used to measure its content in the atmosphere. We discuss the importance of understanding and modeling water vapor for environmental purposes and for remote sensing applications. We also present the models used to quantify the content of IWV based on meteorological parameters.

Chapter 3 presents the test site and the reasons behind its selection. We present also a brief description of the data sets investigated in this work and the spatial and temporal properties of each set.

Chapter 4 presents the interpolation and estimation techniques we used in this work. We describe the difference between the geostatistical and non-geostatistical techniques and the

reasons to use both of them. We discuss the limitations of the ordinary kriging and how they are overcome by using the fixed-rank kriging. We present an example of using ordinary kriging and fixed-rank kriging to interpolate scattered IWV data from single data sets.

In Chapter 5, we discuss the methods exploited to obtain estimations of the IWV from GNSS and InSAR observations. We present the method of precise point positioning as a processing strategy of GNSS observations. Then the Stanford Method for Persistent Scatterers (StaMPS) applied to retrieve wet delay maps from the interferogram is presented. We present a comparative analysis of our results with IWV maps from MERIS and WRF.

In Chapter 6, we present the method we developed to combine wet delay observations from PSI and GNSS to build absolute maps of the wet delay. We present comparisons of the results with IWV measured by MERIS.

In Chapter 7, we discuss the method of spatial statistical data fusion which is applied to fuse maps of absolute wet delay derived in Chapter 6 with IWV maps from WRF. We present the mathematical background of the algorithm and the possible difficulties. We discuss the results of the fusion and compare them with available data from MERIS.

Chapter 8 presents a summary of the results achieved in this work followed by suggestions for possible future investigations.

2 Physics of the Earth's atmosphere

In this chapter, we describe the effect of the Earth's atmosphere on microwave signals transmitted by space-borne transmitters such as GNSS and Synthetic Aperture Radar (SAR). The atmosphere is a heterogeneous medium extending to altitudes of 1000 km above the Earth's surface, with significant variations in its physical parameters such as temperature, pressure and humidity, particularly in the vertical direction. It is classified based on the electrical characteristics into two layers, the ionosphere and the neutrosphere. The ionosphere, which extends from altitudes of about 50 to 1000 km above the Earth's surface [Misra and Enge, 2001, p. 162], contains electrically charged particles. The neutrosphere is, however, electrically neutral and represents approximately the lowest 50 km of the atmosphere.

Electromagnetic signals transmitted from space-borne sensors encounter inevitable changes when they propagate through the atmosphere. The atmosphere changes the propagation speed of the microwave signals in a non-uniform way along the traveling path. The ratio of the signal speed in vacuum and the its speed in a medium defines a quantity called the *refractive index* of the medium [Misra and Enge, 2001, p. 169]. The medium is said to be dispersive, if its refractive index depends upon the frequency of the signal. For radio signals of frequencies below 30 GHz the ionosphere is a dispersive medium [Dach et al., 2007, pp. 37], while the neutrosphere is not [Davis, 1986, pp. 17]. In Section 2.1, we briefly describe the effect of the ionosphere on radio signals. We explain how the ionospheric delay is effectively reduced by a linear combination of multiple frequencies in the GNSS data processing, and its negligible effect in the InSAR data. We describe the neutrospheric layer and its constituents in Section 2.2. In Section 2.2.1, the neutrospheric water vapor and the existing methods to measure its content are introduced. We explain the refraction of radio signals due to dry gases and water vapor in Section 2.2.2. In 2.2.3, we discuss how the time delay is quantified based on empirical models. Then we describe the temporal and spatial properties of the time delay caused by water vapor and dry gases in Section 2.2.4. In Section 2.3, we describe the conversion factor required for translating the delay into water vapor content and vise versa. The mapping function techniques used for modeling the zenith-directed delay in terms of the line-of-sight delay are presented in Section 2.4.

2.1 The ionosphere

The ionosphere, the upper and largest layer of the atmosphere, contains free electrons and ions. The degree of ionization is variable with the solar activity that is characterized by the sunspot number. According to [Dach et al., 2007, pp. 255], the sunspot number shows a repeat cycle of about 11 years and an 80-100 years super cycle. **Figure 2.1** Depicts the sunspot numbers from the year 1999 to 2012. In the most recent full cycle, the maximum solar activity occurred in the years 2000-2002 and the minimum solar activity occurred in the years 2008-2010. Within the period of our study (2003-2009), the ionospheric activity is low.
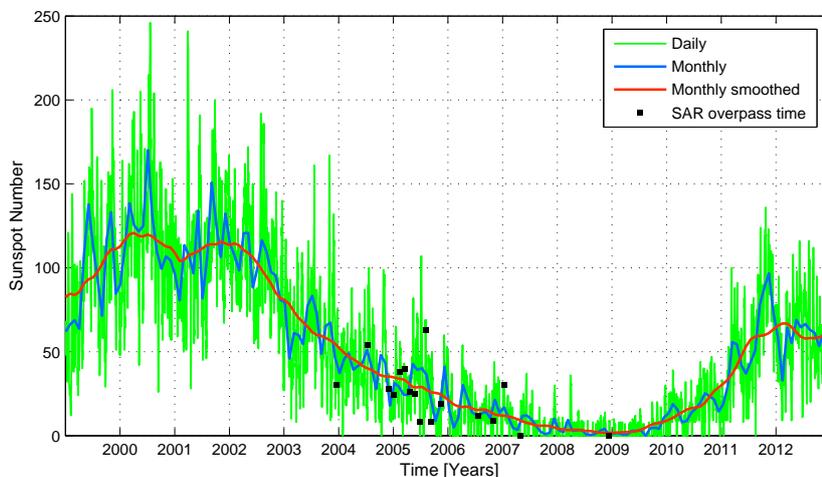
Figure 2.1: Daily, monthly, and monthly-smoothed sunspot numbers for the past 14 years as obtained from [`http://www.sidc.oma.be`]. The little black squares specify the sunspot numbers at the SAR overpass time.

The ionosphere refracts radio signals transmitted from space delaying the time required by these signals to reach the ground. This delay is determined by the number of free electrons and ions in the ionospheric layers. In order to quantify the ionospheric delay in microwave signals, we define the slant total electron content (TEC) over the signal path from the receiver (R) to the satellite (S):

$$TEC = \int_R^S n_e(s)ds \tag{2.1}$$

where TEC is measured by TEC units (TECU), and n_e is the electron density in units of electrons/m^3. The ionospheric refractive index (n_I) is a function of n_e and the carrier frequency (f), such that

$$n_I = 1 - \frac{40.3\,n_e}{f^2} \tag{2.2}$$

where the higher order terms in the equation above are neglected. The ionospheric delay

(ΔL_{iono}) is then given by [Hofmann-Wellenhof et al., 2008, pp. 120]:

$$\Delta L_{iono} = \int_s (n_I - 1)ds = -\frac{40.3 \, \text{TEC}}{f^2} \tag{2.3}$$

The ionosphere is a dispersive medium which means that the ionospheric delay depends on the carrier frequency, eq. (2.3). Therefore, applying a linear combination of two or more microwave signals transmitted at different frequencies succeeds to significantly reduce the effect of the ionosphere on the signal transit time. GPS, for example, which transmits at several frequencies in the L-band, employs linear combinations of L1, L2, and L5 phase observations to effectively reduce the ionospheric refraction in the observations [Hofmann-Wellenhof et al., 2008, pp. 126]. In repeat-pass InSAR with wavelength less than 6 cm (C- and X-band), the ionospheric effects are generally smaller than those observed in the L-band[Gray et al., 2000; Meyer et al., 2006]. Also, the sunspot numbers show low solar activity at the SAR overpass time as observed from **Figure 2.1**, which makes the ionospheric effect in the InSAR data negligible. For these reasons, we put the potential of this work on the neutrospheric propagation delay in the GNSS and InSAR observations.

2.2 The neutrosphere

The neutrosphere extends to heights of about 50 km above the Earth's surface. It is a mixture of dry gases, mainly nitrogen and oxygen, and water vapor. The refractive index of the neutrosphere is slightly larger than unity (the value in vacuum); therefore, the speed of propagation of GNSS and SAR signals in the neutrosphere differs from than that in vacuum. The excess in the traveled path is in the approximate range of 2.5-25 m depending on the satellite elevation angle [Misra and Enge, 2001, pp. 169]. The time delay of traveling signals is originating from both dry gases and water vapor. The delay caused by dry gases is called the *dry delay* and it shows smooth variations over time and space. The *wet delay* defines the excess in the signal path due to water vapor. The neutrospheric delay can also be divided into a hydrostatic component (contains the delay and a part of the wet delay) and a non-hydrostatic component. In the following, we present an overview of water vapor in the neutrosphere and the existing measurement techniques. Then, we describe the influence of dry gases and water vapor on the propagation time of radio signals.

2.2.1 Neutrospheric water vapor

Water can exist in the neutrosphere in one of three different phases: solid, liquid, or gas. Snow and ice crystals observed in clouds represent the solid phase. The liquid phase is observed in the rain falling from sky and clouds which are made up of tiny droplets of water. The largest

amount of water is suspended in the atmosphere in a gaseous form called water vapor. Water vapor is concentrated in the layer stretching from the Earth's surface to altitudes of about 12 km, measured from sea level, and most of its amount is found in the lowest 4 km. Water vapor is present even under clear sky conditions, but its concentration deviates from minimum over deserts to maximum over oceans. It is characterized by high variations in time and space.

Although water vapor constitutes a small fraction of the neutrosphere compared to the dry gases, it has a crucial impact on atmospheric processes. First, it is a major player in the hydrological cycle. The content and distribution of water vapor is highly interconnected with clouds distribution and rainfall [Bevis et al., 1992]. Second, water vapor is known to be the Earth's most abundant greenhouse gas, which plays an important role not only in weather but also in climate. Greenhouse gases absorb the thermal energy radiated from the Earth's surface and re-radiate it in all directions. A part of the re-radiated energy goes back to the surface and the lower atmosphere, which results in increasing the surface temperature. Water vapor has a large latent energy associated with phase change of water; this energy redistributes continuously in the atmosphere with the help of wind that transports water vapor horizontally and vertically. The transmission and absorption of this heat through the circulation in the atmosphere (e.g., condensation and evaporation) is considered as an important component of Earth's surface energy budget. It is observed that a 1% increase of the water vapor content increases the average surface temperature by more than 4°C [Physorg, 2006]. Without greenhouse gases, the temperature of the planet's surface would be below freezing [Seidel, 2002].

A variety of means have been developed to continuously monitor the vertical and horizontal distributions of water vapor in the atmosphere. These devices are used either from the ground such as radiosondes and ground-based water vapor radiometers or from space such as space-based water vapor radiometers and infra-red sensors. Radiosondes are balloon-borne instrument packages that provide measurements of temperature, humidity, and pressure at a fine vertical resolution. Radiosonde measurements are, however, inadequate to resolve horizontal variations of water vapor. Due to high costs the number of radiosonde launching is restricted to twice daily as mentioned in Bevis et al. [1992]. Accordingly, temporal variations within time scales of less than 12 hours cannot be resolved using radiosonde observations. Space-based downward-looking water vapor radiometers are passive remote sensing instruments that provide measures of water vapor content over wide spatial areas [Bevis et al., 1992]. They determine the distribution of water vapor by measuring the amount of thermal energy radiated from the Earth and absorbed by the moisture in the atmosphere. The recovery of IWV by these devices is complicated, since it requires information of the background temperature that is quite variable and difficult to determine over land. The use of space-based water vapor radiometers is more suitable over oceans and their performance is degraded in the presence of clouds [Bevis et al., 1992]. Space-borne Infra-Red sensors, such as MODIS and MERIS, have been used to provide measurements of IWV at a high horizontal resolution. MODIS provides IWV measurements at a resolution of 1 km, while MERIS IWV observations have a spatial resolution of 300 m (full resolution mode). Measurements of MODIS and MERIS can only

be provided during the day and they are degraded by the presence of clouds. In regions with cloud cover, the measured IWV value corresponds to water vapor integrated from the sensor to the could top. Hence, in cloudy regions, IWV measurements are significantly underestimated.

The importance of quantifying the content of water vapor in the neutrosphere is not only arising from being a key element in the atmospheric processes, but also as a source of distortion in high precision geodetic and imagery remote sensing applications of GNSS and InSAR. The neutrosphere is a non-dispersive medium, for frequency bands less than 30 GHz,such that the neutrospheric delay in general and the wet delay in particular cannot be reduced through a linear combination of dual frequencies, rather it has to be modeled and removed from the observations. One of the key tasks of the GNSS data processing software is to accurately estimate and correct the neutrospheric delay [Bevis et al., 1992]. Phase observations from, for example, InSAR exhibit temporal and spatial fluctuations most likely due to water vapor [Zebker et al., 1997]. The wet delay caused by water vapor represents less than 10% of the total neutrospheric delay [Misra and Enge, 2001, pp. 170], nevertheless it is a significant source of limitation in GNSS and InSAR geodetic and remote sensing applications. That is because water vapor content varies with weather and can change quickly in time and space, which makes the corresponding wet delay not easily modeled. Unlike the dry delay that can be accurately estimated from surface meteorological observations, it is not easy to predict the wet delay [Askne and Nordius, 1987; Elgered et al., 1991].

2.2.2 Wet and dry delays

The extent of the path delay experienced by a microwave signal due to the propagation through the neutrosphere depends on the refractive index of the air mass that varies along the signal path. The neutrospheric refractive index (n) is related to the neutrospheric refractivity (N) by the following formula:

$$n = 1 + 10^{-6}N \tag{2.4}$$

As the signal travels downwards in the neutrosphere, the increment in the signal phase can be related to the incremental path length as follows:

$$d\phi = \frac{2\pi n(z)}{\lambda}dz$$

where λ is the wavelength of the carrier signal, and the factor $\frac{2\pi}{\lambda}$ relates the path length in meters to the phase in radians. The phase shift is then obtained by integrating along the propagation path, i.e.,

$$\phi = \int_z \frac{2\pi n(z)}{\lambda}dz \tag{2.5}$$

From eqs. (2.4) and (2.5), a signal propagating vertically in the neutrosphere arrives the observed point, of local coordinates (x, y), with a phase shift given by:

$$\phi(x,y) = \frac{2\pi}{\lambda}\left(\int_0^L dz + 10^{-6}\int_0^L N(x,y,z)dz\right) \tag{2.6}$$

where L is the total path length through the atmosphere. The first integral term represents the geometrical path from the satellite to the point at the Earth's surface, while the second denotes the excess in the path due to the propagation in the neutrosphere. The neutrospheric zenith total delay (ZTD) is defined as

$$ZTD(x,y,t) = 10^{-6}\int_0^L N(x,y,z,t)dz \tag{2.7}$$

In the above equation the delay due to ray bending is neglected. Bevis et al. [1992] observed that the bending component is small compared to the total path excess for signals traveling at an elevation angle greater than about $15°$, while for signals traveling in the zenith direction the bending term most likely vanishes. The dry and wet constituents of the neutrosphere affect microwave signals differently; hence, it is convenient to classify the neutrospheric refractivity and delay into a dry term and a wet term. eq. (2.7) can then be written as a superposition of the two components,

$$ZTD(x,y,t) = 10^{-6}\left(\int_0^L N_{dry}(x,y,z,t)dz + \int_0^L N_{wet}(x,y,z,t)dz\right) \tag{2.8}$$

where N_{dry} is the dry refractivity (for dry gases) and N_{wet} is the wet refractivity (for water vapor). The widely used formula for the refractivity as a function of meteorological parameters is [Smith and Weintraub, 1953; Thayer, 1974]

$$N = k_1\frac{P_d}{T} + \left(k_2\frac{e}{T} + k_3\frac{e}{T^2}\right) + 1.4W \tag{2.9}$$

P_d: Pressure of dry air in units of hecto-Pascal [hPa]

e: Partial pressure of water vapor in [hPa]

T: Air temperature in Kelvin [K]

W: Liquid water content in [g/m^3]

k_1, k_3, k_3: Physical constants

The physical constants k_1, k_2, k_3 have been determined in different laboratory experiments, for example [Bevis et al., 1994; Smith and Weintraub, 1953; Thayer, 1974]. Bevis et al. [1994] adopted the values of 77.6 [KhPa^{-1}], 70.4 [KhPa^{-1}], and 3.739×10^5 [K^{2}hPa^{-1}] for k_1, k_2, and

k_3, respectively. In eq. (2.9), the first term refers to the N_{dry}, and the term in the parentheses denotes the N_{wet}. Hence, the zenith dry delay (ZDD) can be calculated from:

$$ZDD = 10^{-6} \int_0^L k_1 \frac{P_d}{T} dz \tag{2.10}$$

and the zenith wet delay (ZWD) is

$$ZWD = 10^{-6} \int_0^L \left(k_2 \frac{e}{T} + k_3 \frac{e}{T^2} \right) dz \tag{2.11}$$

Neglecting the effect of liquid water and summing up eqs. (2.10) and (2.11), the neutrospheric ZTD is:

$$ZTD = ZDD + ZWD \tag{2.12}$$

2.2.3 Empirical neutrospheric a priori models

There are different empirical neutrospheric a priori models that calculate the dry and wet delays in microwave signals based on surface meteorological data. The most popular models were presented by [Saastamoinen, 1973] and [Hopfield, 1969]. In the Saastamoinen model, the neutrospheric ZTD is modeled as a function of the air pressure and the partial pressure of water vapor:

$$ZTD^{model} = 0.002277 D \left[P + \left(\frac{1255}{T} + 0.05 \right) e \right] \tag{2.13}$$

where the factor D depends on the altitude (z) and the geographical latitude (ϕ) at which the neutrospheric delay is computed,

$$D = 1 + 0.0026 \cos(2\phi) + 0.00028 \, z \tag{2.14}$$

The zenith dry delay (ZDD^{model}) and the wet delay (ZWD^{model}), according to Saastamoinen, are given by:

$$ZDD^{model} = 0.002277 \, D(P - 0.155471 \, e) \tag{2.15}$$

$$ZWD^{model} = 0.002277 \, D \left(\frac{1255}{T} + 0.205471 \right) e \tag{2.16}$$

While the dry component is accurately modeled, the model of the wet component is poor [Misra and Enge, 2001, pp. 172]. In order to use the Saastamoinen model for accurately quantifying the dry delay, for example, at a GNSS site, observations of the air pressure, temperature, and the partial pressure of water vapor on the Earth's surface are required. Since e in eq. (2.15)

has a value significantly smaller than P, the dry delay can be calculated based only on the air pressure neglecting the partial pressure of water vapor. A value for e can be computed based on observations of the air temperature and relative humidity using Using eq. (2.20), if direct observations of e are not available. The meteorological data have to be interpolated or extrapolated at the GNSS site, if the site is not equipped with meteorological sensors. In the absence of meteorological observations, calculations can be made by substituting representative values from the standard atmosphere. The meteorological parameters at a GNSS site of an altitude (z_s), in [km], can be extrapolated from the parameters at the mean sea level ($z_0 = 0$ [km]) as follows:

$$T = T_0 - 6.5(z_s - z_0) \qquad\qquad [\text{K}] \qquad (2.17)$$

$$P = P_0(1 - 0.0226(z_s - z_0))^{5.225} \qquad\qquad [\text{hPa}] \qquad (2.18)$$

$$rh = \frac{rh_0}{100} \cdot \exp(-0.6396(z_s - z_0)) \qquad\qquad [\%] \qquad (2.19)$$

$$e = \frac{rh}{100} \cdot \exp\left(-37.2465 + 0.2131665T - 0.000256908T^2\right) \qquad\qquad [\text{hPa}] \qquad (2.20)$$

The standard values of the air temperature (T_0), the air pressure (P_0), and the relative humidity (rh_0) at the mean sea level are:

$$T_0 = 291.15 \ [\text{K}]$$
$$P_0 = 1013.2 \ [\text{hPa}] \qquad (2.21)$$
$$rh_0 = 50\%$$

The eqs. (2.17) to (2.20) and eq. (2.21) are reported in [Dach et al., 2007, p. 243].

2.2.4 Neutrospheric delay in time and space

Both dry and wet delay vary temporally and spatially (horizontally and vertically); however, the wet delay is more variable. Due to the variations in the atmospheric layer thickness and the refractivity profiles of the neutrosphere in the vertical direction, the neutrospheric effects (dry and wet) are proportional to the surface altitude. In **Figure 2.2**, the expected neutrospheric delay at point B, which exists at a higher altitude, is less than that at point A due to variations in pressure, temperature and the thickness of the neutrospheric layer above each point. The delay varying in the vertical direction is called *vertically-stratified* delay.

Variations of dry delay are linked to the variations in the pressure of dry gases, which have smooth spatial variations with long wavelength and so does the dry delay. In time, the pressure of dry gases shows small variations, especially if the variations of the air temperature are small. This makes the dry delay quite repeatable [Zebker et al., 1997]. Vertical profiles of dry delay observations show a clear reduction in the dry delay by moving upwards away from the Earth's surface.
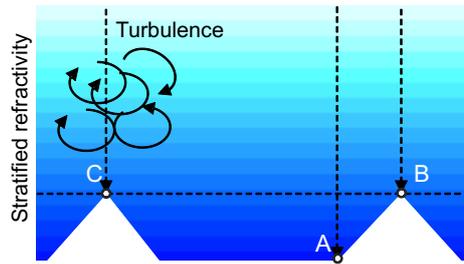
Figure 2.2: Stratified refractivity and turbulent mixing in the neutrosphere. If microwave signals are received at points A, B, the neutrospheric delay at point B is lower. Turbulent mixing above point C makes the delay larger compared to that at point B, despite being at the same altitude.

Short-scale spatial variations of the wet delay results from turbulent mixing of water vapor in the neutrosphere, called *turbulently-mixed* wet delay. Atmospheric turbulence is associated with short-scale, irregular air motions at different speeds and in different directions. This occurs because, when solar radiation heats the Earth's surface, the air above it becomes warmer and more buoyant, and cooler, denser air descends to displace it. The resulting vertical movement of air together with flow disturbances around surface obstacles, makes low-level winds extremely irregular. Atmospheric turbulence up to heights of approximately 2 km causes transport and mixing of water vapor, changing its distribution vertically and horizontally. Due to the strong variations of water vapor distributions and the strong consistency of the dry delay, water vapor is considered as an approximate tracer of the variations in the neutrospheric refractivity. It is also the main source of the temporal and spatial fluctuations of the neutrospheric wet delay observed in GNSS and PSI measurements. Temporally, the wet delay is uncorrelated for time periods longer than one day [Hanssen, 2001; Williams et al., 1998; Zebker et al., 1997]. In **Figure 2.2**, the neutrospheric delay at point C is greater than the delay at point B, although they exist at the same altitude. This is due to the turbulent mixing in the surrounding of point C that raises the value of the wet delay.

Spatial and temporal variations of the wet delay can be described by statistical functions such as the power spectral density or structure functions [Hanssen, 2001; Onn, 2006; Treuhaft and Lanyi, 1987]. Kolmogorov elementary theory of turbulence [Tatarskii, 1961] describes the expected theoretical parameters of these statistical functions and defines the theoretical spatial patterns of water vapor patches (see Section 5.2.6). Water vapor has been considered as an approximate tracer of the atmospheric turbulence [Ishimaru, 1978]; therefore, temporal and spatial fluctuations of water vapor (or wet delay) should also follow the theory of Kolmogorov.

On the basis of the temporal and spatial properties of the dry and wet delays in microwave signals, the neutrospheric delay measured at any point on the Earth's surface can be classified into a homogeneous component and a heterogeneous component. The homogeneous component is a dry delay of a value proportional to the thickness of the neutrospheric layer above the observed point. The heterogeneous component is a sum of dry and wet delays varying with

the weather conditions and the path of the transmitted signal. This classification will be required to detect which components of the water vapor are eliminated by building InSAR interferograms.

In **Figure 2.3**, we show IWV maps measured by MERIS at two different times as well as the corresponding map of surface topography. We observe that the map in (a) is dominated by the vertically-stratified IWV, such that we can observe high correlation between the IWV map and the map of surface topography. In (b), however, we observe the large content of IWV distributing from the south west due to the turbulent mixing in the neutrosphere. Evidence for the high neutrospheric activity during the acquisition time is the presence of clouds, which is visible as white areas in the southern part of the map in **Figure 2.3** (b).

(a) IWV on 07.17.2006

(b) IWV on 04.23.2007

(c) Surface topography
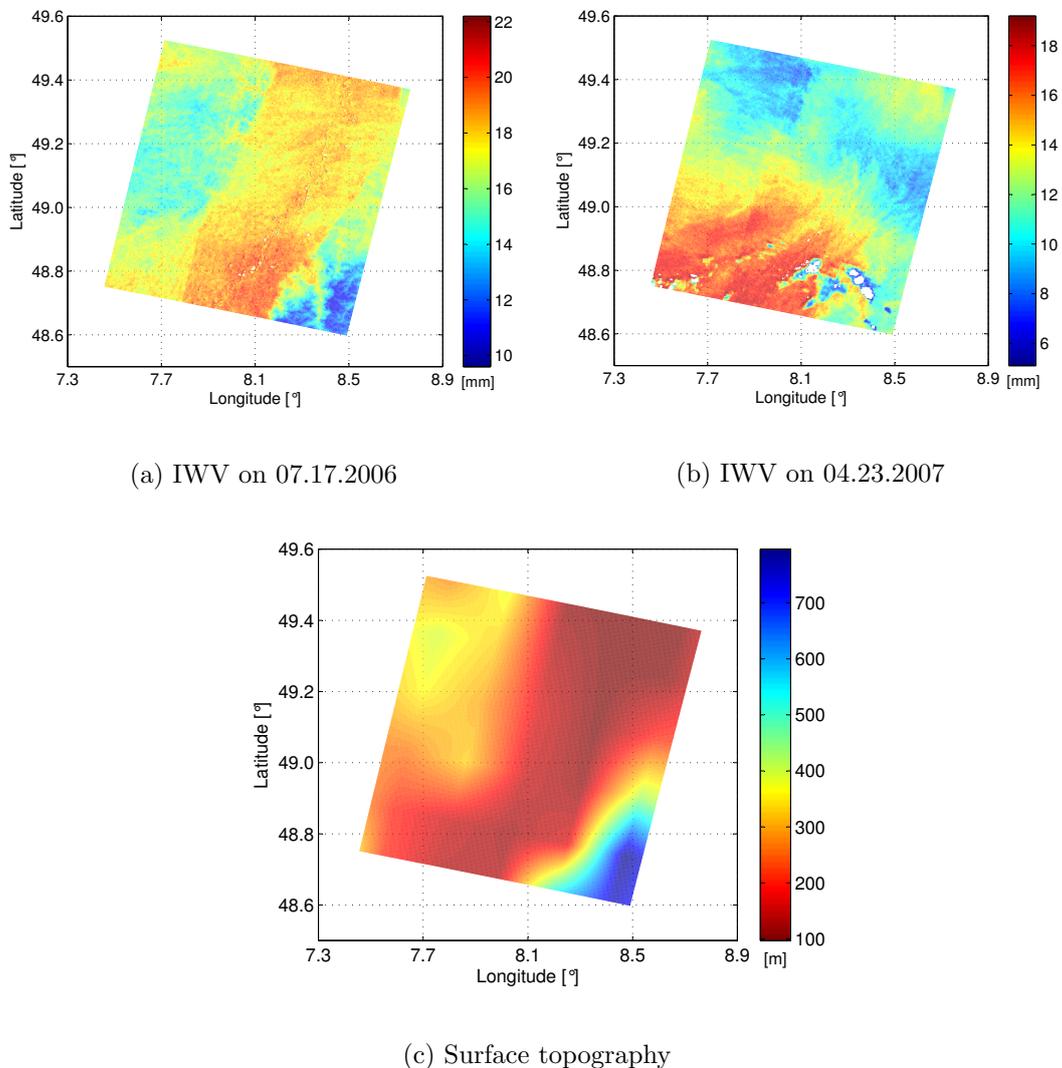
Figure 2.3: IWV maps measured by MERIS and the corresponding surface topography. The maps show the IWV dominated by the stratified water vapor (a) and turbulently-mixed water vapor (b). Empty areas in the map in (b) indicate the presence of clouds.

2.3 Relation between wet delay and water vapor

The relations between vertically integrated water vapor at any point on the surface and the ZWD at that point has been studied by several researchers, such as [Askne and Nordius, 1987; Bevis et al., 1994; Emardson and Derks, 2000; Webley et al., 2002]. Based on their findings, the path delay can be converted to IWV content by the following relation:

$$IWV = \Pi \cdot ZWD \qquad \text{and} \qquad ZWD = Q \cdot IWV \qquad (2.22)$$

Please note that the above relation is given in Bevis et al. [1992] to calculate the total precipitable water. However, taking into account the density of liquid water (approximately 1 g/cm^3), the IWV expressed in kg/m^2 is equivalent to the total precipitable water expressed in mm. $Q = \frac{1}{\Pi}$ and Π is an empirical constant calculated from a formula presented by Askne and Nordius [1987] and reported in Bevis et al. [1994], i.e.,

$$\Pi = \frac{10^6}{\rho_w R_w \left(\dfrac{k_3}{T_m} + k_2 - k_1 \dfrac{M_w}{M_d} \right)} \qquad (2.23)$$

T_m: Weighted mean temperature of the atmosphere in [K]

ρ_w: Density of water

R_w: Specific gas constant of water vapor [8.134 J/mol·K]

M_w: Molar mass of water vapor in [18.0152 g/mol]

M_d: Molar mass of dry air in [28.9644 g/mol]

The values for R_w, M_w and M_d are given in Askne and Nordius [1987]. The factor Q has been derived empirically, and is reported in [Schüler, 2001, pp. 186]:

$$Q \approx 0.10200 + \frac{1708.08}{T_m} \qquad (2.24)$$

Several empirical relations were developed to determine the mean temperature based on surface temperature [Bevis et al., 1994, 1992; Davis et al., 1985; Schüler, 2001]. We compute T_m from the expression presented by Bevis et al. [1992]:

$$T_m \approx 70.2 + 0.72 T_s \qquad (2.25)$$

T_s is the surface temperature in [K]. According to Bevis et al. [1992], this linear regression allows the estimation of the mean temperature with an RMS relative error of less than 2%. As

a rule of thumb, $\Pi = 0.15$, suggested by Bevis et al. [1994], is commonly used in atmospheric research, but the actual value is varying with the surface temperature. In [Leick, 2004, pp. 201], the factor Q has a value in the range of 5.9-6.5. Webley et al. [2002] carried out a study over Europe to compute the value of Π from different models. The analysis was done for two time intervals: the first interval includes ten days in August/September 2000, where Q changed from 6.15 to 6.5. The second interval includes ten days in October 2000 and the range of Q was 6.25-6.5.

We consider the study interval examined by [Webley et al., 2002] to be a short time window and it neither examines low temperature in winter nor high temperature in Summer. In order to test the sensitivity of Q to the surface temperature, we used the model of eq. (2.24) to compute its value over 96 days in the years 2002 and 2004. The values of Q calculated using observations of the surface temperature from three meteorology stations are depicted in **Figure 2.4**. We observed that Q is in the range 6.039-6.633 using the observations of year 2002, while it varies from 6.047 to 6.570 in the year 2004. The value of Q is linearly decreasing with a slope of -0.0236/[K] and -0.0232/[K] for the years 2002 and 2004, respectively. It is evident that the value of Q (and hence Π) depends significantly on T_s, which should have accurate values to obtain accurate values for Q.

(a) Year 2002 (b) Year 2004
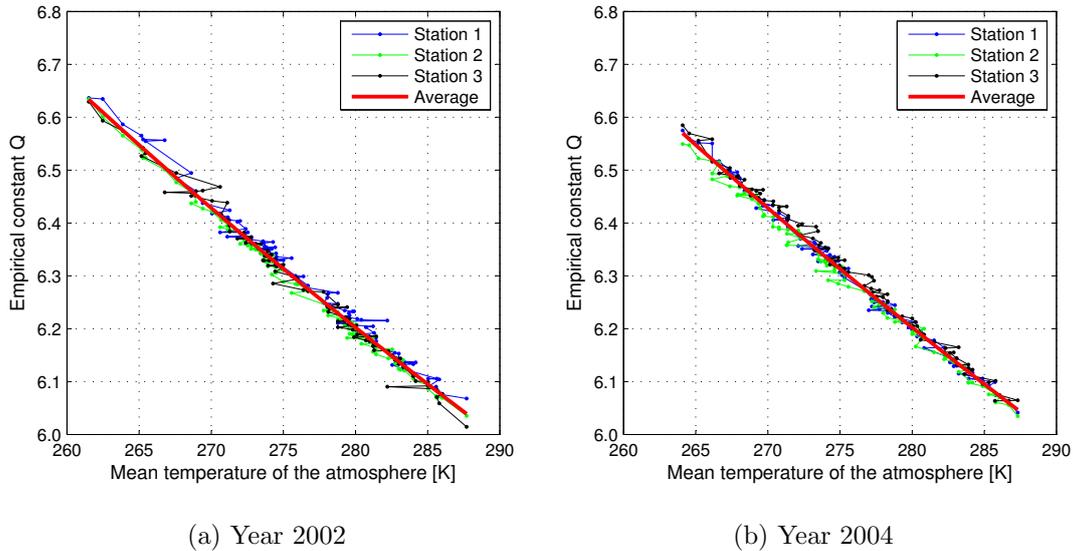
Figure 2.4: The empirical constant $Q = \frac{1}{\Pi}$ calculated based on the surface temperature measured by three meteorology stations and their average. The measurements taken at 10:00 am over 96 days (8 days/month) in the years 2002 and 2004 are used in the calculations. The slope of the red line in (a) is -0.0236/[K], and -0.0232/[K] for the red line in (b)

We found that inaccurate values of Q can cause a significant bias in the ZWD or IWV values after conversion if the amount of water vapor in the atmosphere is large. Hence, we calculated the value of Q from the linear regression in eq. (2.24) on the basis of surface temperature measurements at SAR acquisition times. The result is depicted in **Figure 2.5**, where one value is calculated at each SAR acquisition time. As observed from the figure, the value of

Q determined from the temperature measurements is less than $\frac{1}{0.15}$ and there is a significant change with the surface temperature. We concluded that the value of $\Pi = 0.15$ is not suitable for our test site, in the Upper Rhine Graben (URG).
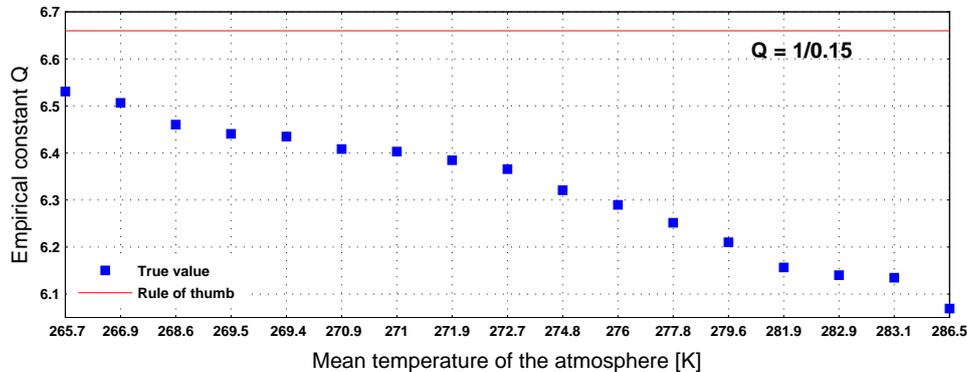
Figure 2.5: Empirical constant Q calculated based on measurements of the surface temperature. Each blue square refers to the value of Q used at each SAR acquisition time. The red line represents the value of $Q = \frac{1}{0.15}$.

2.4 Mapping functions

The dry or the wet neutrospheric delay along the satellite LOS is modeled by the product of the zenith delay and a geometric factor, referred to as the *mapping function* [Niell, 1996]. The mapping function (MF) describes the dependence of the delay on the elevation angle assuming azimuthal symmetry [Niell, 1996]. Scientists have proposed a number of MFs starting from simple models to sophisticated ones. The simplest model of an MF is

$$\text{MF}(\theta_{el}) = \frac{1}{\sin \theta_{el}} \tag{2.26}$$

where θ_{el} is the satellite elevation angle (see **Figure 2.6**). This model is reasonable if the satellite is transmitting signals at high elevation angles, while it is poor for low-elevation satellites. A more sophisticated model was proposed by Marini [1972] by expanding in a continued fraction in terms of $\sin \theta_{el}$

$$\text{MF}(\theta_{el}) = \cfrac{1}{\sin \theta_{el} + \cfrac{a}{\sin \theta_{el} + \cfrac{b}{\sin \theta_{el} + \cfrac{c}{\dots}}}} \tag{2.27}$$

where a, b, and c are constant coefficients. The form in eq. (2.27) was the basis for the most further MFs [Boehm and Schuh, 2004; Herring, 1992; Niell, 1996], except for [Lanyi, 1984]. Moreover, forms were developed for mapping the wet and dry delays independently. In this

20

work we use the model suggested by Niell [1996] for mapping wet and dry delays at elevation angles down to $3°$, which has the following form:

$$\mathrm{MF}_{Niell,i}(\theta_{el}) = \frac{1 + \dfrac{a_i}{1 + \dfrac{b_i}{1 + c_i}}}{\sin\theta_{el} + \dfrac{a_i}{\sin\theta_{el} + \dfrac{b_i}{\sin\theta_{el} + c_i}}} \tag{2.28}$$

where the subscript i in the equation defines either a wet or a dry MF. The coefficients a_i, b_i, and c_i are empirically determined and depend on the latitude and altitude of the observing site for the dry MF, while the wet MF requires only the site latitude [Niell, 1996]. Using a proper MF, the SWD can be obtained from the IWV as follows:

$$SWD = \mathrm{MF}_{wet} \cdot Q \cdot IWV \tag{2.29}$$

where MF_{wet} is the wet MF. We remind the reader that eq. (2.29) does not consider the azimuthal anisotropy of the neutrosphere. Therefore, we describe in Chapter 5 a method to model the total LOS delay by using neutrospheric gradients and phase residuals.

The need for sophisticated MFs grows if the satellite flies at low elevation angles. For Envisat SAR system, the acquisitions are made with a side looking satellite at incidence angles (off-nadir angles) of about $23°$. Thus, for mapping the delay estimated over the radar LOS to the zenith direction, the following MF is suitable:

$$\mathrm{MF}(\theta_{el}) = \frac{1}{\cos\theta_{inc}} \tag{2.30}$$

with θ_{inc} the incidence angle of the SAR.
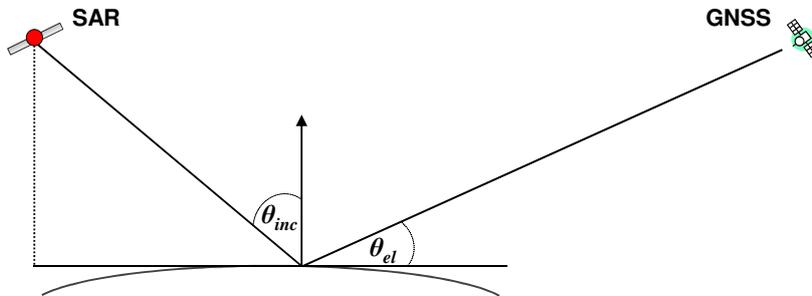
Figure 2.6: Elevation and incidence angles for a GNSS satellite and a SAR, respectively.

3 Data sets and study region

In this chapter, we present the study region and the data sets used in this work either for estimating water vapor or for comparing and evaluating our estimates. We first describe the study region and the reasons behind its choice. SAR data and their location are presented in Section 3.2. In Section 3.3, we describe the GNSS database and present the neutrospheric ZTD time series, which are used for computing the wet delay. In Section 3.4, we describe the importance of the meteorological data for precisely extracting the wet delay from the total neutrospheric delay. We describe the extraction of water vapor information from MERIS data in Section 3.5, which are used later for evaluating our estimates of water vapor. Meteorological data, including water vapor maps, are received from the model of WRF as presented in Section 3.6. The temporal and spatial properties of all data sets are then summarized at the end of the chapter.

3.1 Study region: Upper Rhine Graben

We selected the area of Upper Rhine Graben (URG) as a test region for this study for two reasons: the first reason is, the high availability of data, i.e., the region is well-covered by GNSS sites and it is increasingly acquired by synthetic aperture radar. Second, it has been shown in earlier studies that the tectonic deformation rate in the area of URG does not exceed 0.5 mm/year [Fuhrmann et al., 2012]. Hence, we can consider surface deformation to have a minimal contribution in the interferometric phase. This is important for the InSAR processing, where it is often challenging to separate the atmospheric phase from the crustal deformation phase. **Figure 3.1** shows the study region, the locations of the GNSS sites, the SAR image location, and the location of meteorological stations.

3.2 InSAR data

SAR data acquired by the ESA (European Space Agency) Envisat satellite are collected over the region indicated by the gray box in **Figure 3.1** The captured images contain urban areas and natural terrain with smooth topography. We built a stack of 17 co-registered advanced SAR (ASAR) scenes covering an area of 100×100 km^2 that is centered on 49 09' 38"N, 8 04' 45"E. The images are acquired in the time period of 2003 to 2008 during descending passes at
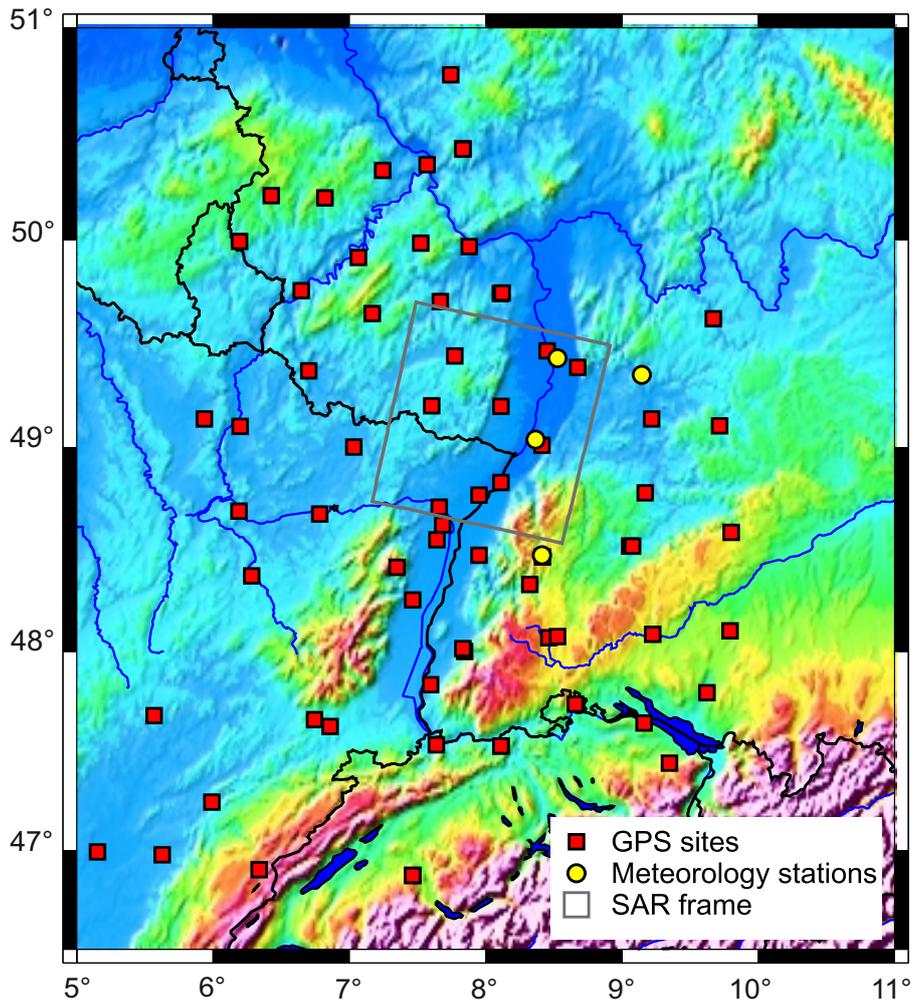
Figure 3.1: Study region in the Upper Rhine Plain (Graben). The little red squares indicate the locations of the GNSS sites in Germany, France, and Switzerland, while the yellow circles indicate the locations of the meteorology stations. The frame of the SAR image is shown by the gray box.

time interval of multiples of 35 days. Detailed information about the ASAR images is given in **Table 3.1**. Interferometric SAR (InSAR) processing is applied with respect to a single reference image (called the master) acquired on 06.27.2005 at 09:51 (UTC). The master scene has been selected such that the neutrospheric effects in the image are minimum. The temporal and spatial distribution of the images is shown in **Figure 3.2**. A map for surface topography is required in the InSAR data processing; therefore, we use a laser digital elevation model (DEM) of a spatial resolution of 10×10 m^2. Persistent scatterer InSAR (PSI) processing is applied to the built interferograms in order to identify stable points that are used to estimate the neutrospheric phase maps. The processing of InSAR data for atmospheric mapping will be described in Chapter 6.

ID	Acquisition day	Acquisition time (UTC)	Orbit	Frame	Processing
1	12.15.2003	09:51:28.09	9366	2614	SLC
2	07.12.2004	09:51:32.83	12372	2614	SLC
3	11.29.2004	09:51:30.57	14376	2614	SLC
4	01.03.2005	09:51:24.50	14877	2614	SLC
5	02.07.2005	09:51:27.67	15378	2614	SLC
6	03.14.2005	09:51:26.18	15879	2614	SLC
7	04.18.2005	09:51:32.54	16380	2614	SLC
8	05.23.2005	09:51:33.44	16881	2614	SLC
9	06.27.2005	09:51:35.22	17382	2614	SLC
10	08.01.2005	09:51:32.54	17883	2614	SLC
11	09.05.2005	09:51:29.05	18384	2614	SLC
12	11.14.2005	09:51:31.06	19386	2614	SLC
13	07.17.2006	09:51:31.20	22893	2614	SLC
14	10.30.2006	09:51:29.90	24396	2614	SLC
15	01.08.2007	09:51:24.39	25398	2614	SLC
16	04.23.2007	09:51:23.21	26901	2614	SLC
17	12.08.2008	09:51:18.55	35418	2614	SLC

Table 3.1: Available SAR scenes.
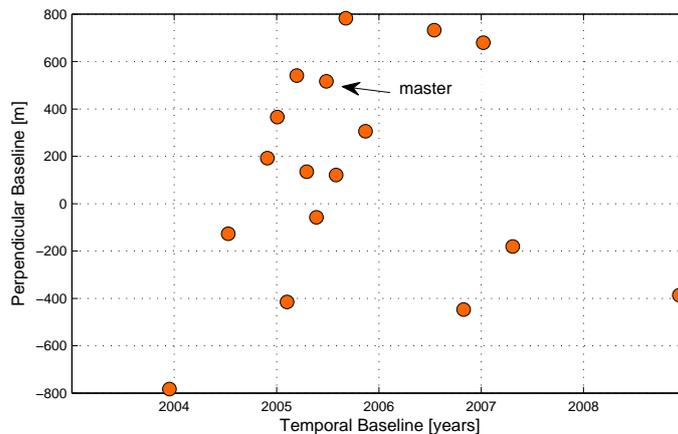
Figure 3.2: Temporal and perpendicular baselines corresponding to each ASAR scene.

3.3 GNSS data

The GNSS data are received from a set of permanent GNSS sites located in the Upper Rhine region, as illustarted in **Figure 3.4**. These sites belong to the GNSS Upper Rhine Graben Network (GURN). GURN is a cooperative project between the Geodetic Institute Karlsruhe (GIK, Karlsruhe, Germany) and the Institut de Physique du Globe de Strasbourg (Ecole et Observatoire des Sciences de la Terre, Strasbourg, France). The network contains Global Positioning System (GPS) and GLONASS sites [Mayer et al., 2012]; However, only GPS data are used in this study. Most of the German sites of GURN are serviced by the Satellite Positioning

Service of the German State Survey, SA*POS*®, in Baden-Württemberg and Rheinland-Pfalz. GURN contains beside the German sites, French and Swiss sites. Phase observations are available at the GPS sites at a temporal resolution of 30 seconds. The sites were installed at different times, for example, some sites have data profiles since 2002, but others have profiles since 2007. The distribution of GURN sites and the corresponding altitudes are shown in **Figure 3.4**. Names are added to the sites used for the combination with InSAR wet delay maps. Some GNSS sites within the image frame are not named since they have short profiles and are not available for the entire SAR observation period. The exact geographical coordinates and altitudes of these sites are given in **Table 3.2**. As will be presented in Chapter 5, the ZTD observations are used to derive values for the wet delay or the IWV.

The GNSS phase and code observations were processed at the GIK using the Bernese GNSS Software Version 5.0 (BS5) to extract time series of absolute neutrospheric ZTD at the GNSS sites. In **Figure 3.3**, we show the 24-hours time series of the ZTD determined at the available GNSS sites for 05.23.2005 (late sparing, warm weather) and 02.07.2005 (winter, cold weather) at 09:51 UTC (SAR overpass times). The warm weather results in evaporating a huge amount of water that increases the water vapor content in the atmosphere. Therefore, the ZTD values in (a) are larger than those values in (b) due to the higher content of water vapor in the neutrosphere and hence a larger wet delay on 05.23.2005. The temporal variations observed in the time series are associated mainly with the variations in water vapor content. Since URG region has dry weather conditions in winter (low water vapor content), the ZTD values are almost constant in plot (b). The ZTD measurements are used to retrieve the wet delay observed at each GNSS site as presented in Chapter 5.

Site name	City	Longitude [°]	Latitude [°]	Altitude [m]
Freud	Freudenstadt	8.4158	48.4645	784.4000
Heid	Heidelberg	8.6753	49.3889	168.8000
Heil	Heilbronn	9.2183	49.1385	234.8000
Iffe	Iffezheim	8.1126	48.8301	185.4000
Kais	Kaiserslautern	7.7740	49.4441	307.4000
Karl	Karlsruhe	8.4113	49.0112	182.9000
Land	Landau	8.1094	49.1998	208.0000
Ludw	Ludwigshafen	8.4506	49.4687	158.3000
Offe	Offenburg	7.9510	48.4730	233.5000
Pirm	Pirmasens	7.6025	49.2021	448.4000

Table 3.2: Names, locations, and ellipsoidal altitudes of the GNSS sites located within and close to the SAR scene.
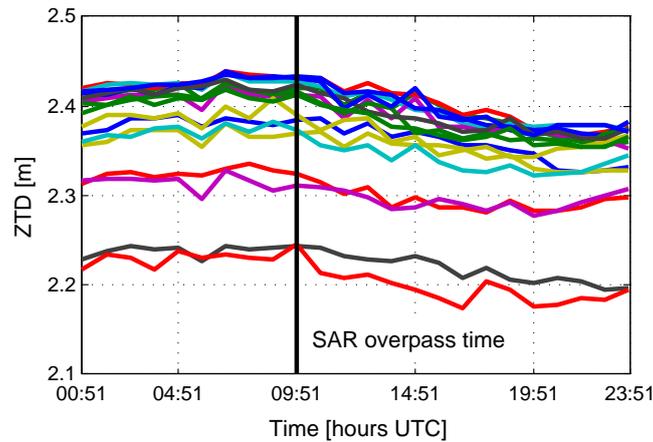
(a) 05.23.2005
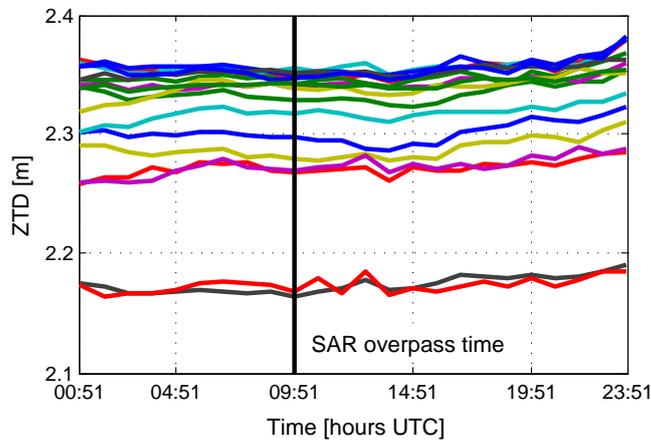
(b) 02.07.2005

Figure 3.3: Time series of the ZTD estimated from GNSS observations at 17 receivers on 05.23.2005 (a) and 02.07.2005 (b). In general, the higher GNSS site is, the lower the ZTD value. The vertical black line in each plot indicates the acquisition time of SAR.

3.4 Meteorological measurements

Accurate retrieval of the wet delay requires measurements of the air temperature, pressure, and relative humidity. Hence, Meteorological data received from the Landesanstalt für Umwelt, Messungen und Naturschutz Baden-Württemberg (LUBW) and from WRF are used to support the calculation the wet delay. The spatial locations and altitudes of the meteorological stations are shown in **Figure 3.4**. The LUBW meteorological stations provides measurement of the air pressure, temperature and dew point temperature. The measurements of the meteorological stations are available at a temporal resolution of 30-60 minutes over different time periods as specified in **Table 3.3**. Measurements of the air temperature and dew point temperature are used to compute the relative humidity and the partial pressure of water vapor

as described in Section 5.1.3. Since the meteorological stations are not located directly at the location of the GNSS site, spatial interpolations or extrapolations are essential (see, Section 5.1.3). LUBW stations provide point measurements at 4 meteorological stations, while the model WRF produces grids of pressure and temperature. The WRF simulations of surface temperature are used to calculate the constant Π presented in Section 2.3.
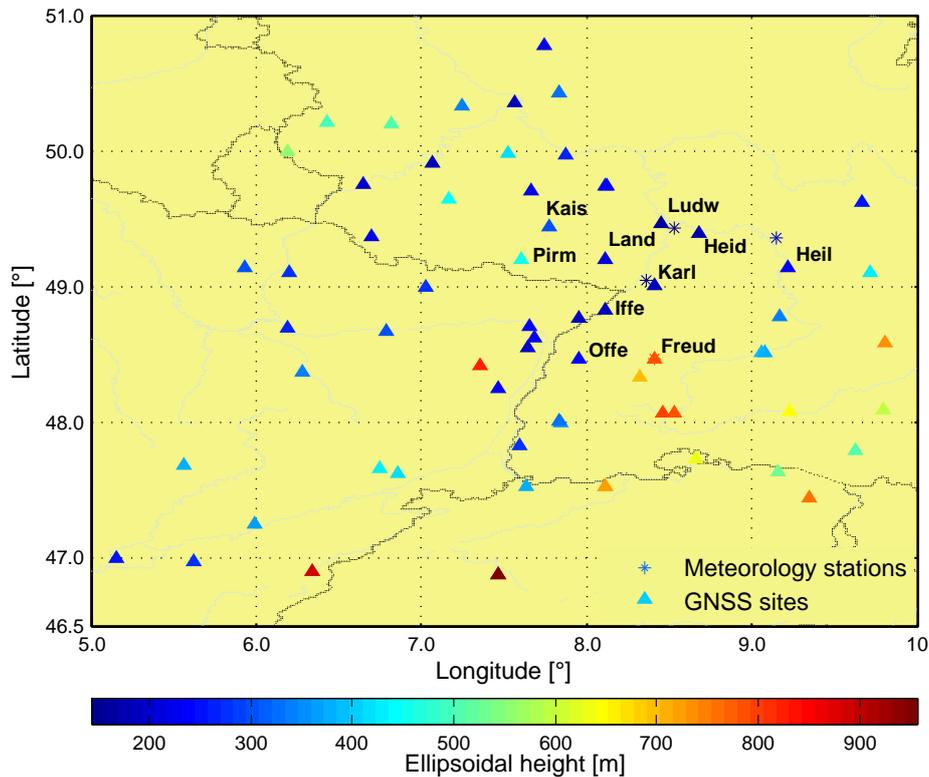
Figure 3.4: Spatial distribution and ellipsoidal altitudes of the GNSS sites and meteorological stations. Names are added to the sites with data profiles that extends over the period covered by the SAR.

3.5 MERIS data

An independent source for atmospheric water vapor measurements is required to evaluate and compare water vapor content obtained from GNSS and InSAR data. Since water vapor is highly variable, it is necessary to have IWV maps simultaneous to our estimates. The MEdium resolution Imaging Spectrometer (MERIS) is located together with the advanced SAR (ASAR) sensor on board of the Envisat platform, so that the two sensors can acquire data simultaneously during daytime. Thus for the collected ASAR scenes, simultaneous water vapor maps are available from MERIS. MERIS is a passive imaging spectrometer and it performs spatial and spectral imaging of the Earth, by looking in the nadir direction [ESA, 2006, pp. 16]. Beside the primary mission to measure ocean color, MERIS made also a significant contribution

to atmospheric and land surface related studies [ESA, 2006, pp. 26]. The swath width covered by MERIS is around 1150 km with a spatial resolution of 260×290 m^2 at full mode. MERIS allows for the global retrieval of total atmospheric water vapor content of the Earth every 3 days.

MERIS measures the solar radiation reflected from the Earth's surface and clouds. Fischer and Bennartz [1997] proposed a general algorithm to relate the integrated water vapor content to the ratio of the radiance values measured at the channels 14 and 15 that are located at 885 nm and 900 nm, respectively. We use the IWV maps produced by MERIS as a reference data for the purpose of comparison and evaluation. In cloudy regions, however, the columnar water values correspond to the water vapor integrated from the cloud top up to the sensor, which means that a large part of the perceptible water is missing since it is concentrated at low altitudes. We mask out IWV measurements in cloudy regions, which can result in removing more than 90% of the observations when clouds are covering wide areas of the map. Hence for the defined time period, only five MERIS IWV maps were observed under cloud-free conditions. An example of the IWV maps measured by MERIS, under clear sky weather, in the region covered by the ASAR image is shown in **Figure 3.5**.

The largest amount of IWV in this example is coming from the stratified water vapor in the neutrospheric layers.
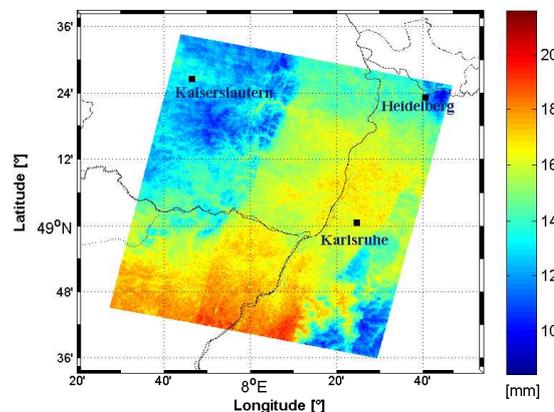
Figure 3.5: IWV as measured by MERIS on 09.05.2005 at 09:51 UTC. The scene covers approximately the area covered by the SAR image in the Upper Rhine region.

3.6 Weather Research and Forecasting Modeling System

Measurements of the water vapor from GNSS are available at spatially-sparse points, and MERIS provides measurements only under cloud free conditions. We investigated meteorological data produced by the weather research and forecasting modeling System (WRF) for two reasons. The first is, to have continuous grids of meteorological data required for wet delay

computations or for the sake of comparison. The second reason is that, IWV fields generated by WRF are used in Chapter 7 for data fusion to provide continuous grids of IWV. NWP models supply products of atmospheric processes at different spatial and temporal scales. WRF is the state-of-the-art atmospheric modeling system that bestows simulations of the atmosphere at local, regional, and global scales. WRF provides simulations of different atmospheric quantities at 2D or 3D continuous grids and the corresponding geographic locations and altitudes. Compared to its predecessors, WRF has been improved in the accuracy level and grid scales. However, the initial boundary conditions are ingested based on the output of a global atmospheric model, which is critical for predicting local atmospheric phenomena. This means that WRF can show scenarios at which the generated grids of IWV deviate from the real state of the atmosphere.

WRF data are received from the Institute of Meteorology and Climate Research, Atmospheric Environmental Research (IMK-IFU) at Karlsruhe institute of technology. We obtain, for example, simulations of air temperature, pressure, and relative humidity in 3D and 2D maps of the IWV. The model is run with 3 nested domains at spatial resolutions of 27, 9, and 3 km. The time-variant boundary conditions are addressed by ingesting the ERAInterim reanalysis data [Dee et al., 2011] received from the European Center for Medium-range Weather Forecasts (ECMWF) into the model. For the inner domain (3 km), 2D IWV fields are generated on a horizontal coverage of 500×500 km^2 centered on Karlsruhe (Germany). The data are simulated over a period of 15 months (July, 2004 - September, 2005) at a temporal rate of 10 minutes. Every point on the horizontal grid has a value for the water vapor integrated over 38 vertical layers, extending from the Earth's surface to heights of approximately 14 km. **Figure 3.6** shows the region over which the simulations were made and a map of the WRF-derived IWV for 06.27.2005 at 10:00 am. The IWV content is relatively high due to the high activity of the atmosphere in summer. IWV is strongly dependent of surface elevation such that the higher the elevation, the lower the IWV. The correlation of the water vapor content with the topography is clearly observed from the map.

Data set	System	Spatial resolution	Temporal resolution	Availability
GNSS	GPS	point-wise	30 sec	since 2002
PSI	Envisat	11 points/km^2	35 days	2003-2009
WRF	WRF 3.1	3×3 km^2	10 min	July, 2004 - September, 2005
MERIS	Envisat	260×290 m^2	3 days	2003-2009
LUBW	Meteorological sensors	point-wise	30-60 min	since 1991

Table 3.3: Spatial and temporal characteristics of the data sets used in this study.
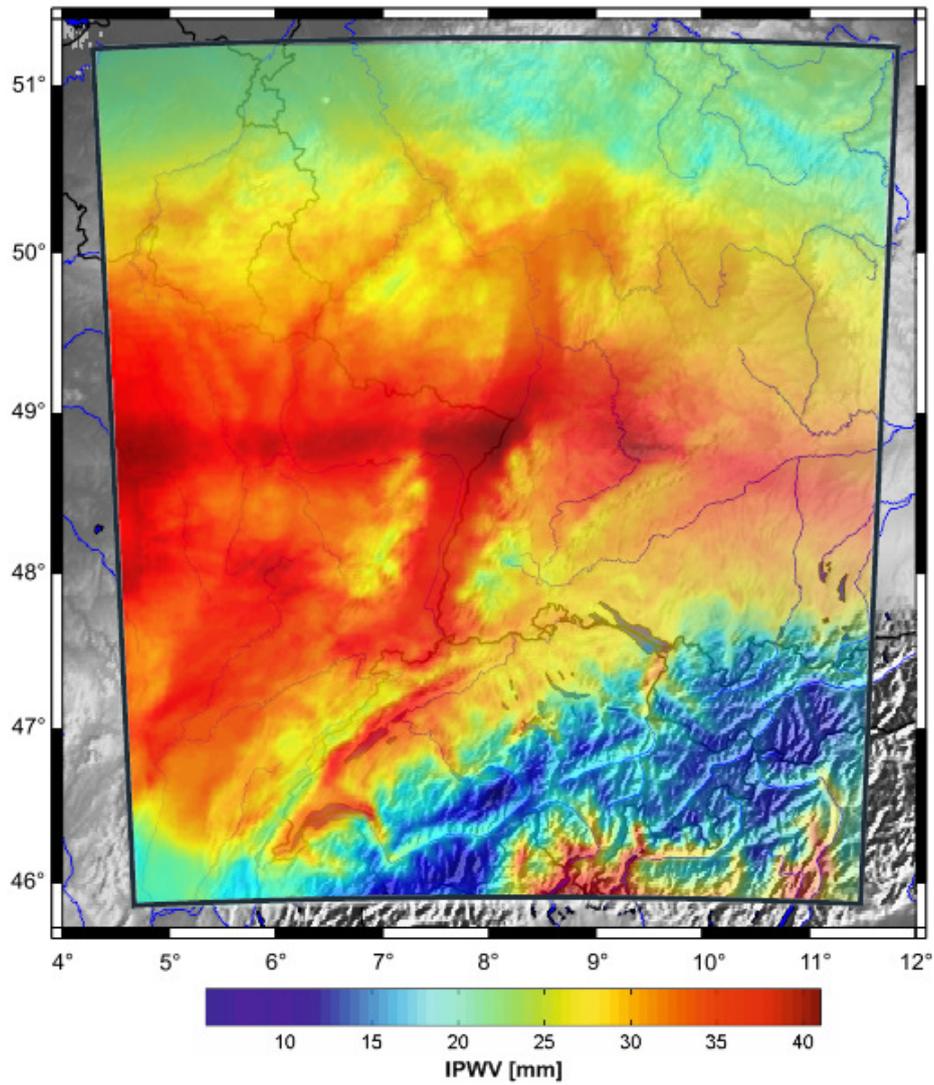
Figure 3.6: The map shows the test site URG, and the black frame specifies the grid over which the WRF simulations are made. An example of the IWV produces by the model on 06.27.2005 at 10:00 (UTC) is shown.

4 Spatial interpolation and estimation techniques

This chapter presents the spatial interpolation and estimation techniques applied to single data sets for the purpose of generating regular and continuous grids of all observed variables. First, we discuss the support of the data and the change of support problems, Section 4.1. In Section 4.2, we discuss the non-geostatistical interpolation techniques, moving-window averaging (local mean), curve fitting, and inverse distance weighting (IDW), then we discuss geostatistical techniques, i.e., kriging in Section sec: kriging. In Section 4.2.2.1, we present the ordinary kriging (OK) method and explain its limitations. We describe the variogram (also called structure functions) as a geostatistical tool used to describe the spatial relations between the observations. We present the method of FRK in Section 4.2.2.3, and we discuss how this method is used to overcome the limitations of OK. In Section 4.2.3, the method of block kriging is presented as a geostatistical solution for the point-to-area change of support problems. We describe the theoretical covariance model associated with the FRK method and the algorithm used for estimating its parameters in Section 4.3. In Section 4.4, we show a case study for using OK and FRK for spatial interpolations of neutrospheric parameters. At the end of the chapter, we give a summary of the presented techniques.

4.1 Change of support problems

One important aspect of remote sensing data is their support. In geostatistics, the term support reflects the geometrical size, shape, and orientation of the regions over which the measurement is available [Gelfand et al., 2001]. In practice, spatial data could be collected exactly at points or associated with areal units. The former are called *point-level* data or simply *point* data while the latter are *areal-level* or *block* data [Gelfand et al., 2001]. The *change of support problem* (COSP) is associated with the inference of values correctly at support levels different from the support level at which the data were observed. The COSP may result when studying single variables or when relating two or more spatial variables of different supports. If we inspect the different data sets investigated in this work, we find that PSI observations are point-level data, while WRF and MERIS provide areal-level data. GNSS also provide areal-level observations of the neutrospheric parameters. Accordingly, the change of support effect should be considered when comparing or assimilating these data. We should also resolve the COSP, if we are interested in estimating block-level data from data at point-level support

(point-to-area COSP) or estimating point-level processes but the available data are at areal-level (area-to-point COSP).

Changing the support of a variable, for example by aggregation, creates a new variable that is related to the original one but is expected to have different statistical and spatial properties [Gotway and Young, 2002]. For block data that can be expressed by block average, the following model should be appropriate. Consider a random variable $Y(\cdot)$ that captures a geophysical variable (also called field or process) over the spatial domain D. Let $\mathbf{s}$ be a point location in D, and suppose that, instead of observing $Y(\cdot)$ at a point-level ($\mathbf{s}_i$), it is observed at an areal-level (B_i), then the data model is defined as:

$$Y(B_i) = \frac{1}{|B_i|} \int_{B_i} Y(\mathbf{s}) d\mathbf{s} \tag{4.1}$$

$|B_i|$ is the observations volume within the footprint B_i. Assuming error-free observations, this model defines the observed process for the cell B_i as an average of the data as if observed at a point-level within the coverage of B_i. The assumption about block data as an average of point data holds, for example, for rainfall, temperature, pollutant level, and Earth's surface elevation [Gelfand et al., 2001]. We found that this model also holds for the wet path delay or IWV measurements. It is worth mentioning that the observations are in practice discrete, hence the integral in Eq. (4.1) is replaced with a summation.

Let the point-to-point or simply point covariance of the variable of interest be $C(\mathbf{s}_i, \mathbf{s}_j) = \mathrm{cov}(Y(\mathbf{s}_i), Y(\mathbf{s}_j))$, for all $\mathbf{s}$ in the domain of Y. Based on the data model in eq. (4.1), the covariance between aggregated data (areal-level) can be related to the point-level covariance as follows [Gotway and Young, 2002]

$$C(B_i, B_j) = \mathrm{cov}\left(\frac{1}{|B_i|} \int_{B_i} Y(\mathbf{u}) d\mathbf{u}, \frac{1}{|B_j|} \int_{B_j} Y(\mathbf{v}) d\mathbf{v}\right)$$
$$= \frac{1}{|B_i||B_j|} \int_{B_i} \int_{B_j} C(\mathbf{u}, \mathbf{v}) \, d\mathbf{u} \, d\mathbf{v} \tag{4.2}$$

where $C(B_i, B_j)$ is the block-to-block or block covariance function and $C(\mathbf{u}, \mathbf{v})$ is the point covariance function. Since the system adds a random noise to the measurements, we rewrite the model in (4.1), such that

$$Z(B_i) = \frac{1}{|B_i|} \int_{B_i} Y(\mathbf{s}) d\mathbf{s} + \epsilon(B_i) \tag{4.3}$$

where $\epsilon(B_i)$ is the measurement error, which is assumed uncorrelated Gaussian process of mean zero and a variance σ_ϵ^2. The noise is also assumed independent of the measurements.

The block covariance of the measurements can be written in terms of the point covariance as

$$
\begin{aligned}
\text{cov}(Z(B_i), Z(B_j)) &= \text{cov}\left(\frac{1}{|B_i|}\int_{B_i} Y(\mathbf{u})d\mathbf{u} + \epsilon(B_i), \frac{1}{|B_j|}\int_{B_j} Y(\mathbf{v})d\mathbf{v} + \epsilon(B_j)\right) \\
&= \frac{1}{|B_i||B_j|}\int_{B_i}\int_{B_j} C(\mathbf{u},\mathbf{v})\,d\mathbf{u}\,d\mathbf{v} + \text{cov}\left(\frac{1}{|B_i|}\int_{B_i} Y(\mathbf{u})d\mathbf{u}, \epsilon(B_i)\right) \\
&\quad + \text{cov}\left(\frac{1}{|B_j|}\int_{B_j} Y(\mathbf{v})d\mathbf{v}, \epsilon(B_j)\right) + \text{cov}\left(\epsilon(B_i), \epsilon(B_j)\right) \\
&= \frac{1}{|B_i||B_j|}\int_{B_i}\int_{B_j} C(\mathbf{u},\mathbf{v})\,d\mathbf{u}\,d\mathbf{v} + \sigma_\epsilon^2 \cdot \delta_{ij}
\end{aligned}
\tag{4.4}
$$

where δ_{ij} has a value of one only if $i = j$, and zero otherwise. The eqs. (4.2) and (4.4) form a basis for the COSP solution [Gotway and Young, 2002]. The point-level covariance should be known or estimated from the data. It is then used to define a valid positive definite theoretical covariance model, which is fitted for the purpose of prediction. These expressions will be required in the method of block kriging, which is used to solve point-to-area COSP (Section 4.2.3), to relate point-level covariance to block-level covariance.

4.2 Spatial interpolation techniques

Spatial continuous fields of atmospheric water vapor are not readily available from remote sensing satellites. PSI, for example, provides clusters of measurements which are most likely absent in vegetated areas (as presented in Chapter 5), while MERIS measurements are removed in cloudy regions since they are significantly underestimated. For the sake of comparison or making justified interpretations about a geophysical phenomenon, spatial interpolation techniques are essential for estimating continuous grids of the variable of interest from the discrete measurements. There are numerous spatial interpolation methods which can fall into one of three categories, namely: non-geostatistical interpolators, geostatistical interpolators, or combined methods [Li and Heap, 2008, pp. 4]. The non-geostatistical methods are based on weighting the surrounding measured values and use mathematical formulas that determine the smoothness of the resulting surface. The geostatistical methods are based on statistical models that include spatial correlations among the measurements [Li and Heap, 2008, pp. 11].

The majority of spatial interpolation methods computes estimations as a weighted average of sampled data. The general estimation formula is written as:

$$
\hat{Z}(\mathbf{s}_0) = \sum_{i=1}^{N} \lambda_i Z(\mathbf{s}_i)
\tag{4.5}
$$

where $\hat{Z}(\mathbf{s}_0)$ is the estimated value at the point $\mathbf{s}_0$, $Z(\mathbf{s}_i)$ is the observed value at the location

$\mathbf{s}_i$, λ_i is the weighting coefficient, and N is the number of observations. In the method of local mean or also called moving-window averaging, a value is computed at the location $\mathbf{s}_0$ using eq. (4.5) by averaging the observations falling within the neighborhood of a predefined cell centered at the location $\mathbf{s}_0$, i.e., $\lambda_i = 1/N$. The method of curve fitting models the spatial variations in the data by creating an interpolant that has the best fit to the scattered data. This interpolant can then be utilized to compute a value of the variable at any location.

4.2.1 Inverse distance weighting

Inverse distance weighting (IDW) is a method that infers a value at a certain location by a weighted mean of the observations available within a predefined radius. The weights are selected such that they are inversely proportional to the squared distance to the prediction location. Given the observations $Z(\mathbf{s}_i) = [Z(\mathbf{s}_1), \ldots, Z(\mathbf{s}_N)]'$ at the locations $\mathbf{s}_i = [\mathbf{s}_1, \ldots, \mathbf{s}_N]'$, a new value can be inferred at the location $\mathbf{s}_0$ using

$$\hat{Z}(\mathbf{s}_0) = \frac{\sum_{i=1}^{N} \frac{1}{D_{i0}^2} Z(\mathbf{s}_i)}{\sum_{i=1}^{N} \frac{1}{D_{i0}^2}}, \qquad D_{i0} > 0 \qquad (4.6)$$

where D_{i0} is the Euclidean distance between the location of the ith observation and the prediction location. IDW is a simple interpolation method, and it is not a model-based technique. If IDW is used for interpolation in regions where the measurements are missing or outside the measurements coverage, the uncertainty increases and the quality of predictions is poor.

4.2.2 Kriging

Kriging is a geostatistical interpolation technique that provides predictions of the quantity of interest by considering the spatial correlations between the measurements. IDW and kriging are similar in that they predict a new value by weighting the surrounding measured values. Kriging differs, however, in that it requires not only the distance to the predicted location, but also the spatial correlations. Kriging is used in this work to generate estimates of the wet path delay. PSI, for example, provides measurements of the wet delay at scattered locations and sometimes there could be gaps in the map due to the lack of coherence in vegetation regions or over water surfaces. Taking spatial correlation between observations into account, the values of the wet delay can be inferred at a regular grid by applying, for example, the technique of OK. This is a widely used prediction technique since the predictions are made by looking for the best linear unbiased estimator and it provides measures of the quality of estimates. In the following, we present the ordinary kriging (OK) method, its limitations, and the alternative method Fixed-Rank Kriging (FRK). Most of the material presented to explain the OK method draws from Deutsch and Journel [1998] and Li and Heap [2008].

The approach of kriging divides the signal into a component that is mostly deterministic and can be modeled by a linear trend and a component that is more random. Consider a geostatistical field $\mathbf{Z}$ of N point-level observations. The data are observed at a finite number of spatial locations $[\mathbf{s}_1, \ldots, \mathbf{s}_N]'$ such that $\mathbf{Z} \equiv [Z(\mathbf{s}_1), \ldots, Z(\mathbf{s}_N)]'$. The Gaussian forward model of $\mathbf{Z}(\mathbf{s})$ is a sum of a real-valued true spatial process $\mathbf{Y}(\mathbf{s})$ and a spatial white noise process $\boldsymbol{\epsilon}$

$$\mathbf{Z}(\mathbf{s}) = \mathbf{Y}(\mathbf{s}) + \boldsymbol{\epsilon}(\mathbf{s}) \tag{4.7}$$

$\mathbf{Y}(\mathbf{s}) = [Y(\mathbf{s}_1), \ldots, Y(\mathbf{s}_N)]'$, $\mathbf{s} \in D$ and D is the spatial domain of the geophysical process. The measurement error is assumed to be uncorrelated zero-mean Gaussian process with a finite variance σ_ϵ^2. The hidden process $Y(\mathbf{s})$ is modeled by a linear combination of a deterministic trend and a spatially-random component [Cressie and Johannesson, 2008]. The former reflects the long-scale spatial behavior of $\mathbf{Y}$, while the latter contains the spatial variations that cannot be modeled by a linear trend. The model for the hidden process $Y(\mathbf{s})$ has the following [Cressie and Johannesson, 2008]:

$$Y(\mathbf{s}) = \mathbf{T}(\mathbf{s}) \cdot \boldsymbol{\alpha} + \boldsymbol{\nu}(\mathbf{s}) \tag{4.8}$$

where $\mathbf{T}(\mathbf{s}) \cdot \boldsymbol{\alpha}$ indicates a deterministic linear trend, $\mathbf{T}$ has a size of $N \times 3$ and each row has three entries: a one and the longitude and latitude of the location $\mathbf{s}$. $\boldsymbol{\alpha}$ is the vector of least squares regression coefficient. $\boldsymbol{\nu}(\mathbf{s})$ is a spatial process of mean zero and has generally a non-stationary covariance function. Applying the method of kriging requires removing the deterministic signal from the data ahead of prediction, we use the estimate of $\boldsymbol{\alpha}$ to calculate the residuals,

$$\tilde{\mathbf{Z}} = \mathbf{Z} - \mathbf{T}\hat{\boldsymbol{\alpha}} \tag{4.9}$$

In the absence of prior information about the spatial covariance, $\boldsymbol{\alpha}$ is estimated from the measurements by means of ordinary least squares adjustment as follows

$$\hat{\boldsymbol{\alpha}} = (\mathbf{T} \ \mathbf{T}')^{-1}\mathbf{T}'\mathbf{Z} \tag{4.10}$$

The residual field contains the signal of medium to fine spatial random variations and the noise. We proceed with the analysis using the detrended field, which is assumed to have a mean zero. From eqs. (4.7) and (4.8), the residual signal can be written as:

$$\tilde{\mathbf{Z}}(\mathbf{s}) = \boldsymbol{\nu}(\mathbf{s}) + \boldsymbol{\epsilon}(\mathbf{s}) \tag{4.11}$$

Kriging estimates the value of $Z(\mathbf{s}_0)$ at the known location $\mathbf{s}_0$ from Eq.(4.5), where the weights λ_i are selected such that the mean squared prediction error (MSPE) between the estimated

value and the true process is minimized, i.e.,

$$\min_{\lambda_i} \quad \mathrm{E}\left\{ \left(\sum_{i=1}^{N} \lambda_i Z(\mathbf{s}_i) - Z(\mathbf{s}_0) \right)^2 \right\} \tag{4.12}$$

To ensure unbiased prediction, it is required that $\sum_{i=1}^{N} \lambda_i = 1$.

4.2.2.1 Ordinary Kriging

In the ordinary kriging, the weights in $\boldsymbol{\lambda}$ are derived on the basis of a covariance function or a semivariogram, which should characterize the spatial properties of the residual signal $\tilde{\mathbf{Z}}$. In order to estimate $\boldsymbol{\lambda}$, the Lagrange multipliers are utilized to solve the constrained minimization and the solution is given by:

$$\boldsymbol{\lambda} = \mathbf{c}_0'(\mathbf{s}_0)\boldsymbol{\Gamma}^{-1} \tag{4.13}$$

$$\text{with: } \boldsymbol{\Gamma} = \begin{bmatrix} \gamma_{11} & \cdots & \gamma_{1N} & 1 \\ \vdots & \ddots & \vdots & \vdots \\ \gamma_{N1} & \cdots & \gamma_{NN} & 1 \\ 1 & \cdots & 1 & 0 \end{bmatrix}, \mathbf{c}_0 = \begin{bmatrix} \gamma_{10} \\ \vdots \\ \gamma_{N0} \\ 1 \end{bmatrix}, \text{ and } \boldsymbol{\lambda} = \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_N \\ m \end{bmatrix}$$

γ_{ij} denotes the semi-variance corresponding to the distance separating the observations i and j, and its value is computed from the semivariogram model (presented in Section 4.2.2.2). m is the unknown Lagrange parameter associated with the unbiasedness constraint. γ_{i0} denotes the modeled semivariance corresponding to the distance separating the ith observation and the prediction location. Let $\mathbf{k} = [\lambda_1, \lambda_2, \cdots, \lambda_N]$, then the formula of the OK predictor is:

$$\hat{Y}(\mathbf{s}_0) = \mathbf{T}(\mathbf{s}_0)\hat{\boldsymbol{\alpha}} + \mathbf{k}\tilde{\mathbf{Z}} \tag{4.14}$$

It is conventional to use the semivariogram to form the matrix $\boldsymbol{\Gamma}$ rather than the covariance although kriging systems are easily solved with the covariance matrices [Deutsch and Journel, 1998, pp. 13]. Assuming that $\boldsymbol{\nu}$ and $\boldsymbol{\epsilon}$ are independent, the covariance function of $\tilde{\mathbf{Z}}$ is given by

$$\boldsymbol{\Sigma} = \mathbf{C} + \sigma_\epsilon^2 \mathbf{V}_\epsilon \tag{4.15}$$

$\mathbf{C}$ is the covariance matrix of $\boldsymbol{\nu}$, where $C(\mathbf{u}, \mathbf{v}) = \mathrm{cov}(\nu(\mathbf{u}), \nu(\mathbf{v}))$. $\sigma_\epsilon^2 \mathbf{V}_\epsilon$ is the covariance matrix of the noise and since the noise is assumed independent and identically distributed $\mathbf{V}_\epsilon$ is an identity matrix of the observations size. The covariance matrix can be estimated from the detrended field as described later in this chapter. The formula for the kriging predictor in

terms of the full covariance function has the following form [Cressie and Johannesson, 2008]

$$\hat{Y}(\mathbf{s}_0) = \mathbf{T}(\mathbf{s}_0)\hat{\boldsymbol{\alpha}} + \mathbf{k}(\mathbf{s}_0)\tilde{\mathbf{Z}} \tag{4.16}$$

where $\hat{\boldsymbol{\alpha}}$ is determined from eq. (4.10) and $\mathbf{k}(\mathbf{s}_0)$ is calculated from:

$$\mathbf{k}(\mathbf{s}_0) = \mathbf{c}(\mathbf{s}_0)'\boldsymbol{\Sigma}^{-1} \tag{4.17}$$

$\mathbf{c}(\mathbf{s}_0) \equiv [C(\mathbf{s}_0, \mathbf{s}_1), \dots, C(\mathbf{s}_0, \mathbf{s}_N)]'$ is a vector that contains the covariance values between the process at the observed locations and the process at the prediction location $\mathbf{s}_0$. These covariance values are derived from the theoretical covariance model used by the kriging algorithm.

4.2.2.2 Semivariograms

The *variogram*, also called *structure function* [Wald, 1983], is a geostatistical tool that has been widely used to measure the spatial variability of the phenomenon under study. Considering the stationary random function $Z(\mathbf{u})$, the theoretical variogram $D(\mathbf{h})$ is defined as the expected squared difference between two or more data values separated by a displacement vector $\mathbf{h}$, i.e.,

$$D(\mathbf{h}) = 2\gamma(\mathbf{h}) = E\{[Y(\mathbf{u} + \mathbf{h}) - Y(\mathbf{u})]^2\}, \tag{4.18}$$

where $\mathbf{u}$ is the spatial location of the measurement and $\mathbf{h}$ is the displacement vector. It is conventional to denominate $\gamma(\mathbf{h})$ a *semivariogram*. If there exist N pairs of observations separated by a displacement vector $\mathbf{h}$, then the classical semivariogram estimator as presented in [Gringarten and Deutsch, 2001] is

$$\gamma(\mathbf{h}) = \frac{1}{2|N(\mathbf{h})|} \sum_{i=1}^{|N(\mathbf{h})|} [Z(\mathbf{u}_i) - Z(\mathbf{u}_j)]^2 \tag{4.19}$$

$N(\mathbf{h}) = \{(\mathbf{u}_i, \mathbf{u}_j) : \mathbf{u}_i - \mathbf{u}_j = \mathbf{h}\}$ and $|N(\mathbf{h})|$ is the cardinality of $N(\mathbf{h})$. The semivariogram is related to the stationary covariance $C(\mathbf{h})$ as

$$\gamma(\mathbf{h}) = C(0) - C(\mathbf{h}) \tag{4.20}$$

with $C(0)$ the a priori variance of the random function, i.e., $C(0) = \mathrm{var}\{Z(\mathbf{u})\}$.

The variogram illustrates the spatial variability as the observations get more distant. In general, the semivariogram should have an initial value of zero at the origin that increases with $\boldsymbol{h}$ until a certain spatial range. On the contrary to the semivariance, the spatial correlation between the observations decreases with the separation distance until a distance at which

no spatial correlation is evident. This is called the *range of correlation*. The semivariance corresponding to the range is called the *sill*, which is simply the a priori variance of the random function, $C(0)$. Another important measure of the semivariogram is the short-scale semivariance or the *nugget effect*. It is a discontinuity in the semivariogram at the origin. The empirical semivariogram should show no nugget effect, but sometimes it exists due to errors in the data or spatial structures with correlation ranges shorter than the minimum sample spacing (1-lag separating distance) [Gringarten and Deutsch, 2001]. **Figure 4.1** shows typical and experimental semivariograms. The experimental semivariogram is computed for a PSI neutrospheric delay-difference map using (4.19). From the plot we observe that the range is approximately 25 km and the corresponding sill is 18.66 mm^2. The spatial correlation between the observations, which are used to fine the kriging solution, are derived from the semivariogram.

(a) Typical semivariogram

(b) Experimental semivariogram
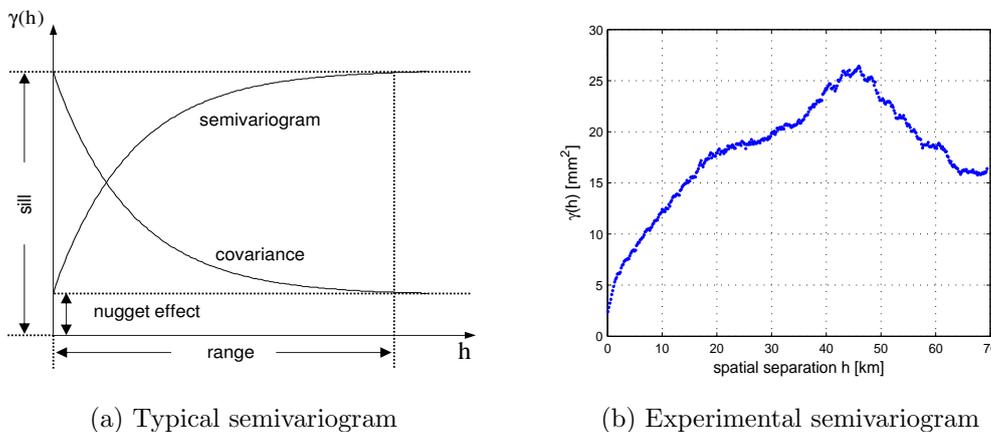
Figure 4.1: Typical and experimental semivariograms. The experimental semivariogram is computed for a PSI neutrospheric delay-difference map, the range is approximately 25 km and the sill is 18.66 mm^2.

The estimator in expression (4.19) can behave poorly if the data contain outliers; thus, a more robust semivariogram estimator was proposed by Cressie and Hawkins [1980],

$$2\gamma(\mathbf{h}) = \left(\frac{1}{|N(\mathbf{h})|} \sum_{N(\mathbf{h})} \left| Z(\mathbf{u}_i) - Z(\mathbf{u}_j) \right|^{1/2} \right)^4 \bigg/ \left(0.457 + \frac{0.494}{|N(\mathbf{h})|} \right) \qquad (4.21)$$

Prediction of the spatial process at uncovered locations using ordinary kriging requires a theoretical semivariogram or a covariance model. Therefore, it is necessary to fit a theoretical semivariogram model to the empirical one. There are various methods for model fitting, such as least squares, maximum likelihood, and robust methods [Lark, 2000]. There are different standard theoretical semivariogram models such as spherical, exponential, Gaussian, and other models. Once the data are explored, the experimental semivariogram is obtained, a variogram model is fitted, and the output grid is defined. We use eq. (4.13) to obtain the λ_i's after fitting

a semivariogram model depending on the performance of the experimental semivariogram. The weights are substituted in eq. (4.5) to get predictions at the predefined locations.

Two common assumptions are made in the above presented OK method, these are stationarity and isotropy. Stationarity implies that the covariance of the random function at two different locations is a function of the displacement vector between them, i.e.,

$$C(Z(\mathbf{u}_i), Z(\mathbf{u}_j)) = C(\mathbf{u}_i - \mathbf{u}_j)$$

Isotropy indicates that the process is uniform in all directions, so that one variogram is representative for the spatial variations of the signal. Hence the covariance is a function of the distance between the two locations (h),

$$C(Z(\mathbf{u}_i), Z(\mathbf{u}_j)) = C(h)$$

These assumptions do not necessarily hold for geophysical data [Braverman et al., 2009; Nguyen, 2009]. However, the most critical limitation of conventional kriging is the lack of scalability. As observed from eq. (4.13), calculating kriging weighting coefficients requires constructing and inverting the matrix $\boldsymbol{\Gamma}$ or $\boldsymbol{\Sigma}$, which have the size $(N+1) \times (N+1)$, for N observations. For massive data sets, where N could be on the order of hundreds of thousands or more, traditional kriging is not possible even with high-end consumer-oriented computers [Nguyen, 2009]. There are a number of different investigations to make kriging feasible for huge data sets, for more details the reader is referred to Nguyen [2009]. An approach called fixed-rank kriging introduced by Cressie and Johannesson [2008] is investigated in this work.

4.2.2.3 Fixed-rank kriging and multi-scale process modeling

Generally, kriging requires a theoretical covariance (or variogram) model to compute the covariance $C(\mathbf{s}_0, \mathbf{s}_i)$ for any output location $\mathbf{s}_0$. We look for a covariance model that does not make any assumptions of stationary and isotropy. Also, it is obvious that inverting the matrix $\boldsymbol{\Sigma}$ in expression (4.17) is unavoidable; however, direct inversion is not always achievable specially for large N. We address these problems by using the fixed-rank kriging (FRK) algorithm presented by Cressie and Johannesson [2008]. In this algorithm, the authors suggested the following covariance model for the signal $\boldsymbol{\nu}(\mathbf{s})$:

$$C(\mathbf{s}_i, \mathbf{s}_j) = \mathbf{S}(\mathbf{s}_i) \mathbf{K} \mathbf{S}(\mathbf{s}_j)' \tag{4.22}$$

$\mathbf{K}$ is a positive-definite covariance matrix of size $r \times r$ and $\mathbf{S}$ has the size $N \times r$ with $r \ll N$.

This covariance model implies that the term $\boldsymbol{\nu}(\boldsymbol{s})$ in the expression (4.11) is extended into:

$$\boldsymbol{\nu}(\mathbf{s}) = \mathbf{S}(\mathbf{s})\boldsymbol{\eta} \tag{4.23}$$

where $\boldsymbol{\eta} = [\eta_1, \ldots, \eta_r]'$ is a zero-mean vector of r spatial random effects with a covariance matrix $\mathbf{K} = \text{var}(\boldsymbol{\eta})$. According to Cressie and Johannesson [2008], the model of $\boldsymbol{\nu}(\mathbf{s})$ is called the *spatial random effects* (SRE) model, which represents a multi-scale modeling of the geophysical process. The models in the eqs. (4.22) and (4.23) assumes that the true process is smooth, so we use the most general model presented in [Kang and Cressie, 2011] that considers not only the smooth variations but also the fine-scale variations of rough processes,

$$C(\mathbf{s}_i, \mathbf{s}_j) = \mathbf{S}(\mathbf{s}_i)\,\mathbf{K}\,\mathbf{S}(\mathbf{s}_j)' + \sigma_\zeta^2 \cdot \mathbf{I}(\mathbf{s}_i = \mathbf{s}_j), \qquad \forall \mathbf{s}_i, \mathbf{s}_j \in D \tag{4.24}$$

where $\sigma_\zeta^2 \cdot \mathbf{I}(\boldsymbol{s_i} = \boldsymbol{s_j})$ is the covariance matrix of the fine-scale signal. The corresponding SRE model is

$$\boldsymbol{\nu}(\mathbf{s}) = \mathbf{S}(\mathbf{s})\boldsymbol{\eta} + \boldsymbol{\zeta}(\mathbf{s}) \tag{4.25}$$

$\boldsymbol{\zeta}(\mathbf{s})$ is assumed uncorrelated Gaussian with mean zero and a variance σ_ζ^2. $\boldsymbol{\zeta}$ accounts for the fine-scale spatial variations and is assumed independent of $\boldsymbol{\eta}$.

From the SRE model in (4.25), we observe that the detrended signal is represented by a linear combination of the elements of the vector $\boldsymbol{\eta}$ added with a random signal due to the fine-scale variations. Each row in the matrix $\mathbf{S}$ contains r weights used to encode each location $\mathbf{s}$ such that we can find a vector $\boldsymbol{\eta}$ for which $\sum_{j=1}^r S_j(\boldsymbol{s})\eta_j$ gives an approximation of the detrended component of $\mathbf{Y}$ at that location. The weights depends on the distance between the location $\mathbf{s}$ and a set of predefined nodes, and their values are obtained from a set of basis functions (BF) with their centers located at the r nodes. **Figure 4.2** (a) shows the observation domain and the nodes at the locations $\mathbf{m}_i$, where each observation location $\mathbf{s}$ in the observation domain is encoded by four weights depending upon its distance d_i from $\mathbf{m}_i$. The radius of the blue circle defines the radius within which the BF has a positive value.

The nodes are selected to cover multiple resolutions so that they can capture spatial variations at different spatial scales. At each level of spatial resolution, the BFs should have a certain effective radius and the number of the resolution levels is determined according to the spatial variations in the observed field. There are several choices of the BFs that could be used such as smoothing spline BFs, wavelet BFs, and bisquare BFs [Cressie and Johannesson, 2008]. The coefficients $\mathbf{S}(\mathbf{s}) = [S^{(1)}(\mathbf{s}), S^{(2)}(\mathbf{s}), \ldots, S^{(r)}(\mathbf{s})]$ are calculated based on the locations of the observations and not the observations themselves. In **Figure 4.2** (b) is shown the bisquare BF, which is a bell-shaped function with a peak one at the center. The local bisquare

(a) Nodes setup

(b) Bisquare BF
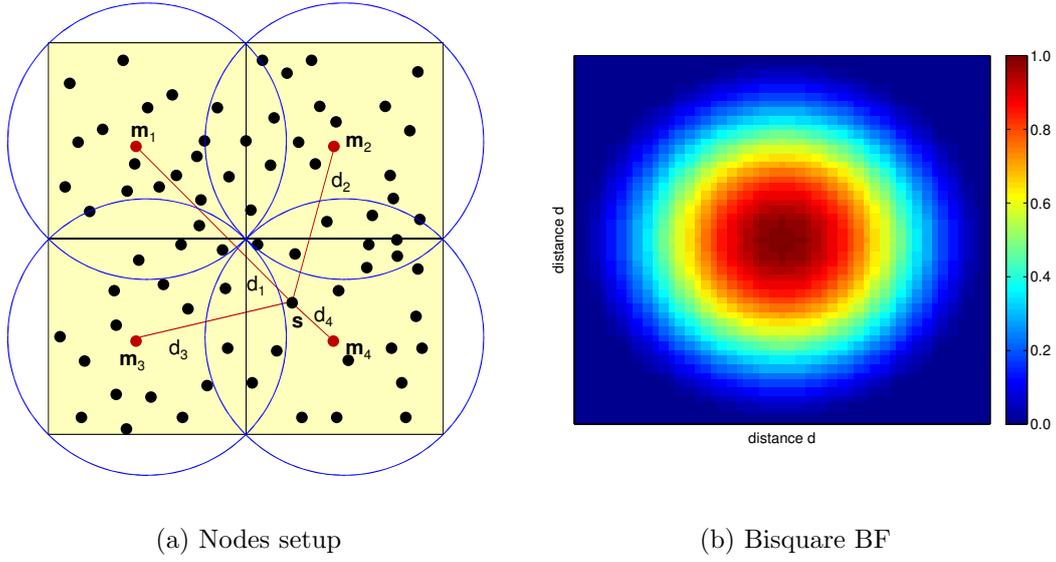
Figure 4.2: The observation domain is shown in (a), where the black dots define the locations at which the data were observed. The red dots indicate the nodes of the encoding scheme. The weights for each location **s** are related to the distances d_i and their value is calculated from BFs centered at the nodes $\mathbf{m}_i$. The blue circles define the radius of the BF. In (b) is shown the bisquare BF.

function at each location **s** has the form

$$
S^{(l)}(\mathbf{s}) = \begin{cases} \left[1 - (||\mathbf{s} - \mathbf{m}_l||/r_l)^2\right]^2, & \text{for } ||\mathbf{s} - \mathbf{m}_l|| \leq r_l, \\ 0 & \text{otherwise} \end{cases} \tag{4.26}
$$

$\mathbf{m}_l$ is the node (center point) of the BF of the lth resolution, and r_l is the effective radius of the BF at the lth resolution. The bisquare function is a bell-shaped function that has its maximum value at $\mathbf{m}_l$ and decreases smoothly until it reaches zero at widths larger than r_l. The bisquare BF curve agrees with the decrease in correlation as the distance between observations increases. It is worth mentioning that the matrix **S** is known and remains fix since the encoding nodes and the BFs are selected by the user. $\boldsymbol{\eta}$, however, is not known and has be estimated from the observations. Using the SRE model in (4.25), the model for $Y(\boldsymbol{s})$ will be

$$
Y(\mathbf{s}) = \mathbf{T}(\mathbf{s})\boldsymbol{\alpha} + \mathbf{S}(\mathbf{s})\boldsymbol{\eta} + \boldsymbol{\zeta}(\mathbf{s}) \tag{4.27}
$$

The covariance matrix ($\boldsymbol{\Sigma}$) of $\tilde{\mathbf{Z}}$ is expressed as

$$
\begin{aligned}
\boldsymbol{\Sigma} &= \text{cov}(\mathbf{S}\boldsymbol{\eta}) + \text{cov}(\boldsymbol{\zeta}) + \text{cov}(\boldsymbol{\epsilon}) \\
&= \mathbf{S}\,\mathbf{K}\,\mathbf{S}' + \sigma_\zeta^2 \mathbf{V}_\zeta + \sigma_\epsilon^2 \mathbf{V}_\epsilon
\end{aligned} \tag{4.28}
$$

Note that σ_ζ^2 and σ_ϵ^2 are not summed as one component, because the first is a part of the predicted process, while the other is the noise variance. $\mathbf{V}_\zeta$ and $\mathbf{V}_\epsilon$ are identity matrices. The covariance between the true process at the prediction location and the observations is given by

$$
\begin{aligned}
\mathbf{c}(\mathbf{s}_0) &= \operatorname{cov}\{Y(\mathbf{s}_0), \mathbf{Z}\} = \operatorname{cov}\{(\mathbf{S}(\mathbf{s}_0)\boldsymbol{\eta} + \boldsymbol{\zeta}(\mathbf{s_0})), (\mathbf{S}\boldsymbol{\eta} + \boldsymbol{\zeta} + \boldsymbol{\epsilon})\} \\
&= \mathbf{S}_p(\mathbf{s}_0)\mathbf{K}\,\mathbf{S}' + \sigma_\zeta^2 \mathbf{1}(\mathbf{s}_0 = \mathbf{s})
\end{aligned}
\tag{4.29}
$$

$\mathbf{S}_p(\cdot)$ is the weighting matrix for the prediction locations. $\mathbf{1}(\mathbf{s}_0 = \mathbf{s})$ is a vector of the size N with a value one at the location $\mathbf{s}_0 = \mathbf{s}$ and zeros elsewhere. In a similar way to OK, the FRK estimator of the process Y has the following form:

$$
\hat{Y}(\mathbf{s}_0) = \mathbf{k}(\mathbf{s}_0)'\tilde{\mathbf{Z}}
\tag{4.30}
$$

with

$$
\mathbf{k}(\mathbf{s}_0) = \mathbf{c}(\mathbf{s}_0)'\boldsymbol{\Sigma}^{-1}
\tag{4.31}
$$

We can rewrite Eq. (4.16) in a more insightful form

$$
\hat{Y}(\mathbf{s}_0) = \mathbf{T}(\mathbf{s}_0)\hat{\boldsymbol{\alpha}} + \left(\mathbf{S}_p(\mathbf{s}_0)\,\mathbf{K}\,\mathbf{S}' + \sigma_\zeta^2 \cdot \mathbf{1}(\mathbf{s}_0 = \mathbf{s})\right)\boldsymbol{\Sigma}^{-1}\tilde{\mathbf{Z}}
\tag{4.32}
$$

and the corresponding MSPE of $\hat{Y}(\mathbf{s}_0)$, $E\{(\hat{Y}(\mathbf{s}_0) - Y(\mathbf{s}_0))^2\}$, is:

$$
\sigma^2(\mathbf{s}_0) = C(\mathbf{s}_0, \mathbf{s}_0) - \mathbf{k}(\mathbf{s}_0)'\boldsymbol{\Sigma}\mathbf{k}(\mathbf{s}_0)' + (\mathbf{T}(\mathbf{s}_0) - T'\mathbf{k}(\mathbf{s}_0)')'(\mathbf{T}\,\mathbf{T}')^{-1}(\mathbf{T}(\mathbf{s}_0) - \mathbf{T}'\mathbf{k}(\mathbf{s}_0}'))
\tag{4.33}
$$

So far, all components required to obtain the predictions and predictions errors can be found either from one of the above expressions of as explained in Section 4.3. Table 4.1 summarizes the procedures followed to calculate predictions based on the FRK technique.

It remains to find a feasible way to achieve the matrix inversion in eq (4.32) for large N. Let the matrix $\mathbf{D}_i = \sigma_\zeta^2\mathbf{I} + \sigma_\epsilon^2\mathbf{I}$, then the matrix $\boldsymbol{\Sigma}$ can be inverted by applying the Sherman-Morrison-Woodbury formula [Hager, 1989],

$$
\begin{aligned}
\boldsymbol{\Sigma}_{ii}^{-1} &= (\mathbf{D}_i + \mathbf{S}_i\,\mathbf{K}_i\,\mathbf{S}_i')^{-1} \\
&= \mathbf{D}_i^{-1} - \mathbf{D}_i^{-1}\mathbf{S}_i(\mathbf{K}^{-1} + \mathbf{S}_i'\,\mathbf{D}_i^{-1}\,\mathbf{S}_i)^{-1}\mathbf{S}_i'\mathbf{D}_i^{-1}
\end{aligned}
\tag{4.34}
$$

The computations require the inversion of the matrices $\mathbf{K}$ and $(\mathbf{K}^{-1} + \mathbf{S}_i'\,\mathbf{D}_i^{-1}\,\mathbf{S}_i)$, both of which have the size $r \times r$. Note that $\mathbf{D}_i$ is a diagonal matrix and it is inverted by taking the reciprocal of the diagonal elements. Using FRK, the matrix $\boldsymbol{\Sigma}^{-1}$ can be obtained by inverting matrices of the dimension $r \times r$ with a fixed r; therefore, it is called fixed-rank kriging. The

In the FRK technique, we need to:
1. define the input data and spatial locations
2. define the output grid (prediction locations)
3. estimate $\boldsymbol{\alpha}$ and detrend the data
4. compute the robust semivariogram and estimate the error variance σ_ϵ^2, if unknown
5. fix r and set up the nodes and the BFs
6. compose $\mathbf{S}$ for the input locations and the output locations
7. determine the empirical covariance matrix of the detrended data
8. compute the maximum likelihood estimates of $\mathbf{K}$ and σ_ζ^2
9. obtain the predictions and the corresponding MSPE

Table 4.1: Procedure of the FRK method

computational burden for inverting the matrix Σ reduces from $O(N^3)$ to $O(N)$ [Cressie and Johannesson, 2008]. This means that the computation costs for the FRK method is linear with the observations size N, which makes FRK feasible for huge data sets.

4.2.3 Block kriging

Spatial data could be observed exactly at point locations or as an average over a given footprint which is called the support of the data, as mentioned earlier. In many occasions, we are interested in finding predictions over a block (cell) rather than at point locations. In fact, this a change of support problem, where point-level observations are available, while we are interested in block-level predictions. A naive solution is to allocate a regularly-spaced grid over the block and predict the values at all points and then calculate the average over the block, as shown in **Figure 4.3**. This is conceptually simple but if utilized with OK, this procedure may be computationally extensive if the number of point estimates is very high. In geostatistics, the method of block kriging has been developed to solve the point-to-area COSPs [Gotway and Young, 2002], such that it achieves the same result by constructing and solving only one kriging system for each block estimate. In this section, we describe how the method of block kriging is applied under FRK and OK.

Suppose that the data $\mathbf{Y} = [Y(B_1), Y(B_2), \ldots, Y(B_N)]'$, at a cell size B_i, should be predicted from point-level observations by means of kriging. Applying kriging requires the covariance function of the observations and a covariance model used to obtain the covariances between the predictions and the observations. As presented before, the point-level covariance function can be related to the block-level covariance. Using the FRK covariance structure (4.24), the

point-to-block covariance in eq. (4.4) can be written as:

$$
\begin{aligned}
\mathrm{cov}(Z(B_i), Z(B_j)) &= \frac{1}{|B_i|}\frac{1}{|B_j|}\int_{B_i}\int_{B_j}\mathbf{S}(\mathbf{u})\,\mathbf{K}\,\mathbf{S}(\mathbf{v})du\,dv \\
&= \frac{1}{|B_i|}\int_{B_i}\mathbf{S}(\mathbf{u})d\mathbf{u}\,\mathbf{K}\,\frac{1}{|B_j|}\int_{B_j}\mathbf{S}(\mathbf{v})d\mathbf{v} \\
&= \tilde{\mathbf{S}}(B_i)\,\mathbf{K}\,\tilde{\mathbf{S}}(B_j)' + \sigma_\zeta^2\delta_{ij} + \sigma_\epsilon^2\delta_{ij} \qquad (4.35)
\end{aligned}
$$

where

$$
\delta_{ij} = \begin{cases} 1 & \text{for } i = j \\ 0 & \text{for } i \neq j \end{cases}
$$

with the matrix $\mathbf{S}(\cdot)$ aggregated within the block B_i,

$$
\tilde{\mathbf{S}} = \left[\tilde{S}_1(B_i), \tilde{S}_2(B_i), \ldots, \tilde{S}_r(B_i)\right]
$$
$$
\tilde{S}_j(B_i) = \frac{1}{|B_i|}\int_{B_i}S_j(\mathbf{u})d\mathbf{u}
$$

We should notice that the covariance terms due to fine-scale signal ($\boldsymbol{\zeta}$) and the noise disappeared from the expression, since both signals are assumed to be spatially uncorrelated and independent from other components. Expression (4.35) relates the block covariance with the point covariance on the basis of the FRK covariance model. The covariance between two point locations $\mathbf{s}_1$ and $\mathbf{s}_2$ is $\mathbf{S}(\mathbf{s}_1)\,\mathbf{K}\,\mathbf{S}(\mathbf{s}_2)'$. The covariance parameter $\mathbf{K}$ can be used to obtain the covariance function at any aggregation level by changing the structure of the matrix $\mathbf{S}$ from point-level to aggregated area-level. In a similar way, the covariance function between the input signal $Z(B_i)$ and the true process at a point location $\mathbf{s}$ is given by

$$
\begin{aligned}
\mathrm{cov}(Z(B_i), Y(\mathbf{s}))) &= \mathrm{cov}\left(\left(\frac{1}{|B_i|}\int_{B_i}Y(\mathbf{s})d\mathbf{s} + \epsilon(B_i)\right), Y(\mathbf{s})\right) \\
&= \tilde{\mathbf{S}}(B_i)\,\mathbf{K}\,\mathbf{S}_p(\mathbf{s})' \qquad (4.36)
\end{aligned}
$$

where $\mathbf{S}_p(\mathbf{s})$ is the weighting matrix for the prediction locations, which are at point-level in this case. The block kriging weights $\boldsymbol{\lambda}$ are obtained by solving the minimization problem in the same way as with FRK at point-level, i.e.,

$$
\min_{\boldsymbol{\lambda}} \quad \mathrm{E}\left\{\left(Y(B) - \boldsymbol{\lambda}'\tilde{\mathbf{Z}}\right)^2\right\} \qquad (4.37)
$$

Note that the vector $\boldsymbol{\lambda}$ has in the last entry an estimate for the Lagrange multiplier.

The estimation of $Y(B)$ (which is considered as a mean value over the cell B) is given by:

$$\hat{Y}(B) = \mathbf{T}(B)\hat{\boldsymbol{\alpha}} + \mathbf{k}(B)\tilde{\mathbf{Z}} \tag{4.38}$$

where

$$\mathbf{k}(B) = \mathbf{c}(B)\Sigma^{-1} \tag{4.39}$$

$$\mathbf{c}(B) = C(\tilde{\mathbf{Z}}(\mathbf{s}), \mathbf{Y}(B)) = \bar{\mathbf{S}}_p(B)\,\mathbf{K}\,\mathbf{S}(\mathbf{s})', \quad \bar{\mathbf{S}}_p(B) = \frac{1}{|B|}\int_{\mathbf{s}\in B}\mathbf{S}_p(\mathbf{s})d\mathbf{s} \tag{4.40}$$

$\bar{\mathbf{S}}_p(B)$ is the weighting matrix constructed for all prediction blocks, and $|B|$ is the number of points within the area B. Note that the difference between point-level kriging and block kriging is only the calculation of the weighting matrix $\bar{\mathbf{S}}_p(B)$ rather than $\mathbf{S}_p(\mathbf{s})$, while other parameters stay unchanged.
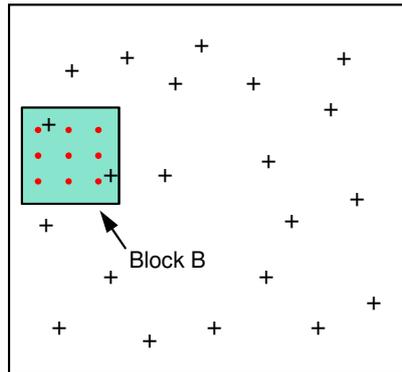
Figure 4.3: Kriging by block where the small black box shows the area at which the new value should be predicted. The plus signs denote the locations of the observed samples, while the red dots represent the regularly-spaced grid set up within the block.

In a similar way, OK can be applied over a block B. The weights are calculated as in eq. (4.13) by replacing the vector $\mathbf{c}_0$ with the vector $\mathbf{c}_0 = [\bar{\gamma}_{1B}, \ldots, \bar{\gamma}_{NB}, 1]'$, where

$$\bar{\gamma}_{iB} = \frac{1}{|B|}\sum_{j\in B}\gamma_{ij} \tag{4.41}$$

It remains to mention that the grid allocated within the block being estimated should be always regular. The spacing between points can be larger in one direction than the other, if the spatial continuity is anisotropic [Deutsch and Journel, 1998, pp. 107].

4.3 Estimating covariance model parameters

Predicting the stochastic component of the atmospheric signal using kriging requires the estimation of the covariance function and fitting a covariance model. Using the FRK covariance model in eq. (4.24), we need to estimate the matrix $\mathbf{K}$, the noise variance σ_ϵ^2, and the variance of the fine-scale signal σ_ζ^2. The first method proposed to find $\mathbf{K}$ is called binned method-of-moments (MM) [Cressie and Johannesson, 2008; Nguyen, 2009]. This approach derives the empirical estimator for Σ and finds $\mathbf{K}$ such that $||\hat{\Sigma} - \Sigma||_F$ is minimum, where $|| \cdot ||_F$ refers to the Frobenius norm. Another approach proposed by Katzfuss and Cressie [2009] targets to determine the covariance parameters using the algorithm of maximum likelihood estimation (MLE). Furthermore, they made the estimations using the expectation-maximization (E-M) algorithm to reduce the computational burden. This algorithm provides estimates not only of $\mathbf{K}$ but also of σ_ζ^2, where the solution of MLEs is found iteratively [Dempster et al., 1977]. Within each iteration the algorithm performs two steps, the expectation and maximization. In the following, we present a description on how to obtain the maximum likelihood estimates of the covariance model parameters via the E-M algorithm. In Section 7.4.2, we show the derivation of the MM method and how it is used to estimate the matrix $\mathbf{K}$.

Assuming that the observations in $\tilde{\mathbf{Z}}$ follow a multivariate Gaussian distribution, that is $\tilde{\mathbf{Z}} \sim N(0, \boldsymbol{\Sigma})$ with $\boldsymbol{\Sigma}$ of the form in (4.28). Let the parameters of interest $\mathbf{K}$ and σ_ζ^2 be summarized in the vector Θ, then the likelihood function $L(\Theta)$ [Katzfuss and Cressie, 2009]

$$
\begin{aligned}
-2 \log L(\Theta) &= -2 f(\tilde{\mathbf{Z}}; \Theta) \\
&= \log \det \boldsymbol{\Sigma} + \tilde{\mathbf{Z}}' \boldsymbol{\Sigma}^{-1} \tilde{\mathbf{Z}} + c \\
&= \log \det \boldsymbol{\Sigma} + \operatorname{tr}(\Sigma^{-1} \tilde{\mathbf{Z}} \tilde{\mathbf{Z}}') + c
\end{aligned}
\tag{4.42}
$$

where $c = (N/2) \log 2\pi$ is a constant independent of Θ and hence it cancels out in the maximization step. $\operatorname{tr}(\cdot)$ denotes the trace operator of a square matrix, with $\operatorname{tr}(A) = \sum_{i=1}^n a_{ii}$. In the expectation step of the algorithm, we calculate

$$
Q(\Theta; \Theta^{[t]}) = \mathrm{E}_{\Theta^{[t]}} \{-2 \log L(\boldsymbol{\eta}, \boldsymbol{\zeta}; \Theta) | \tilde{\mathbf{Z}}\}
\tag{4.43}
$$

given that:

$$
\begin{aligned}
-2 \log L(\boldsymbol{\eta}, \boldsymbol{\zeta}; \Theta) &= \log \det \mathbf{K} + \operatorname{tr}(\mathbf{K}^{-1} \boldsymbol{\eta} \boldsymbol{\eta}') + N \log \sigma_\zeta^2 + \sigma_\zeta^{-2} \operatorname{tr}(\boldsymbol{\zeta} \boldsymbol{\zeta}') \\
&\quad + N \log \sigma_\epsilon^2 + \sigma_\epsilon^{-2} \operatorname{tr}(\boldsymbol{\epsilon} \boldsymbol{\epsilon}')
\end{aligned}
$$

Then Eq. (4.43) becomes

$$
\begin{aligned}
Q(\Theta;\Theta^{[t]}) \;=\; & -\frac{1}{2}\Big[\log\det\mathbf{K} + \operatorname{tr}(\mathbf{K}^{-1}\mathrm{E}_{\Theta^{[t]}}\{\boldsymbol{\eta\eta}'|\tilde{\mathbf{Z}}\}) \\
& +N\log\sigma_\zeta^2 + \sigma_\zeta^{-2}\operatorname{tr}(\mathbf{V}_\zeta^{-1}\mathrm{E}_{\Theta^{[t]}}\{\boldsymbol{\zeta\zeta}'|\tilde{\mathbf{Z}}\}) \\
& +N\log\sigma_\epsilon^2 + \sigma_\epsilon^{-2}\operatorname{tr}(\mathbf{V}_\epsilon^{-1}\mathrm{E}_{\Theta^{[t]}}\{\boldsymbol{\epsilon\epsilon}'|\tilde{\mathbf{Z}}\})\Big]
\end{aligned} \tag{4.44}
$$

We should remind the reader that the parameters to be estimated here are $\mathbf{K}$ and σ_ζ^2, while σ_ϵ^2 is already estimated from the robust semivariogram. To proceed with the solution, it is required to quantify the conditional expectations in eq. (4.44). Using the standard formula required for calculating conditional expectations for multivariate normal distribution, given in Appendix A, the expectations will have the following form [Katzfuss and Cressie, 2009]

$$
\begin{aligned}
\mathrm{E}_{\Theta^{[t]}}\{\boldsymbol{\eta\eta}'|\tilde{\mathbf{Z}}\} &= \boldsymbol{\Sigma}_\eta^{[t]} + \boldsymbol{\mu}_\eta^{[t]}\boldsymbol{\mu}_\eta'^{[t]} \\
\mathrm{E}_{\Theta^{[t]}}\{\boldsymbol{\zeta\zeta}'|\tilde{\mathbf{Z}}\} &= \boldsymbol{\Sigma}_\zeta^{[t]} + \boldsymbol{\mu}_\zeta^{[t]}\boldsymbol{\mu}_\zeta'^{[t]}
\end{aligned}
$$

with

$$
\begin{aligned}
\boldsymbol{\mu}_\eta^{[t]} &= \mathrm{E}_{\Theta^{[t]}}\{\boldsymbol{\eta}|\tilde{\mathbf{Z}}\} = \mathbf{K}^{[t]}\mathbf{S}'\Sigma^{t-1}\tilde{\mathbf{Z}} \\
\boldsymbol{\mu}_\zeta^{[t]} &= \mathrm{E}_{\Theta^{[t]}}\{\boldsymbol{\zeta}|\tilde{\mathbf{Z}}\} = \sigma_\zeta^{2[t]}\mathbf{V}_\zeta\Sigma^{[t]-1}\tilde{\mathbf{Z}} \\
\boldsymbol{\Sigma}_\eta^{[t]} &= \operatorname{cov}_{\Theta^{[t]}}(\boldsymbol{\eta}|\tilde{\mathbf{Z}}) = \mathbf{K}^{[t]} - \mathbf{K}^{[t]}\mathbf{S}'\Sigma^{[t]-1}\mathbf{S}\mathbf{K}^{[t]} \\
\boldsymbol{\Sigma}_\zeta^{[t]} &= \operatorname{cov}_{\Theta^{[t]}}(\boldsymbol{\zeta}|\tilde{\mathbf{Z}}) = \sigma_\zeta^{2[t]}\mathbf{V}_\zeta - \sigma_\zeta^{2[t]}\mathbf{V}_\zeta\Sigma^{[t]-1}\sigma_\zeta^{2[t]}\mathbf{V}_\zeta
\end{aligned}
$$

After the expectation step, we perform a maximization step. The parameters $\mathbf{K}$ and σ_ζ^2 in Eq. (4.44) should be selected such that $Q(\cdot)$ is maximized. The partial derivative is taken with respect to both parameters and the result is assigned to zero. Finding the derivative here is rather simple since $\boldsymbol{\eta}$ and $\boldsymbol{\zeta}$ do not show dependency on each other, as observed from Eq. (4.44). The updating scheme of the E-M algorithm in each iteration is

$$
\mathbf{K}^{[t+1]} = \mathbf{K}^{[t]} + \mathbf{K}^{[t]}\left(\mathbf{S}'\Sigma^{[t]-1}\left(\tilde{\mathbf{Z}}\tilde{\mathbf{Z}}'\Sigma^{[t]-1} - \mathbf{I}_N\right)\mathbf{S}\right)\mathbf{K}^{[t]} \tag{4.45}
$$

$$
\sigma_\zeta^{2[t+1]} = \sigma_\zeta^{2[t]} + \sigma_\zeta^{2[t]}\operatorname{tr}\left(\frac{1}{N}\Sigma^{[t]-1}\left(\tilde{\mathbf{Z}}\tilde{\mathbf{Z}}'\Sigma^{[t]-1} - \mathbf{I}_N\right)\mathbf{V}_\zeta\right)\sigma_\zeta^{2[t]} \tag{4.46}
$$

We keep updating the solution until the algorithm converges. One criterion to monitor convergence is to calculate the norm of the difference between the current and last update of the vector $\boldsymbol{\Theta}$ (which is of size r^2+1). That means $||\boldsymbol{\Theta}^{[t+1]} - \boldsymbol{\Theta}^{[t]}|| < b$ should hold for small enough and positive value of b. Following Katzfuss and Cressie [2009], b is assigned a value of $10^{-6}r^2$. The starting choice of $\mathbf{K}$ and σ_ζ^2 should be valid; strictly speaking, $\mathbf{K}^{[0]}$ must be symmetric and positive-definite and $\sigma_\zeta^{2[0]}$ must be positive. Default values as suggested by Katzfuss and

Cressie [2009] are: $\mathbf{K}^{[0]} = 0.9 \cdot \text{var}(\tilde{\mathbf{Z}})\mathbf{I}_r$ and $\sigma_\zeta^{2[0]} = 0.1 \cdot \text{var}(\tilde{\mathbf{Z}})$.

The measurement error variance σ_ϵ^2 is estimated separately from the empirical semivariogram of the data. Estimating both σ_ϵ^2 and σ_ζ^2 from the data is not a trivial task. That is because the nugget effect in the semivariogram reflects not only the error variance but also the fine-scale variance. Therefore, having information about the error distribution and variance is worthwhile. In our case we estimate the sum $\sigma_\epsilon^2 + \sigma_\zeta^2$ using the method of robust semivariogram, eq. (4.21).

We estimate the semivariogram as a function of the separation distance $\mathbf{h}$. In order to obtain values of σ_ϵ^2 and σ_ζ^2, we follow the approach presented in [Kang et al., 2010]. To obtain an estimate of σ_ϵ^2, a strait line is fitted to the estimated semivariogram at short h. Since the slope of the structure function (variogram) describing atmospheric turbulence is expected to vary with h [Treuhaft and Lanyi, 1987], we made the line fitting based on the estimates of the first 3 km (empirically defined). Let the line fit be $\hat{\gamma}(h) = \hat{\gamma}(0+) + bh$, then the estimate of σ_ϵ^2 is

$$\hat{\sigma}_\epsilon^2 = \hat{\gamma}(0+) \tag{4.47}$$

Should $\hat{\gamma}(0+)$ have a negative value, $\hat{\sigma}_\epsilon^2$ is set to zero. σ_ζ^2 is estimated from the 1-lag semivariance after removing the noise effect, i.e.,

$$\hat{\sigma}_\zeta^2 = \frac{1}{2|N(\mathbf{h}_1)|} \sum_{|N(\mathbf{h}_1)|} \left\{ (Y(\mathbf{u}_i) - Y(\mathbf{u}_j))^2 - \hat{\sigma}_\epsilon^2(v_\epsilon(\mathbf{u}_i + v_\epsilon(\mathbf{u}_j)) \right\} \tag{4.48}$$

where 1-lag distance $\mathbf{h}_1$ equals the minimum distance separating two pixels. The noise was assumed to have a covariance function of the from $\sigma_\epsilon^2 \mathbf{V}_\epsilon$ with $\mathbf{V}_\epsilon = \text{diag}[v_\epsilon(\mathbf{u}_1), \ldots, v_\epsilon(\mathbf{u}_j)]$. We assumed $\mathbf{V}_\epsilon$ to be an identity matrix.

4.4 A case study

In this section, we present an example of spatial data prediction by applying the algorithms of FRK and OK to single data sets. SWD fields derived in Chapter 6 are used as input to the two algorithms. The input data are at point-level support, where the spatial density of the measurements varies over the region acquired by the radar. The density is high in urban areas; however, it decreases dramatically in forests and agricultural regions, which results in the gaps observed in **Figure 4.4**. As mentioned before, FRK and OK are able to infer SWD values on a regular grid in spite of the lack of observations in some areas. That is possible since these methods learn the spatial properties of the signal and fit a proper covariance model prior to prediction.

We perform predictions of SWD on a continuous field using PSI observations. Predictions

are achieved at point-level and areal-level. We do not have concrete information about the accuracy of the master and slave atmospheric components and the error term remains after the PSI data processing. It is necessary to explore the statistical properties of atmospheric data received by processing InSAR and GNSS data. We use the method of FRK as a geostatistical prediction technique that inspects the spatial dependence of the data, the fine-scale variance, and the error variance prior to prediction. As observed from eq. (4.32), the parameters we need to estimate are $\{\boldsymbol{\alpha}, \sigma_\epsilon^2, \sigma_\zeta^2, \mathbf{K}\}$. $\boldsymbol{\alpha}$ is estimated from the absolute SWD observations as shown in Eq. (4.10). A linear trend is then calculated and subtracted from the data as the other parameters should be estimated from the detrended data (zero mean).

We estimate the parameter $\boldsymbol{\alpha}$ using eq. (4.10) and compute the long-wavelength component. The observations are detrended by subtracting the estimated component. **Figure 4.4** shows the zero-mean SWD field (05.23.2005, 9:51 am). The field contains 76841 scattered observations within an area of about 100×100 km^2. We can see the low density of observations in the area between longitude lines $7°30'$ and $8°06'$, and in the south-eastern part. For the sake of comparison with the SWD fields obtained from WRF, we predict at a regular grid of 3×3 km^2. Because the input data are at point-level support and the output data have block-level support, we should solve point-to-area COSP. Therefore, we apply block kriging (OK and FRK) to derive SWD fields at the required spatial resolution.

To proceed, first with the OK algorithm, we should compute the empirical semivariogram and fit a theoretical model (exponential function for this example). The weights are computed based on the semivariogram model and the predictions are then derived by applying eq. (4.5). **Figure 4.7(a, b)** shows the SWD inferred on a regular grid of resolution 3×3 km^2 using OK and the corresponding MSPE maps.

Next, we perform prediction using the FRK algorithm. At the beginning, we estimate a
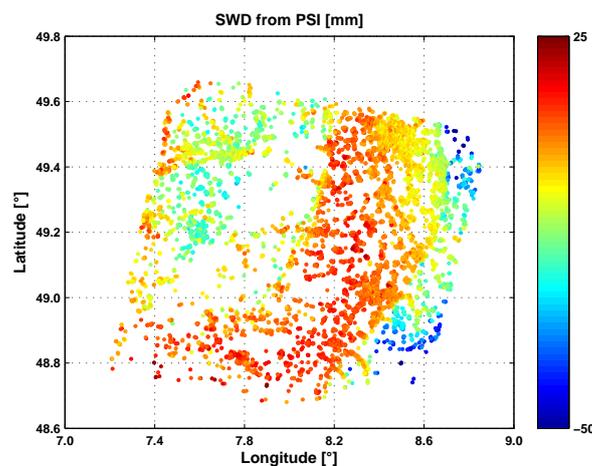
Figure 4.4: SWD observations derived in Chapter 6, where the long-wavelength component is subtracted. Forest areas in the middle and the south east contain almost no PS points. Acquisition date: 05.23.2005

robust semivariogram using eq. (4.21). By line fitting, the error variance σ_ϵ^2 is estimated from the robust semivariogram method, eq. (4.47). We specify a set of nodes at different spatial resolutions such that they cover the entire observation domain. The number of nodes is $r = 93$, and to capture spatial variations on several scales we distributed the nodes at three spatial resolutions. In the first group, 4 nodes are put at a resolution of about 40 km, while the resolution of the second group is 20 km and there exist 16 nodes. The last group contains 64 nodes at spatial spacing of 10 km. The nodes represent the center point of the BFs required to encode all locations in the observation domain. We use the bisquare BF in eq. (7.34) to assign weights to each observation and prediction location and form the matrix $\mathbf{S}(\mathbf{s})$ and $\bar{\mathbf{S}}_p(\mathbf{s})$. The set-up of the nodes is depicted in **Figure 4.5**. We then estimate $\mathbf{K}$, and σ_ζ^2 using the E-M algorithm as described in Section 4.3.
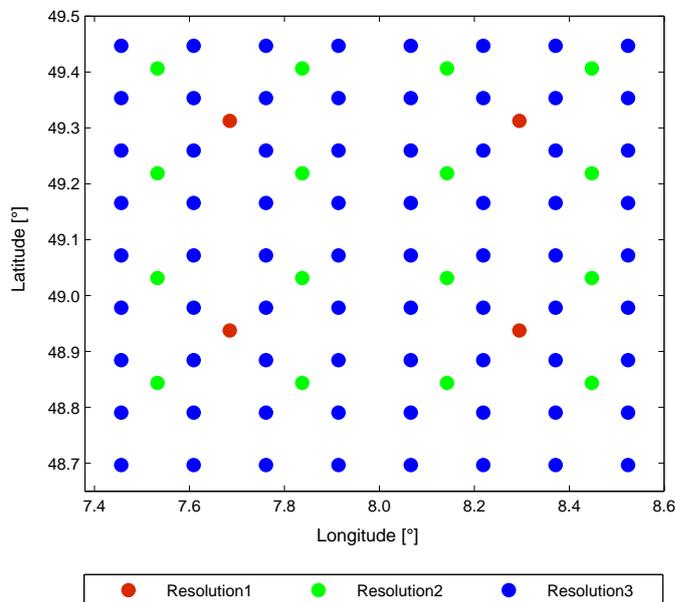
Figure 4.5: FRK nodes or center locations of 93 basis functions at three spatial resolutions. The first resolution is 40 km, the second resolution is 20 km, and the third resolution is 10 km.

The result of the FRK predictions and MSPEs are exhibited in **Figure 4.7** (c, d). WRF data are shown in **Figure 4.6**, in which the IWV values are mapped to SWD values using eq. (2.22). The long-wavelength component is reduced to achieve a proper comparison. From the plots in **Figure 4.7**, we observe similar results from both OK and FRK that agree with the original WRF data. The predictions based on FRK agree with those obtained using the OK method, and the correlation coefficients with the original data of WRF are approximately 84% and 83% for FRK and OK, respectively. The MSPE plots in **Figure 4.7** (b, d) show that the error increases in areas where observations are sparse. In general, the MSPE values of the FRK method are smaller compared with those of the OK method. This ensures that the weights found based on the covariance model of the FRK method are more appropriate.
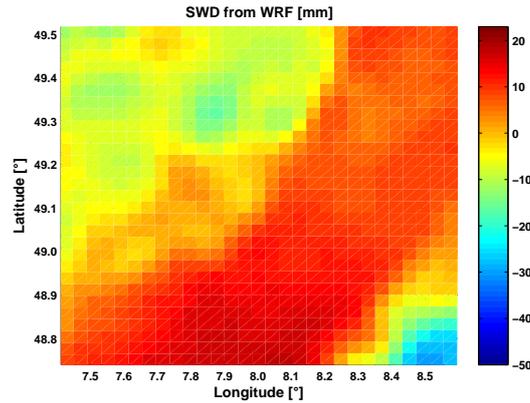
Figure 4.6: Detrended SWD maps received from WRF. The resolution of the grid is $3{\times}3$ km^2.

In the lower right corner of **Figure 4.7** (b), we notice larger uncertainty in OK predictions compared with FRK values in (d) and the penalty of single 'outliers' when the observations are sparse (inspect the change from red to blue in the lower part of the figure).

The most impressive point here is the computational time reported for both algorithms. The FRK algorithm is fast so that it requires a short time to compute the predictions. Most of the time is invested in the calculations of the covariance model parameters and constructing the matrices **S** and $\boldsymbol{\Sigma}$. On the other hand, we implement the OK algorithm so that the predictions are found iteratively to reduce computational time. We also consider the observations which are located within a predefined radius. Nevertheless, the algorithm requires computational time with an order of magnitude higher than that required by the FRK method to obtain predictions on the same grid, using the same machine. The computational time should rise proportional to the growth of prediction locations. If we apply the traditional OK, where $\boldsymbol{\Sigma}^{-1}$ is constructed for N observations, the computational time grows significantly.

4.5 Summary

In this chapter, we presented the spatial interpolation and estimation techniques used in the study. Due to lack of SWD (or IWV) fields on contentious grids and the different spatial resolutions of the data sets, it is required to apply spatial interpolations to bring the data into a common basis. We used moving-window averaging, IDW and surface fitting as non-geostatistical interpolation methods, which obtain interpolations within the data coverage. Due to the presence of gaps in SWD fields from, derived for example from PSI, we used geostatistical interpolators, namely kriging, to get predictions on any predefined grid. These interpolators are based on the observation that the variability of a specific spatial process should have a particular structure modeled by a variogram. We used OK with the assumptions of isotropy and stationary of the SWD fields. Generally, these assumptions do not hold for geophysical quantities, Moreover, the OK requires high computational costs and as the size of

(a)

(b)

(c)

(d)
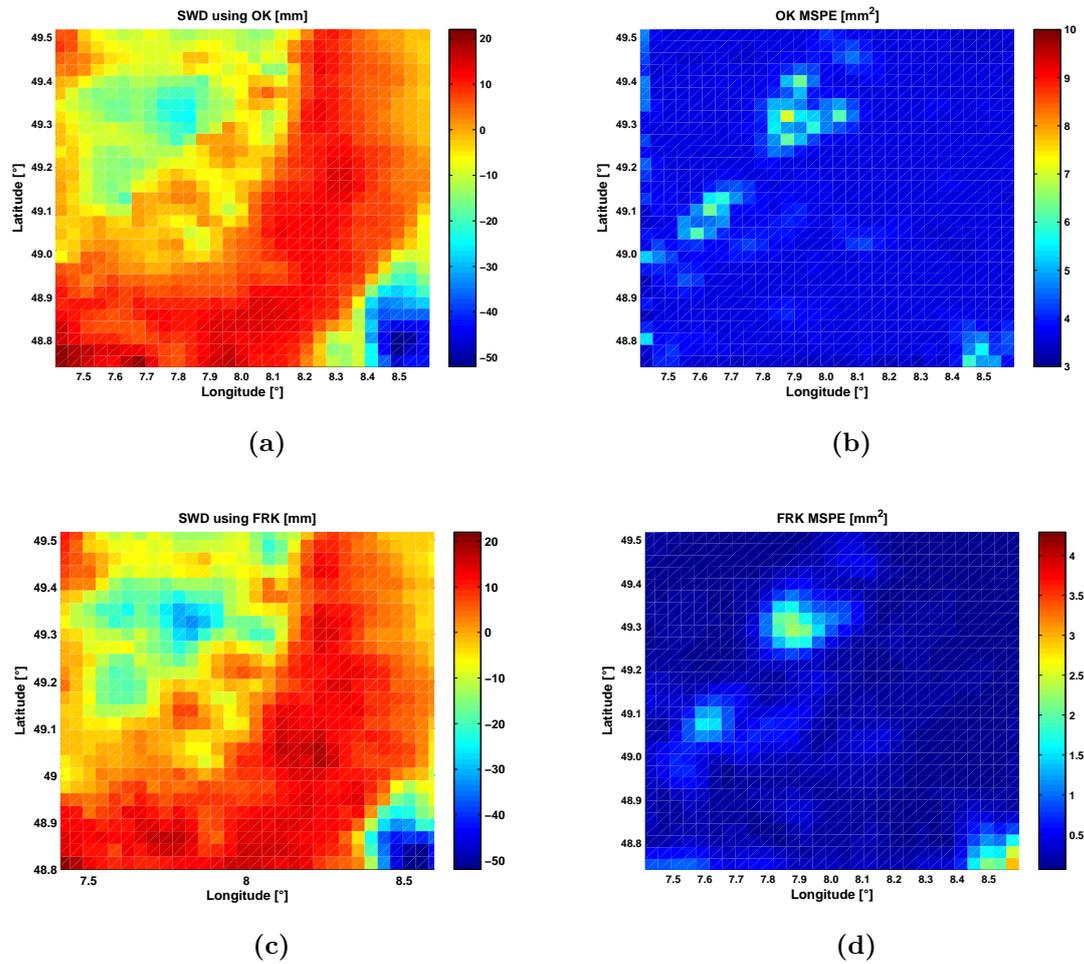
Figure 4.7: In (a) is shown the prediction map using the block OK method, and in (b) are the corresponding MSPEs. Predictions of the SWD by applying the block FRK algorithm are shown in (c) and the corresponding MSPEs are shown (d). The resolution of the grid is 3×3 km^2. A point-level SWD map, on 05.23.2005 at 09:51 UTC, is used as input to the algorithms.

the observations exceeds a certain limit, OK becomes non-feasible. To overcome the limitations of OK, the method of FRK was applied. Using FRK, it is also possible to solve change of support problems. The FRK method is the basis for spatial data fusion approach presented in Chapter 7.

5 Remote sensing of water vapor

Our primary objective in this work is to derive maps of the atmospheric wet delay by processing InSAR phase observations and to reconstruct the absolute wet delay signal (and hence IWV maps) using supporting data from GNSS. These maps are then assimilated with wet delay maps generated from the WRF model to test the influence of data fusion on the quality of the output maps. **Figure 5.1** shows a setup of the GNSS and SAR systems used for remote sensing of water vapor. We described in Chapter 2 the influence of the atmosphere, in particular the neutrosphere, on the propagation of microwave signals. The neutrospheric delay is considered as the major source of error since it cannot be eliminated by linear combination of multiple frequencies. Different approaches have been developed to model and substantially reduce this error, as presented in Chapter 1. Both GNSS and InSAR are affected by the neutrosphere in a similar manner [Williams et al., 1998]. GNSS have been exploited since the 1990's for water vapor sounding; yet, this is a quite new research field for InSAR. In this chapter, we describe the methods to obtain the neutrospheric delay from GNSS and InSAR observations, such that we can investigate this "error" signal for measuring the content of water vapor in the neutrosphere. Since phase observations from GNSS and InSAR are a superposition of different contributions, our first goal is to separate the neutrospheric delay from other phase components. We present in Section 5.1 the methods used to estimate the neutrospheric delay from GNSS observations. In Section 5.2, we describe the phase measured by InSAR and the contribution of the neutrosphere in the interferograms. We explain the strategy to extract the contribution of the neutrosphere by applying PSI. Furthermore, we present comparisons of our results with IWV maps from MERIS and WRF as well as measures of the spatial structure of the SWD maps by means of the power spectral density and structure function calculations.

5.1 Neutrospheric delay in GNSS phase observations

GNSS, mainly GPS, have been used as a data source for a number highly precise geodetic applications. Because GNSS signals are delayed when penetrating the Earth's atmosphere, one of the main tasks of the processing software is to successfully estimate and remove the atmospheric effect from the signal. Both the ionosphere and the neutrosphere introduce time delay into the signal and there exist a variety of techniques to eliminate them. While the ionosphere is dispersive and the ionospheric delay can be effectively reduced by linear combination of multiple frequencies, the neutrosphere is non-dispersive for L-band frequencies and the its
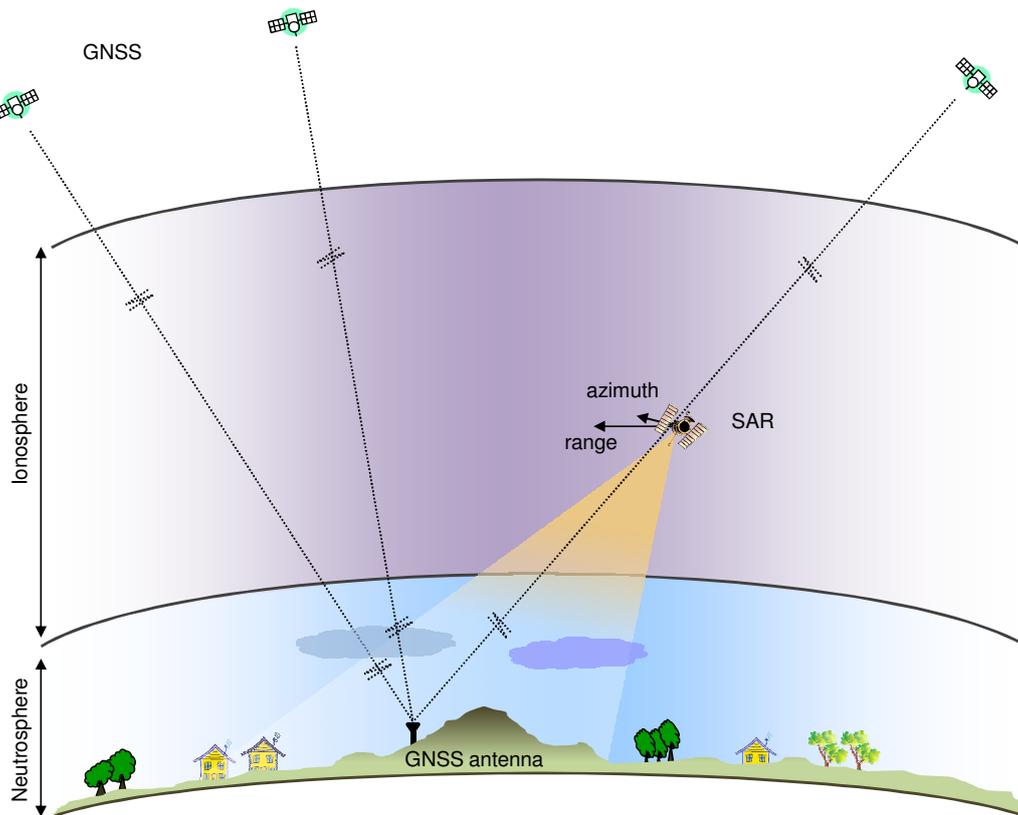
Figure 5.1: Schematic diagram illustrating the geometry and the spatial coverage of GNSS and (In)SAR. GNSS satellites fly at altitudes of 20,000 km, while the SAR flies at 700 km in a direction perpendicular to the page (azimuth) and captures a swath of width 100 km (in range). GNSS measurements are received at single sites distributed within the SAR image on the ground.

delay has to be modeled. Bevis et al. [1992] suggested the investigation of the signal propagation error to characterize atmospheric parameters such as the water vapor content. Since then, different studies have been carried out to estimate water vapor content in the atmosphere using GNSS signals, for example [Bender et al., 2008; Karabatić et al., 2011; Luo et al., 2008]. The growing number of GNSS constellations, such as GPS, GLONASS, and GALILEO, offer promising possibilities to determine several atmospheric parameters at a satisfying quality level. We focus on the neutrospheric delay derived during GNSS data processing as a valuable source to determine the atmospheric water vapor content and to analyze its temporal and spatial properties.

5.1.1 Precise Point Positioning and Differential GNSS

GNSS data processing can be achieved either by Differential Phase GNSS (DPGNSS) or PPP. The method of DPGNSS is based on using a reference GNSS receiver with a known position [Misra and Enge, 2001, p. 49]. Since its position is known, the receiver can provide esti-

mates of the error signals that can be used by close receivers to mitigate the errors in their measurements.

On the other hand, the concept of Precise Point Positioning (PPP) was first introduced in the 1970's, and it has been substantially used since the end of the 1990's when accurate orbit and satellite clock data were made available [Zumberge et al., 1997]. The method of PPP utilizes a linear combination of L1 and L2 frequencies to remove the first order component of the ionospheric delay, which is an excellent approximation as mentioned in [Dach et al., 2007, p. 257]. Using a linear combination of dual frequencies succeeds to effectively reduce the ionospheric delay, but inaccurate information about satellite orbits and clocks can extremely worsen the position estimates. Therefore, the accuracy can be significantly improved by exploiting highly precise satellite orbits and clock information available from the International GNSS Service (IGS) [`http://igscb.jpl.nasa.gov/`]. The remaining propagation error is caused mainly by the neutrosphere. PPP has the following benefits over DPGNSS:

- Only one GNSS receiver
  Unlike the DPGNSS, which requires simultaneous measurements at two or more GNSS sites, the PPP strategy utilizes observations from only one GNSS site. Hence, no simultaneous measurements are necessary, which saves time and costs, and there is no need to examine the limitations on the length of the spatial baselines between sites.

- No correlation between the processed sites
  In the method of DPGNSS, the results at two or more sites are correlated because differences are built with respect to a reference site. If the signal of the reference site is inaccurate, the results in the whole network are affected. This source of failure is avoided when using the method of PPP.

- Absolute measurements of neutrospheric delay
  The most important point about using the PPP method in this work is that it delivers absolute measurements of the neutrospheric delay, while the DPGNSS produces relative measurements of the neutrospheric delay at short baselines.

On the other hand, there are some disadvantages behind using PPP:

- Linear combination of L1 and L2 frequencies
  To eliminate the ionospheric delay, a linear combination of L1 and L2 observations must be applied. Because the linear combination removes the first order ionospheric delay, the remaining error from higher order components may affect the accuracy of the estimates of the neutrospheric delay. Also, under the use of linear combinations, the phase ambiguity is no longer integer number.

- Dependency on the quality of the input data
  The accuracy of the estimated parameters is dependent of the input data of satellite orbits, as well as satellite and antenna clock information. Highly precise orbit and clock data are available several days after the observations; accordingly, the accuracy of the

PPP estimations in real-time is not adequate for the state-of-the-art research.

- Accuracy and observations time window
  In order to achieve the same accuracy as DPGNSS, observations over longer time windows
  are required for PPP processing.

By revealing tradeoffs, uncertainties, and requirements, we made the decision to use data
processed by the PPP strategy, since we are most interested in absolute measurements of the
neutrospheric delay. The aforementioned disadvantages of PPP can still be addressed to attain
a satisfying accuracy. For the current work, real-time data processing is not essential.

A GNSS satellite transmits coded messages modulated onto the carrier signals at L1 (1.57542
GHz) or L2 (1.2276 GHz) frequencies. The carrier phase at L1 and L2 frequencies is given by
[Misra and Enge, 2001, pp. 166]:

$$\phi_r^{gnss,k} = \lambda^{-1}\rho_r^k + \frac{c}{\lambda}(\delta t_r - \delta t^k) - \lambda^{-1}\Delta L_{r,iono}^{gnss} + \lambda^{-1}\Delta L_{r,neu}^{gnss} + N_r^k + \epsilon_r^{gnss} \qquad (5.1)$$

$\phi_r^{gnss,k}$:   Carrier phase measured by the receiver r, from a signal transmitted from the kth GNSS satellite, in units of cycles

ρ_r^k: Geometrical distance from the kth satellite to the receiver r

$\delta t^k, \delta t_r$:   Clock error of the kth satellite and the receiver clock error, respectively

ΔL_{iono}^{gnss}:   Delay, in meters, due to the propagation in the ionosphere

ΔL_{neu}^{gnss}:   Delay, in meters, due to the propagation in the neutrosphere

ϵ_r^{gnss}:   Phase noise, which includes orbit inaccuracies, multipath, and antenna phase center variations

λ: Wavelength of the signal; 19 cm at L1 and 24 cm at L2

N_r^k: Integer number ambiguity in the signal transmitted from the kth satellite

c: Speed of light in vacuum, 2.99792458×10^8 m/s

The phase measurement is converted into a path length in units of meters using the scaling
factor $\frac{\lambda}{2\pi}$. For this work, the interesting term in eq. (5.1) is the delay due to the neutrosphere.

Let the satellite coordinates be given by $[X_s, Y_s, Z_s]$ and the receiver coordinates are $[X_r, Y_r, Z_r]$,
then ρ_r^k is computed from

$$\rho_r^k = \sqrt{(X_r - X_s^k)^2 + (Y_r - Y_s^k)^2 + (Z_r - Z_s^k)^2} \qquad (5.2)$$

The most popular PPP-related linear combination of the carrier phase of L1 and L2 frequencies

is achieved by the following relation

$$\phi_{IF} = \frac{1}{f_1^2 - f_2^2}(f_1^2 \phi_{L1} - f_2^2 \phi_{L2}) \tag{5.3}$$

The ionosphere-free (IF) combination of the carrier phase measurements, $\phi_{r,IF}^{gnss,k}$, reads

$$\phi_{r,IF}^{gnss,k} = \lambda^{-1}\rho_r^k - \frac{c}{\lambda}\delta t_r + \lambda^{-1}\Delta L_{r,neu}^{gnss} + N_{r,IF}^k + \epsilon_r^{gnss} \tag{5.4}$$

$N_{r,IF}^k$:   Ionosphere-free ambiguity, no longer integer

GNSS observations (code and phase) are processed using the BS5 software `http://www.bernese.unibe.ch/` [Dach et al., 2007], where the code measurements are required for the preprocessing step. In the presence of satellite orbits and clock information, GNSS data are processed to estimate: the three location coordinates of the GNSS receiver (antenna), the receiver clock error, the phase ambiguity, and a so-called site specific neutrospheric parameter (SSNP). The estimation of the parameters is achieved by the weighted least squares inversion. For detailed description of GNSS data processing for this work, the reader is referred to [Fuhrmann et al., 2010, Chapter 5].

5.1.2 Estimating the neutrospheric delay

The azimuthally-isotropic neutrospheric zenith total delay (ZTD) caused by the column of air above the GNSS receiver is determined in the BS5 from the following formula:

$$ZTD_{iso}^{gnss} = \Delta^{model} + \Delta^z \tag{5.5}$$

ZTD_{iso}^{gnss}:   Azimuthally-isotropic neutrospheric zenith total delay from GNSS observations in [meters]

Δ^{model}:   neutrospheric delay in [meters], obtained from Saastamoinen model

Δ^z: zenith-directed SSNP (Correction term) in [meters]

Note that Δ^{model} can slightly differ from the value of the ZDD. As we mentioned in Section 2.2.2, the ZDD is accurately determined from empirical neutrospheric a priori models; therefore, it is calculated in the BS5 by substituting representative meteorological parameters from the standard atmosphere in eq. (2.21). Note that by using representative values from the standard atmosphere, the calculated ZDD will remain constant over time since it depends only on the altitude and latitude of GNSS site, eq. (2.15).

In order to make the estimated ZTD more realistic, the BS5 estimates a correction term (SSNP), which is more variable over time. One value for the SSNP is estimated every hour such that for estimating the SSNP at 02:00 (UTC), the observations in the interval from 01:30-02:30 (UTC) are used, and the estimated SSNP value is valid for that interval. A large part of the SSNP corresponds to the wet delay, but it also contains a small fraction of the dry delay. The accuracy of the estimated receiver position as well as the SSNP is affected by different factors, for example, the weighting of the observations and correcting the site-specific multipath effects [Luo et al., 2008]. **Figure 5.2** shows time series obtained from the BS5 for the modeled ZDD as well as the sum of the ZDD and the SSNP at the site KARL on 07.12.2004 (see **Figure 3.4** for site location and altitude).
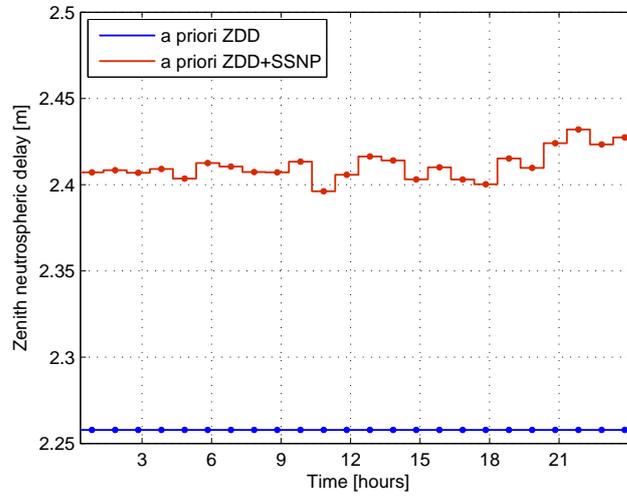
Figure 5.2: The modeled ZDD and the corresponding correction term (SSNP) as received from BS5, site: KARL (49.0112° N,8.4113° E), day: 07.12.2004. The values are estimated at times (UTC): [00:51, 01:51,..., 23:51], and the SAR overpass time is 09:51.

A GNSS receiver records the signals from all visible satellites with elevation angles above the cut-off elevation (7° for this work). The estimated ZTD measures the average effect of the conical neutrosphere above the GNSS receiver. Since most of the water vapor is located near the Earth's surface, we can approximate the radius of the cone. Referring to the schematic diagram shown in **Figure 5.3**, if the minimum cut-off elevation (θ_{min}) is 7° and assuming that water vapor is concentrated in the lower 1 km of the atmosphere, the corresponding cone radius is $r_c \approx 8$ km. The ZTD estimated in the BS5 software contains the delay effects of the azimuthally-isotropic neutrosphere. Estimating the path delay along the satellite LOS is however more challenging as it requires to account for the anisotropy of the neutrosphere. The neutrospheric delay along the satellite LOS, called the slant total delay (STD), is estimated as a sum of four components: the ZDD, the SSNP, the neutrospheric horizontal gradients, and the related phase residuals after least squares, i.e.,

$$STD^{gnss} = \Delta^{model} \cdot \mathrm{MF}_{dry} + \Delta^z \cdot \mathrm{MF}_{wet} + \Delta^n \cdot \frac{\partial \mathrm{MF}_{wet}}{\partial z} \cdot \cos A + \Delta^e \cdot \frac{\partial \mathrm{MF}_{wet}}{\partial z} \cdot \sin A + \Delta^r \quad (5.6)$$
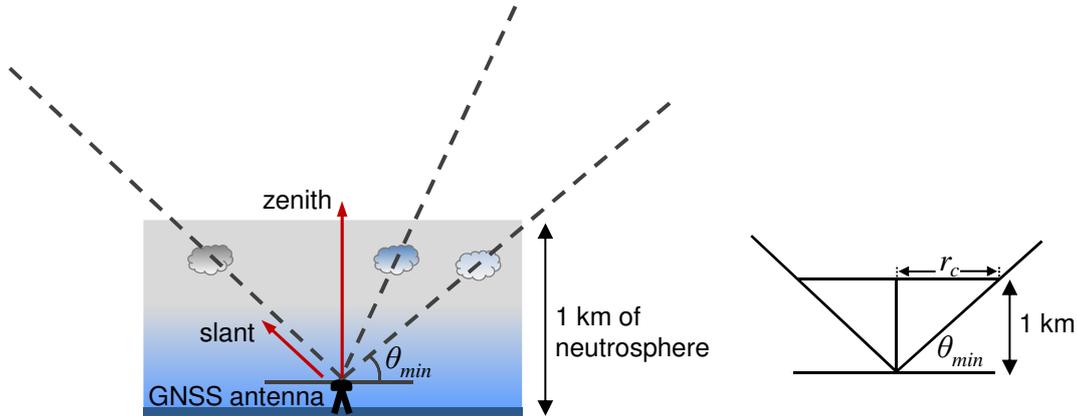
Figure 5.3: A GNSS antenna receives signals from different satellites at a minimum elevation angle (θ_{min}) that defines a cone-like neutrospheric section above the GNSS receiver. Assuming water vapor resides in the lower 1 km of the neutrosphere, $r_c \approx 8$ km for θ_{min} of 7°.

STD^{gnss}:   Neutrospheric slant total delay from GNSS observations in [meters]

Δ^n, Δ^e: Northing and easting neutrospheric gradients, respectively

Δ^r: Anisotropic wet component extracted from the phase residuals in [meters]

A, z: Satellite azimuth and zenith angles

MF: Mapping function (here Niell MF)

The BS5 provides estimates of the northing and easting neutrospheric gradients to account for the neutrospheric tilting [MacMillan, 1995]. This tilting means that zenith direction of the neutrosphere (with minimum neutrospheric delay) can slightly mismatch the ellipsoidal zenith. The neutrospheric gradients have an influence at low elevation angles, while they cancel out at 90° elevation. The neutrospheric gradients depict insignificant variations over one day, hence they are estimated in the BS5 once every 24 hours.

A more significant component to account for the neutrospheric anisotropy is the component of the wet delay extracted using the phase residuals resulting from computing the difference between the estimated GNSS solution and the measurements. The residuals contain, however, other error sources due to multipath (MP) and antenna phase center variations (PCV). Fuhrmann et al. [2010] proposed a spatial stacking approach to compensate errors due to MP and nonmodeled antenna PCV and extract the wet delay component. After filtering out MP and antenna PCV errors, the azimuthally-anisotropic neutrospheric delay Δ^r is retrieved and added to the other estimated components to produce the entire STD.

5.1.3 Quantifying the zenith wet delay

The ZDD is computed in the BS5 using representative meteorological parameters from the standard atmosphere. A fraction of the dry delay depends on the weather conditions, which we classified as a heterogeneous component in Section 2.2.4. Because the SSNP is a correction term that is calculated for the total neutrospheric delay, it is not exactly the wet delay, but it also contains the heterogeneous fraction of the dry delay. In order to retrieve the wet delay, we calculate a value for the ZDD by substituting meteorological observations in eq. (2.15), and subtract it from the ZTD, i.e.,

$$ZWD_{iso}^{gnss} = ZTD_{iso}^{gnss} - ZDD^{meteo}$$
$$= \Delta^{model} + \Delta^z - ZDD^{meteo} \tag{5.7}$$

ZWD_{iso}^{gnss}:   Azimuthally-isotropic zenith wet delay from GNSS observations in [meters]

ZDD^{meteo}:   Zenith dry delay in [meters], obtained from Saastamoinen model using meteorological data

This approach has been exploited in different studies [Fuhrmann et al., 2010; Jade and Vijayan, 2008; Jin and Luo, 2009; Karabatić et al., 2011]. We compute a value for the dry delay from the Saastamoinen model by substituting the values of P, T, and e from surface meteorological observations of 30-60 minutes sampling time. When the meteorological data are not available at the acquisition time of the remote sensing data, temporal interpolations of meteorological data are done by applying cubic spline interpolators, while spatial interpolations are made by kriging as presented in Section 5.1.5. In addition, we use the meteorological data from the WRF model, which are available at a 10 minutes temporal resolution and a spatial resolution of $3{\times}3$ km^2. WRF surface temperature data are mainly used to calculate the weighted mean temperature of the atmosphere, which is required for calculating the conversion factor Pi (Section 2.3). Once the ZWD component is calculated, it is mapped to the LOS of the satellite to add the anisotropic component of the wet delay, Hence, the SWD is:

$$SWD^{gnss} = ZWD_{iso}^{gnss} \cdot \text{MF}_{wet} + \Delta^n \cdot \frac{\partial \text{MF}_{wet}}{\partial z} \cdot \cos A + \Delta^e \cdot \frac{\partial \text{MF}_{wet}}{\partial z} \cdot \sin A + \Delta^r \tag{5.8}$$

The SWD is then mapped back to the vertical direction to obtain the ZWD that contains the azimuthally-isotropic and anisotropic water vapor components as follow:

$$ZWD^{gnss} = \frac{SWD^{gnss}}{\text{MF}_{wet}} \tag{5.9}$$

The procedure followed to determine the SWD from the GNSS observations are summarized
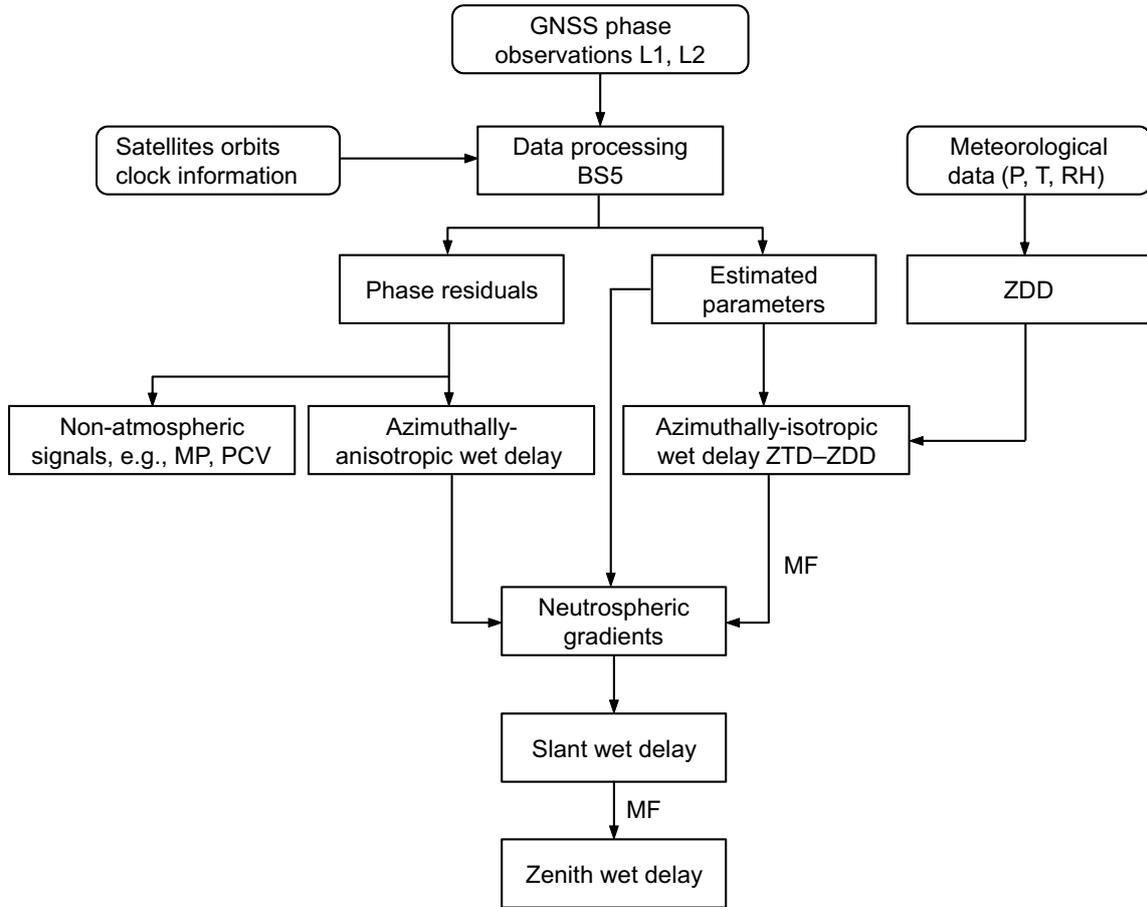
Figure 5.4: A flowchart summarizing the procedures of retrieving the ZWD from GNSS phase observations based on the PPP strategy. Meteorological observations are necessary for accurate separation of the wet delay.

in **Figure 5.4**. It remains to describe how the ZDD is calculated from the Saastamoinen model (eq. (2.15)) using observations of the air pressure, temperature, and relative humidity at the GNSS sites. The relative humidity is necessary to determine the partial pressure of water vapor. Some stations provide measurements of the dew point temperature, which can be related to the relative humidity as presented by Lawrence [2005] using the following formula:

$$rh \approx 100 - 5(T - T_d) \qquad [\%] \tag{5.10}$$

T: Air temperature

T_d: dew point temperature, and both T, T_d are in degrees Celsius.

The partial pressure of water vapor is determined using meteorological data from one of the

following empirical expressions:

$$e = 6.1162 \cdot \exp\left(\frac{17.67\,T_d}{T_d + 243.5}\right), \quad T_d \text{ in } [^\circ\text{C}] \tag{5.11}$$

$$e = \frac{rh}{100} \cdot \exp\left(-37.2465 + 0.2131665T - 0.000256908T^2\right), \quad T \text{ in } [\text{K}] \tag{5.12}$$

Since these data are often not available exactly at the site, we have to interpolate or extrapolate horizontally and vertically. If the meteorological station is located adjacent to the GNSS site, the air pressure at the meteorology station P^{MET} is used to extrapolate the air pressure at the altitude of the GNSS site (P^{gnss}) using a form of the hydrostatic equations [Barry and Chorley, 1987; Byers, 1974]:

$$P^{gnss} = P^{MET} \left(\frac{T_{meteo} - L(z_{gnss} - z_{meteo})}{T_{meteo}}\right)^{g/RL} \tag{5.13}$$

T_{meteo}:   Air temperature at the meteorology station [K]

z_{gnss}:   Ellipsoidal altitude of the GNSS site (antenna) [m]

z_{meteo}:   Ellipsoidal altitude of the Meteorology station [m]

L: Temperature lapse rate ≈ 0.0065 [K/m] [Barry and Chorley, 1987, pp. 56]

R: Universal gas constant ≈ 8.31447 [J/mol K]

g: Earth's gravitational acceleration ≈ 9.80665 [m/s^2]

In a similar way, the partial pressure of water vapor is computed at the GNSS site using eq. (5.13). Given the air pressure and the partial pressure of water vapor at the GNSS site, the ZDD can be calculated using eq. (2.15). **Figure 5.5** shows 24-hours time series of the ZTD, ZDD, and ZWD quantified at 10 sites from GNSS observations on the days 06.27.2005 (summer) and 01.03.2005 (winter). The ellipsoidal altitudes of the GNSS sites are shown in **Figure 5.6** and the geographic locations of the sites are provided in **Table 3.2**. We can observe that the higher the GNSS site is, the lower the neutrospheric delay becomes. The ZTD values in summer are higher than the ZTD values in winter; this increase is resulting from the ZWD component since the air is more humid and the amount of water vapor in the neutrosphere is larger in summer. The ZWD shows higher temporal variations compared with the ZDD. If we examine the ZDD, we notice a small increase in the winter values. This can be explained by cold air temperature in winter that is associated with high air pressure, while in summer warm air is more buoyant and has a lower pressure. Also, in winter the air pressure increases while the partial pressure of water vapor decreases, then the pressure of dry gases ($P_d = P - e$) and hence the ZDD will have greater values.
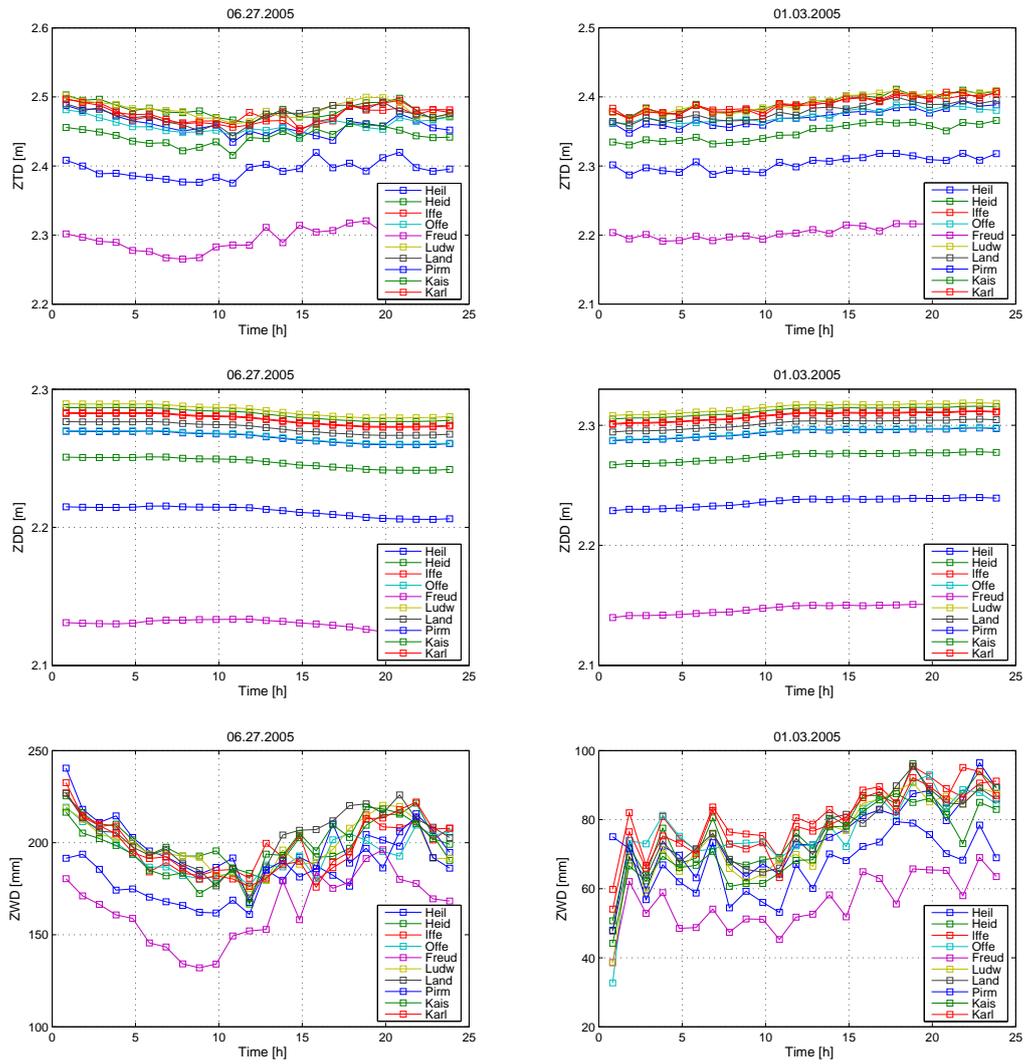
Figure 5.5: 24-hours time series of ZTD, ZDD, and ZWD from GNSS observations at 10 sites on the days 06.27.2005 and 01.03.2005. The exact geographic locations of the sites are given in **Table 3.2**, and their heights are shown in **Figure 5.6**.
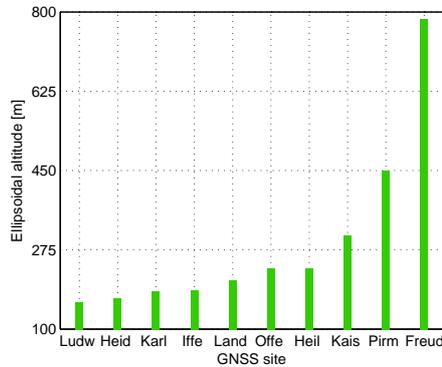
Figure 5.6: Ellipsoidal heights of the GNSS sites.

5.1.4 Comparing ZWD derived from GNSS, MERIS, and WRF data

For the sake of verifying the ZWD estimated from GNSS observations, we use data from MERIS and WRF. MERIS provides maps of IWV, which are mapped into ZWD values to be compared with the absolute values of the ZWD measurements retrieved from GNSS. Because of the aforementioned cone-averaging of GNSS observations, ZWD measurements from MERIS are averaged within 15 km by 15 km blocks centered at the location of the GNSS site (**Figure 5.7**). **Figure 5.8** (a) shows the ZWD from MERIS against the ZWD measured at 10 GNSS receivers on 5 days, for which the cloud-free MERIS maps were available. We observe strong correlation between the ZWD values of GNSS and MERIS for all days. We also compute the difference of ZWD values as depicted in **Figure 5.8** (b), and the root mean squared (RMS) values of the differences are given in **Table 5.1**. The results show low RMS values except for the observations on the day 10.30.2006. We notice that the differences are large at some receivers compared to the others, which is explained by the partial presence of clouds in the area close to those receivers. Because the data are masked out in the cloudy regions using a threshold, the available data decrease and there can be samples in the remaining IWV observations which are underestimated. This can make the IWV observed by MERIS differ from that estimated from GNSS measurements.
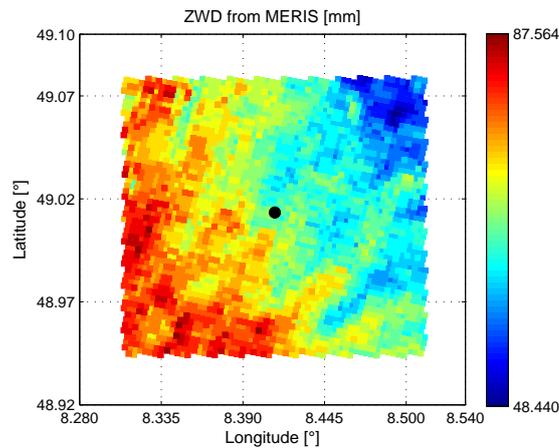
Figure 5.7: MERIS measurements of ZWD in a block of 15×15 km^2, and the little black circle defines the location of the GNSS site (KARL).

Time	06.27.2005	09.05.2005	07.17.2006	10.30.2006	04.23.2007
RMS [mm]	6.3514	5.7456	5.6483	8.1244	3.3830

Table 5.1: RMS of the difference between GNSS and MERIS ZWD observations

We compare in the same way the GNSS ZWD with the IWV simulated by the WRF model, after converting into ZWD values. In **Figure 5.9** the ZWD from GPS and WRF are shown, and the corresponding correlation coefficient (CC) and RMS values are given in **Table 5.2**.

(a) Scatter plot (b) ZWD differences
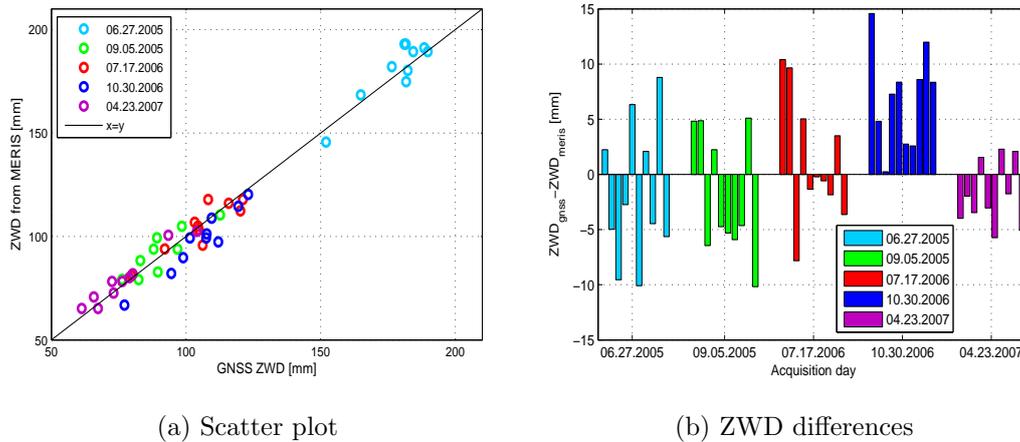
Figure 5.8: (a) MERIS ZWD plotted against GNSS ZWD from 10 receivers at SAR overpass time in 5 days. The MERIS observations are averaged in blocks of 15×15 km^2 to emulate GNSS. The slope of the black line is 1. (b) The difference between the ZWD values from GNSS and MERIS. The difference is calculated at the 10 receivers at 5 days.

Generally, it is expected that the ZWD values at the 10 GNSS sites decrease in the direction of increasing heights. Nevertheless, this reduction of the ZWD values with the increasing surface altitude might not be observed from the data. This is well explained by the spatial variability of water vapor content and the local effects of turbulently-mixing water vapor around each site. These effects may increase the wet delay observed at the site even if it is located at a higher altitude. This effect will be more clear if the GNSS sites are widely separated and the differences of their altitudes are small.

Time	11.29.2004	03.14.2005	05.23.2005	08.01.2005
RMS [mm]	7.862	4.0248	15.3143	12.4249
Correlation coefficient	0.8461	0.8058	0.9279	0.8011

Table 5.2: Correlation coefficients between GNSS and WRF ZWD data and the RMS values.

5.1.5 Interpolation of surface meteorological temperature and pressure data

In many cases, meteorological stations are distant from the location of the GNSS site; for that reason, spatial interpolations of meteorological parameters (e.g., P, T) are imperative. If we inspect the temperature, it can be divided into two components: one (the largest) is elevation dependent and the second is dependent of the spatial location. In order to interpolate the temperature in space, we do the following: the altitude dependent component is removed by line fitting ($T(z) = az + b$), then we build a variogram of the temperature residuals. The spatial variations of temperature in the area of URG (600×600 km^2) can be modeled by an exponential variogram function with a range of about 300 km. To find the value of the temperature at any (x, y, z) location, we calculate a height-dependent component using the

(a) Scatter plot   (b) ZWD differences
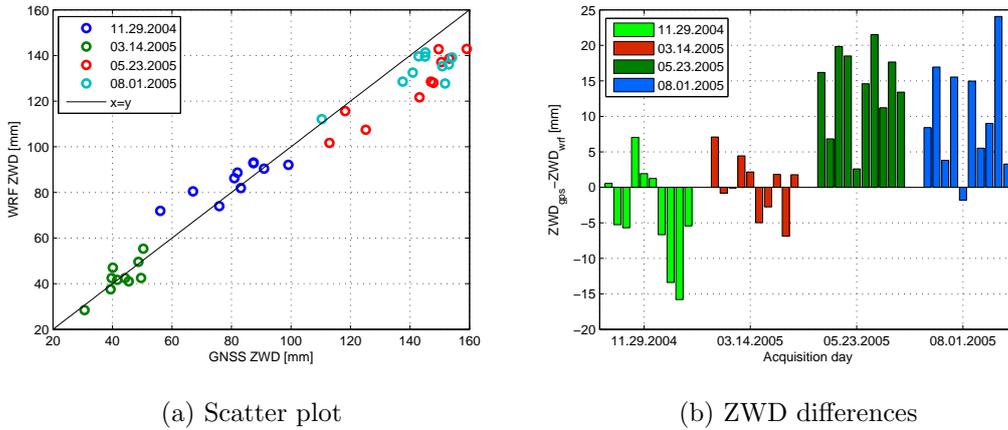
Figure 5.9: ZWD from GPS and WRF at three SAR acquisition times observed by 10 sites. ZWD from WRF is averaged within a box centered at the GPS site. The corresponding correlation coefficient and RMS values are depicted in **Table 5.2**.

coefficients of the line fitting and add a correction value to account for the horizontal variations,

$$T(x, y, z) = T_0(z) + \Delta T_{corr}(x, y) \tag{5.14}$$

where $T_0(z) = az + b$, and $\Delta T_{corr}(x, y)$ is a correction term depending on the spatial location of the site, which is modeled by a semivariogram. We analyzed more than 40 variograms of the temperature residuals in different year seasons. I is observed a similar performance in all of them with a slight difference in the variogram sill but the variogram range remains almost unchanged. Accordingly, we can use a representative variogram model with an average sill from all inspected variograms. The variogram model in **Figure 5.10** can be utilized in the absence of spatially-dense meteorological data, where building a proper variogram is not possible.
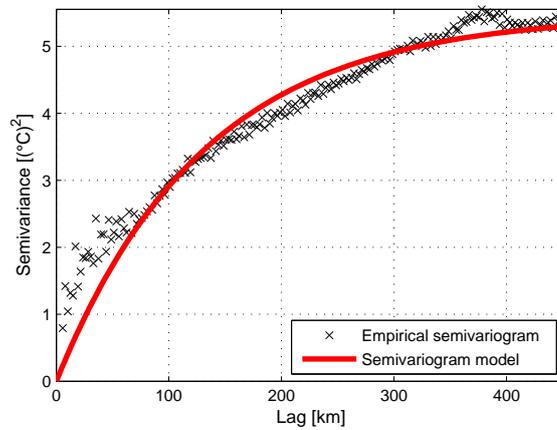
Figure 5.10: The experimental and theoretical semivariogram models for the temperature maps simulated by the WRF model.

We used WRF temperature simulations to test this method. Due to the lack of other 2D data, we used temperature maps of WRF down-sampled by a factor of ten as input for the method. The output grid is defined at a spatial resolution of 9×9 km^2. The original temperature maps are aggregated into 9 km cells to be compared with the interpolated temperature. **Figure 5.11** shows the results for a day in winter and a day in summer. The interpolations made using the semivariogram model shown in **Figure 5.10**, show a strong agreement with the maps simulated by WRF. The difference maps follow a Gaussian distribution with a mean close to zero and a standard deviation of less that 0.3°C.

In order to interpolate the air pressure to different spatial locations, a similar scenario is used. The elevation-dependent component can then be computed at different altitudes using eq. (5.13), and the residuals are spatially interpolated using the method of IDW. Because the partial pressure of water vapor is more variable, we only interpolate or extrapolate the pressure of the dry air. It was not possible to compute one common variogram for the pressure, since the data in different areas depicted different spatial behavior.
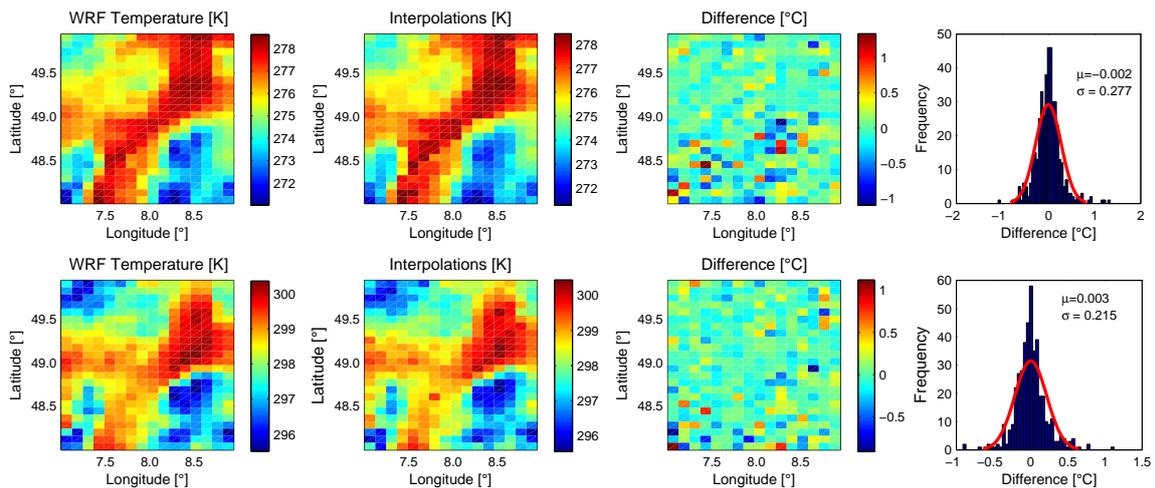
Figure 5.11: Temperature maps constructed by summing a height dependent component and a correction value accounting for spatial variations determined by kriging. The first row shows results using data on 01.03.2005 and the second on 06.27.2005. The mean (μ) and the standard deviation (σ) of the differences are shown in the histogram plots.

5.2 Neutrospheric delay in InSAR

In the previous section, we discussed the procedure followed to estimate the ZWD at the GNSS sites. In this section, we present the method used to derive the wet delay maps from InSAR data. We first give an overview of the SAR system and the InSAR technique.

5.2.1 SAR

Synthetic aperture radar (SAR) is an active remote sensing sensor that has its own illumination source; therefore, it can collect data at day and night times. The operating frequencies are selected properly to allow the operation during all weather conditions, so that the microwave signal can penetrate clouds, haze, rain, fog and precipitation with marginal attenuation. To overcome the limitation of physical antenna aperture, SAR applies signal processing techniques (using the Doppler effect associated with each echo) to achieve high spatial resolution through the combination of many low-resolution but correlated radar images. In order to form an image, the radar transmits pulses to the ground, and the return echos scattered from the objects on the ground are received by the antenna. The image can be acquired in different modes such as stripmap (which is used in this work), ScanSAR, and spotlight by adjusting the radar imaging geometry [Bamler and Hartl, 1998]. The raw data are processed to generate a single look complex (SLC) SAR image. A SAR image can be thought of as a two-dimensional array of complex values, representing the brightness of the scattering objects on the ground and the phase associated with the traveled distance. Each resolution cell observed on the ground returns a signal that is received by the radar and stored in the array as a complex number in range and azimuth coordinates. The size of the ground cell depends on the parameters of the SAR system; for example, Envisat radar has a spatial resolution of about $20{\times}4$ m^2.

The information of the image can be represented in amplitude and phase components. The phase signal of one SAR image is uniformly distributed due to the random superposition of echoes from all scatterers within each resolution cell. This effect masks the useful information about the geometric range between the radar and the ground. To make use of the phase information for measuring different signals such as surface displacement and elevation, the technique of Interferometric SAR (InSAR) is applied.

5.2.2 InSAR

In most InSAR applications, two images acquired for the same region, but with slightly different viewing geometries are used. The images can be captured by the same antenna with a temporal shift (repeat-pass InSAR), or by two different antennas at the same time (single-pass InSAR) [Bamler and Hartl, 1998]. In single-pass InSAR, the change in the atmospheric effect between the two images is insignificant and it cancels out by interferometry. Therefore, we limit this study to repeat-pass InSAR. Interferometry is achieved by multiplying the first SLC image ($S_1 = |S_1| \exp(j\phi_1)$) by the complex conjugate of the other ($S_2 = |S_2| \exp(j\phi_2)$), i.e.,

$$V = S_1 S_2^* = |S_1||S_2| \exp(j(\phi_1 - \phi_2)) \tag{5.15}$$

where V is called *complex interferogram*, ϕ_1 and ϕ_2 are the phase of the first and the second image, respectively. The star in the above equation indicates the conjugate. The phase of each

resolution cell is a sum of a component that is proportional to the geometric distance and a scattering phase. If the scattering characteristics of the resolution cell stay the same during the two acquisitions, then the phase due to the scattering effect nearly cancels out [Hanssen, 2001, p. 36]. Under these conditions, each element in the interferogram contains a value of the interferometric phase that measures the difference of the distances from the satellite to the ground, i.e.,

$$\phi_{int}(x, y) = -\frac{4\pi}{\lambda}\Delta R + \phi_{noise} \tag{5.16}$$

where $\Delta R = R_1 - R_2$ denotes the difference between the observed distance from the ground cell to the radar at the first and the second overpass times, see **Figure 5.12**.
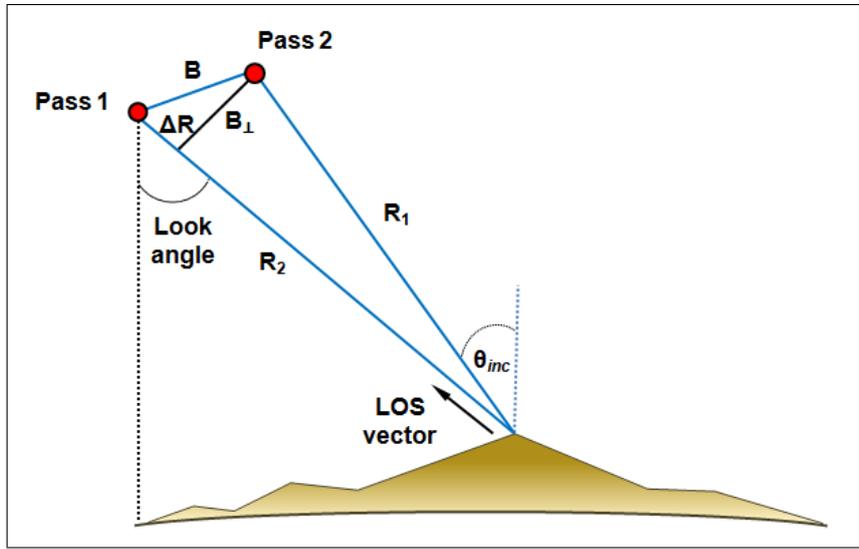
Figure 5.12: Imaging geometry of InSAR. R_1, R_2 measure the distances from the satellite to the ground at the first and the second overpass times. B is the spatial baseline, $B_\perp$ is the perpendicular baseline, and θ_{inc} is the radar incidence angle.

The interferometric phase for each pixel in an interferogram is given by the superposition of different contributions such as topography, Earth surface displacement, and atmosphere. InSAR phase for each pixel in the interferogram is given by:

$$\phi_{int} = \phi_{topo} + \phi_{disp} + \phi_{neu} + \phi_{iono} + \phi_{orbit} + \phi_{flat} + \phi_{noise} \tag{5.17}$$

where ϕ_{topo} is the topographic phase component. ϕ_{disp} is the phase component due to the Earth's surface displacement between the two SAR acquisitions. ϕ_{neu} is defined as differential phase shift caused by the propagation of the signal through the neutrosphere. Similarly, ϕ_{iono} is the differential phase shift due to the signal propagation through the ionosphere. The phase component due to the inaccuracy of the satellite orbit is denoted by ϕ_{orbit} and ϕ_{flat} is the phase component due to the Earth's reference phase. ϕ_{noise} is the component due to the system thermal noise and the loss of coherence between the two observations.

The focus of this work aims at separating the neutrospheric phase from other phase components in eq. (5.17). We will describe later in this chapter the way to reduce the contribution of all other phase components to extract the neutrospheric phase. The ionosphere has a marginal influence on radar signals transmitted at frequencies of C- and X-band [Gray et al., 2000], while it has been shown that the ionosphere has a significant impact in L-band SAR observations [Meyer et al., 2006]. Furthermore, the ionospheric effect is expected to model a trend of long-wavelength that is inseparable from other phase ramps; thus, they are eliminated during InSAR processing [Zebker et al., 1997]. In Chapter 2, we examined the influence of the neutrosphere on microwave signals and classified the encountered delay into wet and dry components. In the same way, the neutrospheric phase in InSAR is given by

$$\phi_{neu} = \phi_{dry} + \phi_{wet} \tag{5.18}$$

The dry delay is characterized by smooth and slow variations in time and space. By applying interferometry, the dry delay is mostly eliminated and the artifacts observed in the interferogram are mainly caused by wet delay [Zebker et al., 1997]. GNSS and InSAR are affected by the neutrospheric delay in a similar way. However, GNSS provide absolute measurements of the delay while the neutrospheric observations from InSAR represent a difference between the neutrospheric states at two acquisition times. GNSS provide observations at spatially sparse locations compared to the high spatial resolution of InSAR observations.

The phase observations in InSAR is given by the difference of two observations, hence the neutrospheric phase observed in InSAR is given by

$$\phi_{neu}(x, y, t_1, t_2) = -\frac{4\pi}{\lambda \cos \theta_{inc}} 10^{-6} \left(\int_0^Z N(x, y, z, t_1) dz - \int_0^Z N(x, y, z, t_2) dz \right) \tag{5.19}$$

where t_1 and t_2 denote the SAR acquisition times. θ_{inc} is the radar incidence angle and the factor $\frac{1}{\cos \theta_{inc}}$ is a mapping function applied to project the neutrospheric delay from the zenith direction to the radar LOS. The theoretical value of θ_{inc} for Envisat is $\approx 23°$. Because the radar signal travels downwards and upwards through the neutrosphere, a factor of 2 apparent in eq. (5.19). It should be noticed that the fraction $-\frac{4\pi}{\lambda}$ is used to convert the delay in meters into phase delay in radians, and its reciprocal is used to convert the radians into meters.

5.2.3 InSAR time-series analysis

It is common that interferograms with long temporal baselines contain noisy areas where no measurement is possible. This noise can originate from temporal decorrelation in vegetation areas where low or even no coherence between the two SAR images is observed [Li and Goldstein, 1990]. Decorrelation can also result from the change in the viewing geometry, where the change in the spatial baselines can change the coherent sum of the wavelets from different

scatterers within the resolution cell resulting in spatial decorrelation [Zebker and Villasenor, 1992]. Spatial decorrelation restricts the number of image pairs suitable for interferometric applications [Ferretti et al., 2001]. Multi-temporal InSAR techniques are extensions of conventional InSAR methods developed to address the problems of decorrelation and optimized to minimize errors caused by the neutrospheric delay. These techniques involve simultaneous processing of multiple SAR acquisitions of the same area to increase the spatial density of extractable information and to reduce the error associated with the estimated deformation signal. Currently, there are two categories of algorithms for processing multiple acquisitions: persistent Scatterers InSAR (PSI) and small BAseline Subset (SBAS) method [Hooper, 2008]. If the scatterers are of comparable strength, the SBAS method mitigates the neutrospheric effect by averaging numerous interferograms of short spatial baselines under the assumption of zero-mean Gaussian properties of the atmospheric signal delays. Through averaging large data stacks, improved information about average surface velocity measurements can be obtained. On the contrary, if the resolution cell is dominated by one scatterer, the influence of other scatterers is minimal and the received signal is stable, it is recommended to use PSI. The PSI technique aims at identifying temporally coherent targets that show phase stability at all acquisition times. These targets are point-wise objects that are plentiful in the city-type environments but are less likely in non-urban regions. PSI was developed to overcome InSAR limitations of temporal and geometrical decorrelation [Ferretti et al., 2001]. Different methods have been developed to identify PS points in the interferograms, which were successful in urban areas where many man-made objects are available and hence many PS points are possible, but tend to provide only few PS points in the areas of natural terrain [Ferretti et al., 2001; Kampes, 2005; Lyons and Sandwell, 2003]. The Stanford Method for Persistent Scatterers (StaMPS) is an alternative approach that was successfully developed to define a larger number of PS candidates even in non-urban areas [Hooper et al., 2007, 2004]. SBAS will be used in future work within the project to extract neutrospheric phase from interferograms and to test the influence of redundancy from difference subsets on the reconstruction of the phase at each SAR overpass time. A new method for PSI analysis, SqueeSAR, has been developed to increase the density of stable targets by analyzing both point-wise PS targets and distributed scatterers [Ferretti et al., 2011]. This technique should be used in future work to increase the number of stable targets particularly in the green areas, which are common in the URG.

5.2.4 StaMPS framework

StaMPS was initially developed for PSI applications in natural regions [Hooper et al., 2007, 2004], and then SBAS analysis was added to the package of StaMPS [Hooper, 2008]. The StaMPS framework contains a collection of spatial and temporal filtering routines that allow to estimate each of the phase components contained in the interferometric phase by assuming its spatial and temporal structure. We use StaMPS in this work to process an Envisat image stack and reconstruct the neutrospheric phase. Data processing in StaMPS, selection of PS pixels, estimation of other phase components and noise, and phase unwrapping are described

in [Hooper et al., 2007], from which we built most of the material in this section.

The selection of the PS pixels in StaMPS is achieved by inspecting the amplitude and the phase of the pixels, where testing the amplitude is done to reduce the computation time. As the amplitude stability is a measure for phase stability of a pixel in the interferogram, the PS candidates are selected in the first iteration such that the amplitude dispersion index is below a threshold value. The amplitude dispersion index is defined as the ratio between the standard deviation and the mean of the SAR pixel amplitude [Ferretti et al., 2001]. We tested different threshold values and the most suitable value for the natural terrain in URG was 0.4. The amplitude analysis provides first PS candidates and then the phase stability is measured for the selected candidates by phase analysis. By applying phase analysis, unstable pixels or partially stable pixels (stable in some interferograms) are rejected. It is required to estimate the phase noise for each PS candidate and the final PS pixels are those pixels that have very little noise. Since the wrapped phase for each pixel has different contributions as given in eq. (5.20), the noise is dominated by the variations of the first four components. Therefore, we first need to separate the useful signal from the noise. The phase noise is temporally and spatially uncorrelated, hence to estimate the phase noise other phase components, that are spatially correlated, are filtered out. However, the phase due to look angle error has a component that is spatially uncorrelated, and it is estimated using least squares fitting taking the benefit of its correlation with the perpendicular baseline. By subtracting the spatially correlated terms and the residual look angle error, the remaining component is the phase noise. The variation of the phase noise is determined for each pixel and by setting a threshold value the pixel can be confirmed to be a PS pixel or not. We followed the above described strategy to define the PS points in the SAR images.

The interferograms are formed at full resolution in range and azimuth in order to maximize the signal-to-clutter ratio of the resolution cells containing a single dominant scatterer [Agram, 2010]. The signal-to-clutter ratio is defined as the ratio between the reflected energy from the dominant scatterer in the resolution cell to that of the energy reflected from the rest of the elements in the cell [Adam et al., 2003; Kampes, 2005]. The interferograms are corrected for the phase trend generated by flat Earth before the StaMPS processing. Also, the phase due to the Earth's topography is subtracted with the aid of a DEM model of 10 m resolution. The residual wrapped phase for each PS point is then given by:

$$\Delta \psi_{ik}^{psi} = W\left\{\phi_{disp,ik} + \phi_{neu,ik} + \Delta\phi_{orbit,ik} + \Delta\phi_{la,ik} + \phi_{noise,ik}\right\} \qquad (5.20)$$

where $\Delta\psi_{ik}$ is the residual interferometric phase for the ith PS point in the kth interferogram. $\phi_{disp,ik}$ if the phase change due to the displacement of the PS in the radar LOS direction. The neutrospheric phase is denoted by $\phi_{neu,ik}$. The residual phase due to the inaccuracies in the satellite orbits is $\Delta\phi_{orbit,ik}$. The slightly different acquisition geometries between the first and the second overpass times and the error in the DEM can result in a residual phase term, called look angle error, denoted by $\Delta\phi_{la,ik}$. $\phi_{noise,ik}$ is the noise due to variability in

scattering, thermal noise, coregistration error, and uncertainty in the phase center in azimuth, and $W\{\cdot\}$ defines the wrapping operator. Note that the term ϕ_{iono} apparent in eq. (5.17) is removed here by assuming that it often has a negligible effect in C-band data.

The retrieval of the neutrospheric phase is achieved in StaMPS by analyzing the spatial and temporal characteristics of each phase component in the interferogram. **Table 5.3** shows the spatial and temporal properties of all phase components appearing in eq. (5.20). The first four terms in the equation are correlated in space while the noise is uncorrelated; therefore, using a low pass filter in space separates the noise from the signals. The noise signal is assessed statistically and the statistical parameters of the noise are used to adapt the filter and the first four terms are re-estimated and subtracted to obtain a new estimate of the noise. This is iterated in a loop until the estimate of the noise converges, and we get a component for the correlated phases. The uncorrelated component of $\Delta\phi_{la,ik}$ is approximated by a linear relation with the perpendicular baseline and it can be estimated by least squares. After subtracting the uncorrelated phase terms, phase unwrapping is applied and the resulting unwrapped phase is:

$$\Delta\phi_{ik}^{psi} = \phi_{disp,ik} + \phi_{neu,ik} + \Delta\phi_{orbit,ik} + \Delta\phi_{la,ik}^{c} + \Delta\phi_{noise,ik} + 2n_{ik}\pi \qquad (5.21)$$

$\Delta\phi_{la,ik}^{c}$ is the spatially-correlated part of the phase $\Delta\phi_{la,ik}$, and $\Delta\phi_{noise,ik}$ is the residual noise term. n_{ik} is the integer ambiguity which should remain unchanged for most PS points in the interferogram if the unwrapping was done correctly.

In this work, we are interested in retrieving the term $\phi_{neu,ik}$ from the unwrapped phase in eq. (5.21). The most challenging point is to separate the neutrospheric phase from the phase due to surface displacement. However, the tectonic surface motion in the test area of URG is observed to have a small, long-term change. Accordingly, the phase component $\phi_{disp,ik}$ can be assumed negligible. For the remaining phase terms, StaMPS attempts to separate the master and slave contributions by temporal filtering. The contribution of the master to the neutrospheric and the orbital phases is correlated in time, while the contributions from the slaves are not correlated in time as well as the component $\Delta\phi_{la,ik}^{c}$. The master contribution to the neutrospheric and orbital phase can be retrieved by applying a low-pass filter in time. The slave contributions are retrieved by applying a high-pass filter in time followed by a low-pass filter in space to remove noise. The spatially-correlated look angle (SCLA) error is dependent of the baseline and it is estimated using least squares. The orbital phase can be modeled by a 2D linear trend and subtracted from the unwrapped phase. The neutrospheric phase in each interferogram is the sum of the phase of the master and the corresponding slave. Estimating the neutrospheric contribution of the slave requires adjusting temporal and spatial parameters of the filter, we found the phase estimates of the slave neutrosphere depend highly on the filter parameters (e.g., temporal and spatial correlation lengths), which are not easy to fix. Because we assumed that the displacement phase is negligible, we filtered out noise and subtracted the estimated orbital trend and the SCLA signal, so that the remaining phase is due to the neutrosphere without the need for splitting the master and slave signals at this stage.

Phase component	Spatial correlation	Temporal correlation
$\phi_{disp,ik}$	correlated	correlated
$\phi_{neu,ik}$	correlated	master signal is correlated, slaves not
$\Delta\phi_{orbit,ik}$	correlated	master signal is correlated, slaves not
$\Delta\phi_{la,ik}$	only partially correlated	uncorrelated
$\phi_{noise,ik}$	uncorrelated	uncorrelated

Table 5.3: Temporal and spatial properties of the phase components in the interferogram

5.2.5 Extracting SWD-difference maps

The neutrospheric phase is converted into slant delay (along the radar LOS) in units of millimeters by multiplying the phase value, in radians, by the factor $-\frac{4\pi}{\lambda}$, i.e.,

$$\Delta STD^{psi}(x,y) = -\frac{4\pi}{\lambda}\phi_{neu}(x,y) \tag{5.22}$$

where $\Delta STD^{psi}(x,y)$ is the slant neutrospheric delay-difference, and λ is the wavelength of the microwave signal, which is about 56 mm in the C-band. As mentioned earlier, the neutrospheric phase is a sum of a phase shift due to dry gases and a phase shift due to water vapor. Since dry gases are quite repeatable [Zebker et al., 1997], their effect is significantly reduced when building the interferogram. If a residual component of the dry delay remains in the interferogram, it will be indistinguishable from the orbital ramps, and it will be eliminated when modeling and subtracting those ramps. Accordingly, we consider the neutrospheric phase extracted from the set of interferograms to be caused by water vapor (i.e., $\Delta STD^{psi} = \Delta SWD^{psi}$). The ZWD at any spatial location (x,y) can be obtained from the SWD measurements using a mapping function that is dependent of the radar incidence angle, i.e.,

$$\Delta ZWD^{psi}(x,y) = \Delta SWD^{psi}(x,y)\cos\theta_{inc} \tag{5.23}$$

where ΔZWD^{psi} is the zenith-directed wet delay-difference and ΔSWD^{psi} is the SWD-difference. θ_{inc} is the incidence angle of the radar (see **Figure 5.12**). Since the SWD-difference maps extracted from the interferograms are relative measurements to a reference master, we call them *SWD-difference maps*, and they can be converted into *IWV-difference maps*.

We built interferograms from an Envisat image stack with respect to a single reference image (the master) acquired on 06.27.2005 at 09:51 (UTC). StaMPS is applied to 16 interferograms corrected for flat Earth trends and topography. The total number of identified stable-phase pixels is 76841, and the average density of PS points in the interferogram is about 11 points/km^2. Please note, that the density is much higher in urban regions, but due to the absence of PS

points in forest regions, the overall average decreases. This density is however good enough for atmospheric studies where we can rely upon the spatial correlation between the observations. After the PS points are identified and all other points are discarded, we started separating the neutrospheric phase from other phase components as described before. **Figure 5.13** shows different examples of the SWD fields extracted from the interferograms. The regions of forests and vegetation west to the Rhine valley and the southern-east show almost no presence of PS targets, while the PS density is high within the cities. The spatial variability of the SWD fields is dependent of the neutrospheric state at the acquisition time of the master and slave scenes. We can observe correlations between the six SWD-difference maps due to the presence of the master contribution in each map. In the map shown in (a,f), the observed signal is dominated by the contribution of the master, that is explained by the low activity and dry weather conditions during December (slave image acquisition time) in the URG region where insignificant water vapor content and variations are expected, in contrary to the master signal. In (b), for example, the SWD signal is caused mainly by water vapor stratified in the neutrospheric layers; however, the map in (e) depicts clearly the effects of turbulent mixing of water vapor.

5.2.6 Comparative analysis

Before proceeding to further investigations, we had to evaluate and analyze the neutrospheric phases derived from the interferograms to confirm if they are appropriate for water vapor mapping. We used the data available from MERIS to compare the SWD-difference maps derived from PSI. IWV maps from MERIS are translated into ZWD values and mapped to the LOS of the radar using eq. (2.29). The empirical constant Q apparent in the equation is determined from eq. (2.24) using surface meteorological observations. For the sake of comparison, we built differences of SWD maps from MERIS (ΔSWD^{meris}). To emulate InSAR, we eliminated the long-wavelength signal (2-dimensional linear trend) by fitting a plane to the data using least squares. The other point that should be considered here is that PSI SWD difference maps are point-level data, while ΔSWD^{meris} has a spatial resolution of approximately 300×300 m^2. Therefore, we aggregate PSI data to the MERIS resolution, which is achieved either by spatial averaging or interpolation.

We made a comparative analysis of MERIS and PSI SWD-difference maps by comparing the data pixel by pixel, and by inspecting the spatial properties. In **Figure 5.14**, are shown the SWD maps derived from the interferograms and the corresponding SWD maps from MERIS for two different dates. InSAR data are interpolated at he MERIS resolution using the method of moving-window averaging. The results show strong spatial correlation between PSI and MERIS, with correlation coefficients of 0.86 and 0.84, respectively for the upper and lower maps. The difference between the PSI and MERIS maps follow a Gaussian distribution with a mean close to zero and a σ of 4-5 mm.

Furthermore, we applied four interpolation methods: moving-window averaging, IDW, surface fitting, and OK. The first three methods are not based on statistical properties of the

(a) 12.15.2003

(b) 07.12.2005

(c) 09.05.2005

(d) 07.17.2006

(e) 04.23.2007

(f) 12.08.2008
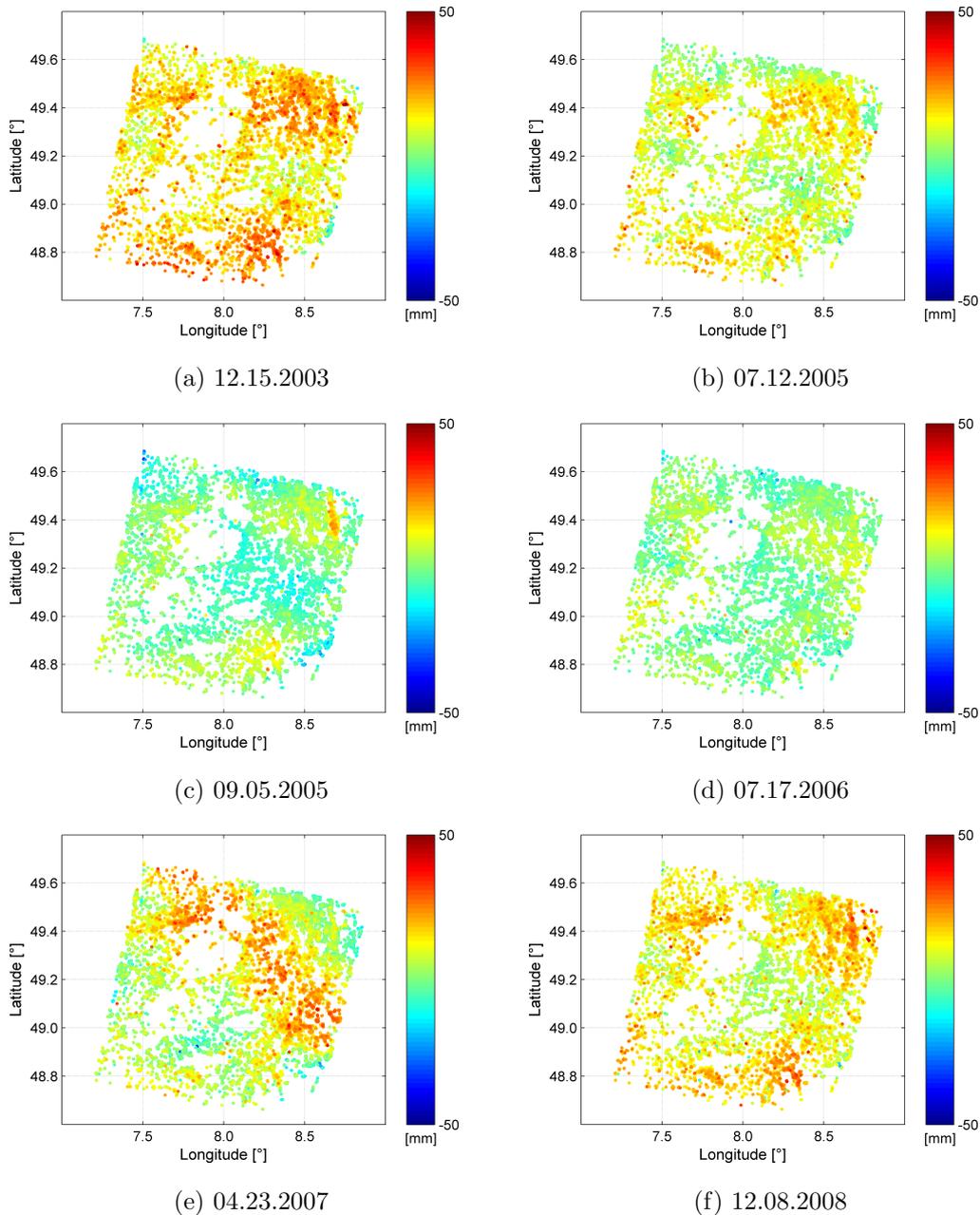
Figure 5.13: Neutrospheric delay difference maps derived from PSI. The master scene (reference) was acquired on 06.27.2005 and the acquisition dates of the slaves are specified in the figure. The spatial variations in each sub-figure depend on the atmospheric state at the acquisition time of the master and slave images.

data, which means that in the areas where no PSI data are available, we cannot provide proper estimates of the SWD. For that reason, we masked out the MERIS data in the corresponding regions. For the moving-window averaging, we simply average the SWD values for the PS points located within a MERIS cell. In a similar way, the IDW method calculates a weighted average of the SWD at the PS points within the cell using eq. (4.6). In the third method,
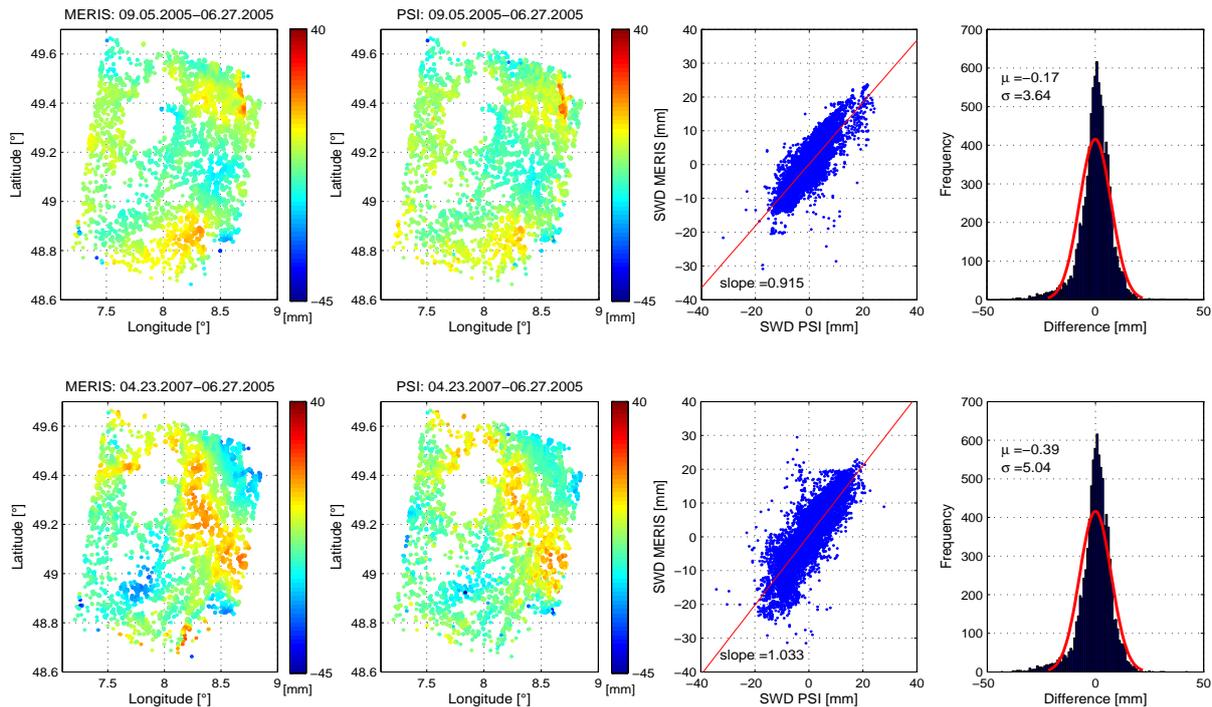
Figure 5.14: Comparison of the SWD-difference maps derived from PSI and MERIS. The master image was acquired on 06.27.2005, while the slaves were acquired on 09.05.2005 (top) and 04.23.2007 (bottom). The spatial correlation coefficients are 0.86 for the upper example and 0.84 for the lower. The difference maps have Gaussian distributions with mean (μ) and standard deviation (σ) as shown within the histogram plots.

we create an interpolant by fitting a surface to the scattered PSI data and use it to make the interpolations to any location. We note that this method, in its simple form, does not account for the data change of support, which might make the results less accurate compared to the other methods. The last method is the OK, details on using OK for interpolation are described in Section 4.2.2.1. **Figure 5.15** exhibits the SWD maps from MERIS, and the PSI SWD maps interpolated at the MERIS locations using the four methods. The corresponding correlation coefficients and the RMS values are summarized in **Table 5.4**. The four methods lead to similar results with slight changes in the values of the correlation coefficients and the RMS of the difference maps between PSI and MERIS data. The advantage of OK over the other methods is that estimates can be made even in the lack of data in some regions because this method creates a model of the spatial correlation between the observations. The computational complexity of kriging can be addressed as described in Section 4.2.2.3. If no gaps are observed in the maps, the other methods can be exploited for generating regular maps at low computation costs.

Microwave signals propagating in a turbulent neutrosphere are expected to have spatial fluctuations described by the power spectral density and structure functions. These functions should follow a power law described by the elementary turbulence theory of Kolmogorov. We inspected the spatial variations of the SWD, as an approximate tracer of the refractivity

(a)

(b)

(c)

(d)

(e)
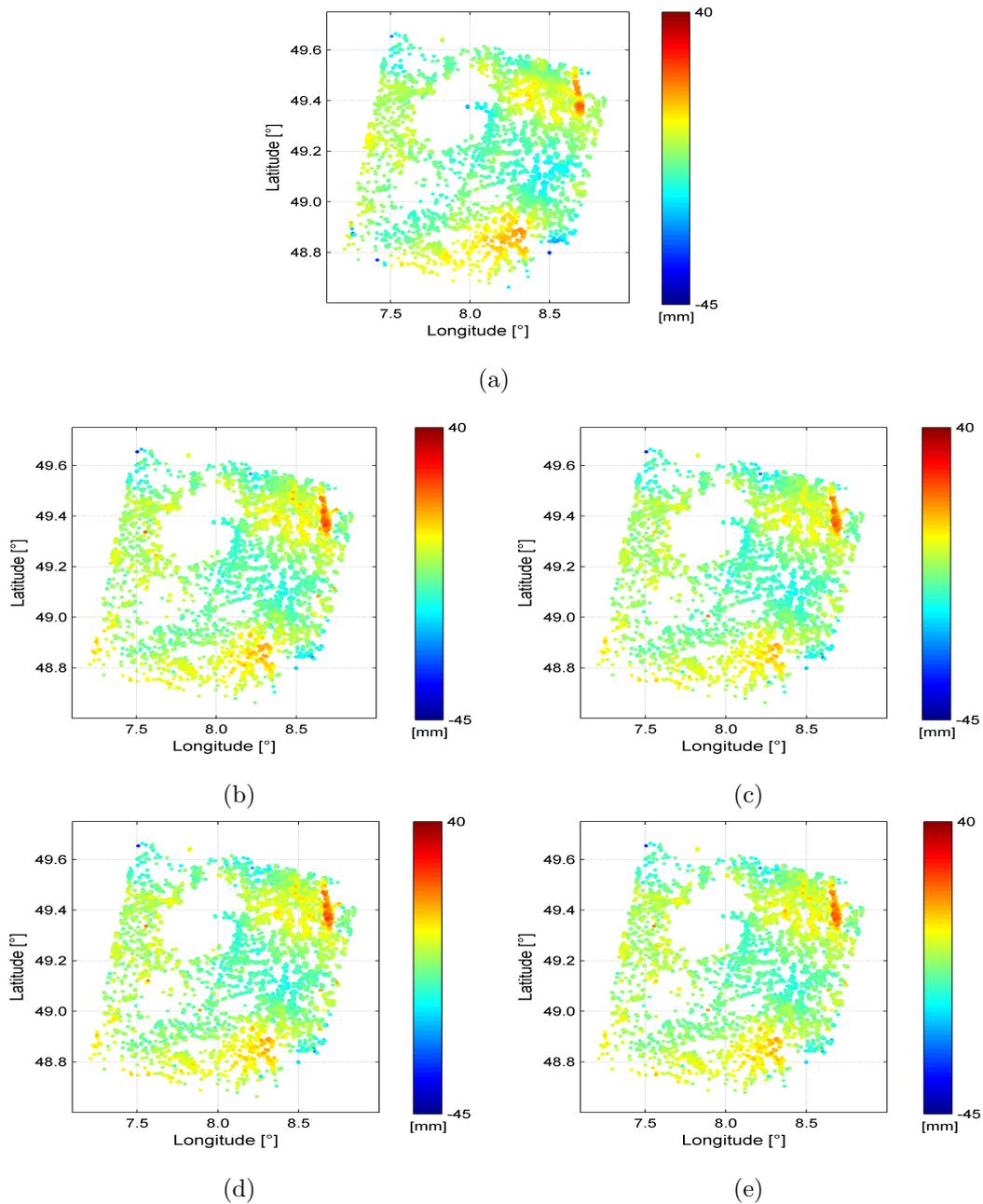
Figure 5.15: (a) SWD-difference maps derived from MERIS IWV, where the difference is built from the observations on 06.27.2005 and 09.05.2005. The long-wavelength component is estimated and subtracted. SWD from PSI phase was interpolated at MERIS locations using ordinary kriging (b), moving-window averaging (c), inverse distance weighting (d), and surface fitting (e).

fluctuations. Assuming geometrical isotropy, we calculate the structure functions of PSI and MERIS SWD-difference maps shown in **Figure 5.14**. The curves of the SFs are depicted in **Figure 5.16**. The structure function predicts the decorrelation of the wet refractivity over the distance. The decorrelation rate is less than 5/3 for distances of 1-2 km, and it decreases to 2/3 for distances larger than 2 km until a range where no spatial correlation is observed. Both data

Quantity\ method	Moving-window	Ordinary kriging	IDW	Surface fitting
Correlation coefficient	0.8415	0.8193	0.8192	0.8078
RMS [mm]	3.6484	3.9226	3.9246	3.9883

Table 5.4: Correlation coefficients between MERIS and PSI data, and the RMS values of the difference signal.

sets show a correlation range of about 20 km. These examples show a good correspondence to the results presented in [Hanssen, 2001, pp. 144] and to the models presented by Treuhaft and Lanyi [1987]. From the plots, we notice that the change of the weather conditions affects the power of the signal, but not the slope of the structure function.

Next we analyzed the power spectra of the zero-mean SWD-difference maps. The power spectrum reflects the energy transfer over distance, which is expected to decrease as the distance increases. **Figure 5.17** depicts the rotationally-averaged power spectra computed for the same maps, assuming geometrical isotropy. The power spectra of InSAR and MERIS follow a decay slope of -8/3 on a log-log scale. [Hanssen, 2001, pp. 145] computed the PSD from 8 interferograms over the Netherlands with a decay at slope varying between -5/3 and -8/3 for different regimes. Also, Goldstein [1995] observed that the PSD of the phase signatures over Mojave Desert decays at a rate of -8/3.
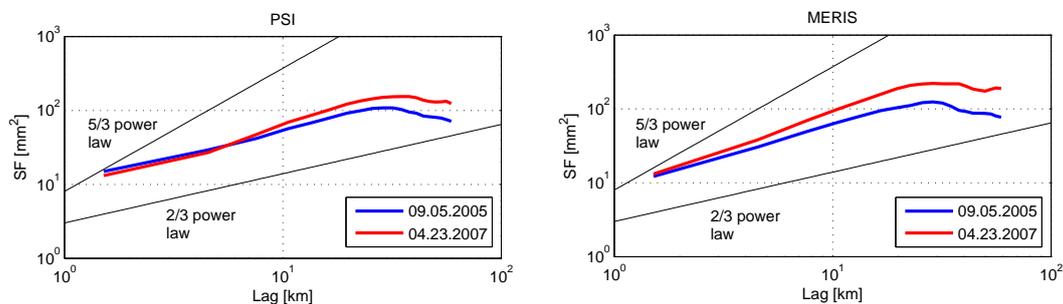
Figure 5.16: Structure functions for two SWD-difference maps from PSI and MERIS. The master image was acquired on 06.27.2005, while the acquisition dates of the slave images are shown in the sub-figures. The black lines indicate the model shown by Treuhaft and Lanyi [1987].

5.3 Summary

In this chapter, we described the methods applied to derive the neutrospheric delay from GNSS and InSAR measurements. We explained the importance behind using meteorological data (pressure, temperature, relative humidity) in the determination of the ZWD from the ZTD estimated based on GNSS observations. The ZWD estimates from the GNSS observations show strong agreement with the data from MERIS and WRF. Since the GNSS
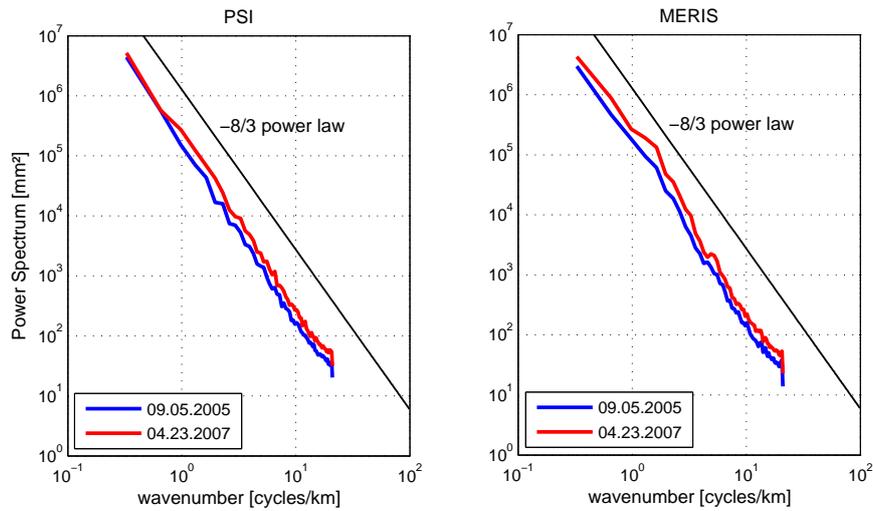
Figure 5.17: Rotationally-averaged spatial power spectra computed for the SWD-difference maps from PSI and MERIS. The master image was acquired on 06.27.2005, while the acquisition dates of the slave images are shown in the sub-figures. The black solid line has a slope of -8/3.

ZWD estimates represent the average of the neutrospheric effect within a cone above the antenna, aggregation of MERIS and WRF data is done to enable a proper comparison. Also, we extracted SWD-difference maps from a coregistered ASAR image stack by applying the technique of PSI. The comparative analysis with MERIS SWD-difference maps demonstrates strong spatial correlation and a standard deviation of the differences of about 4-5 mm.

6 Combination of InSAR and GNSS data

In the previous chapter, we described how the neutrospheric delay is estimated from InSAR and GNSS phase observations and how it is used to retrieve the wet delay. We used meteorological observations of pressure, temperature, and relative humidity to retrieve the wet delay from GNSS observations at each GNSS site. Unlike GNSS which is utilized to estimate absolute values of the wet delay, wet delay maps extracted from InSAR data are the difference between the wet delay maps at two acquisition times. We exploit the wet delay estimates obtained from GNSS observations to reconstruct the missing delay component and generate maps of absolute SWD at each SAR overpass time.

In order to specify which wet delay components were eliminated by building interferograms, we analyzed maps of IWV measured by MERIS. In Chapter 2, we classified atmospheric water vapor into an elevation-dependent (vertically-stratified) component, and turbulently mixing water vapor. Accordingly, we split the maps of IWV measured by MERIS into an elevation-dependent IWV and a residual component, i.e.,

$$IWV_{resid}^{meris} = IWV^{meris} - IWV_{st}^{meris}(z) \qquad (6.1)$$

where IWV^{meris} is the IWV measured by MERIS, IWV_{st}^{meris} is the vertically-stratified IWV, and IWV_{resid}^{meris} is the residual IWV. We analyzed the residuals to test if there exists any deterministic long-wavelength signal that results from the change of air pressure and temperature over distance. This signal, if present, accounts for the neutrospheric anisotropy over large scales. We found that in most cases the residuals contain beside the short-wavelength signal caused by turbulent mixing water vapor, a long-wavelength signal (2-dimensional ramp). We estimated this signal by fitting a 2-dimensional ramp to the residuals using least squares adjustment. The remaining signal after subtracting the vertically-stratified IWV and the ramp is caused by short-scale mixing of water vapor. **Figure 6.1** shows a map of MERIS IWV as a sum of the three components: elevation-dependent component, long-wavelength component (ramp), and short-wavelength component. We call the sum of the first two components the non-turbulent wet delay and the third the turbulent wet delay. The ramp generally contains a signal of small magnitude, but it is essential for building correct water vapor maps from GNSS and InSAR as shown later.

In this chapter, we describe how the SWD-difference maps at each SAR overpass time are estimated using the method of least squares inversion. We present a new method that
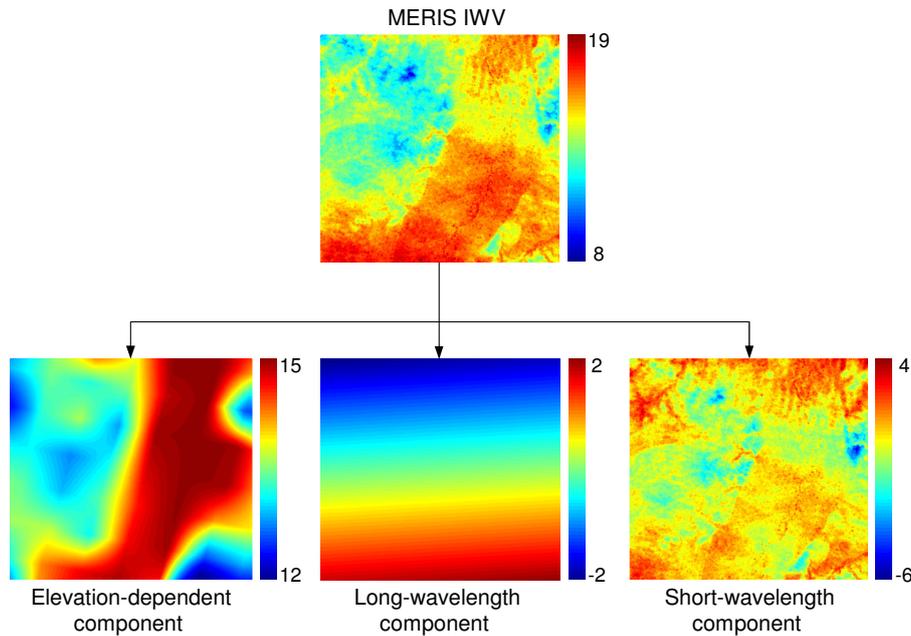
Figure 6.1: A map of IWV is a sum of three components: the first and the largest is the elevation-dependent IWV, the second is a long-wavelength component resulting from the change in air pressure and temperature over distance, and the third is the short-wavelength component due to the turbulently mixing water vapor.

uses GNSS observations to estimates the wet delay component missing due to interferograms formation. We combine these estimates of the missing delay at each PS pixel with those extracted from the interferograms to derive absolute maps of SWD at the density of PS points. We then compare the SWD absolute maps estimated using our method with IWV maps from MERIS.

6.1 Extracting wet delay maps at SAR acquisition time

The SWD-difference maps retrieved from the interferograms are relative measurements, as they measure the difference between the wet path delay at two acquisition times. Our goal is to reconstruct the fields of SWD corresponding to each SAR observation time. By examining PSI observations, there exist N-1 neutrospheric difference maps (interferograms) and N unknown neutrospheric phase maps. Then this system is underestimated. To overcome this problem, we need to either rely upon external data or to make an assumption. It is challenging to have external data at the high resolution of PSI, so it is more convenient to apply least squares inversion (LSQ) by adding a constraint. We used the "zero mean" assumption, in which the average of the neutrospheric phases over time is assumed zero ($\sum_{i=1}^{N} \frac{1}{N}\phi_{i,neu} = 0$). Note that $\phi_{i,neu}$ is the partial neutrospheric phase, where a large signal is eliminated by building interferograms. This assumption is generally used to derive observation residuals

corresponding to each point from difference measurements as in [Liu et al., 2009; Luo et al., 2008]. We construct a linear model to relate $N-1$ observations to N unknowns for each PS pixel such as

$$
\begin{bmatrix}
-1 & 0 & \cdots & +1 & 0 & \cdots & 0 \\
0 & -1 & \cdots & +1 & 0 & \cdots & 0 \\
\vdots & \vdots & \cdots & +1 & \vdots & \cdots & \vdots \\
0 & 0 & \cdots & +1 & 0 & \cdots & -1 \\
\beta_1 & \beta_2 & \cdots & \beta_M & \cdots & \beta_{N-1} & \beta_N
\end{bmatrix}
\begin{bmatrix}
\phi_{1,neu} \\
\vdots \\
\phi_{M,neu} \\
\vdots \\
\phi_{N,neu}
\end{bmatrix}
=
\begin{bmatrix}
\phi_{1M,neu} \\
\vdots \\
\vdots \\
\phi_{(N-1)M,neu} \\
0
\end{bmatrix}
\tag{6.2}
$$

with $\beta_i = \frac{1}{N}$. The equation can be rewritten in a more compact form as:

$$
\mathbf{AB} = \mathbf{C} \tag{6.3}
$$

A: Design matrix of size $N \times N$, N is the number of SAR scenes

B: Unknown phases of size $N \times K$, K is the number of PS points

C: Neutrospheric phase observations of size $N \times K$

The subscript M in eq. (6.2) refers to the "master". The solution for eq. (6.3) is found by applying LSQ inversion,

$$
\mathbf{B} = (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{C} \tag{6.4}
$$

The LSQ inversion is applied for all PS observations at once to reduce errors by taking the advantage of spatial correlations between the PS observations. In **B** are the estimates of the neutrospheric phase partial maps at each SAR overpass time. The phase is converted into SWD, in millimeters, using eq. (5.22). In **Figure 6.2** are shown three maps of the SWD residuals at each SAR overpass time, in (c) is the SWD-partial map corresponding to the "master". Compared with the IWV maps available from MERIS, the reconstructed SWD-partial maps at SAR acquisition time agree in amplitude and spatial variations with MERIS IWV maps, as will be shown later in this chapter.

We examined the dependence of the SWD-partial maps from PSI and temporal-differences of GNSS ZWD on the surface elevation in the study region. We found that the dependence of the SWD differences on the altitude depends on the SWD value at the "master" and "slave" acquisitions. Generally, we can say that the SWD decreases linearly as the surface elevation increases, but by building differences, the correlation of the SWD differences with the surface elevation will be poor or even zero as observed from the plots in **Figure 6.3**. The SWD difference maps extracted from the interferograms have a mean zero and the region of URG is generally characterized by smooth topography; therefore most of the elevation-dependent signal is eliminated when building interferograms or by reducing the topographic phase. On

(a) SWD: 05.23.2007   (b) SWD: 09.05.2005   (c) SWD: 06.27.2005
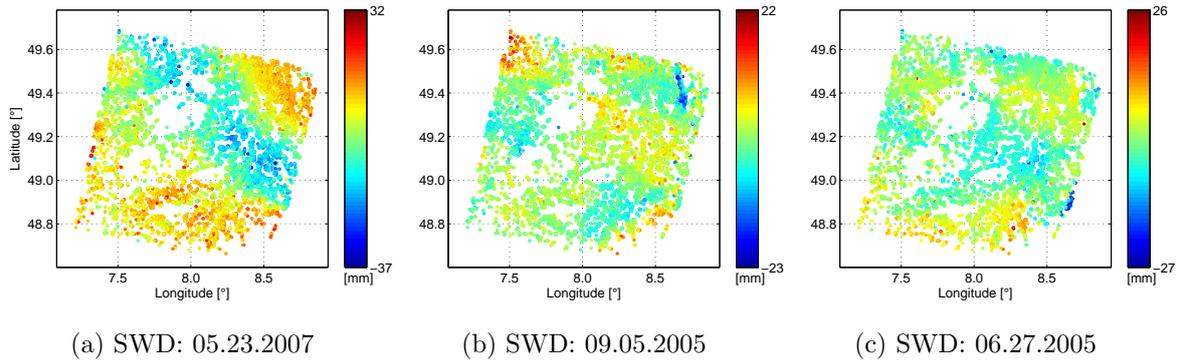
Figure 6.2: Partial maps of SWD at SAR overpass times. The maps are retrieved by least squares inversion of SWD-difference maps derived from the interferograms. To overcome rank deficiency, we added a constraint by assuming the temporal mean of the partial SWD observations is zero.

the other hand, the neutrospheric long-wavelength signal cannot be distinguished from orbital ramps, so this signal is also eliminated during PSI data processing. We consider the SWD-partial maps to measure the short-wavelength signal of the SWD, while the elevation-dependent and the long-wavelength components are missing. For the purpose of reconstructing maps of the absolute SWD along the signal path, we estimate the missing vertically-stratified and long-wavelength SWD components using GNSS ZWD observations as presented in the following section.

(a) CC: -0.0418   (b) CC: -0.1445   (c) CC: 0.3131
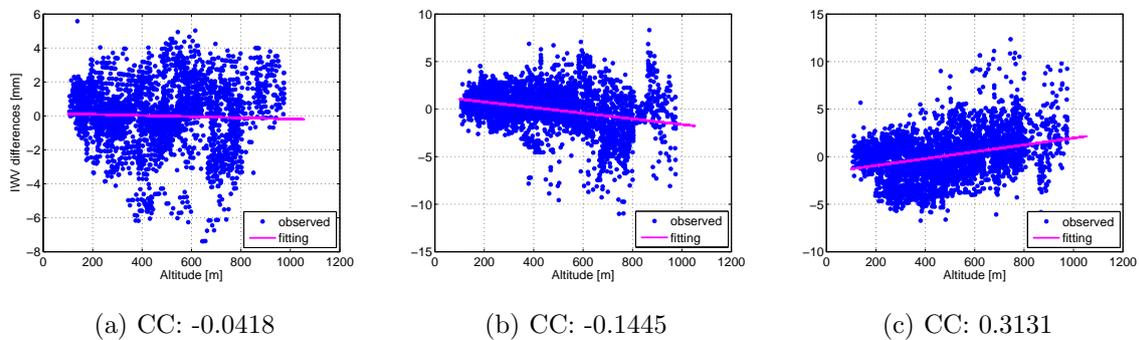
Figure 6.3: Dependence of the IWV-difference maps on the surface elevation.

6.2 combination of SWD observations from PSI and GNSS

In this section, we present a new approach to combine the SWD-difference maps extracted from PSI data and the estimates from GNSS data. First, the eliminated components are defined and modeled using GNSS data, then the results are validated using the IWV maps measured by MERIS.

6.2.1 Modeling missing SWD components using GNSS

In this section, we focus on the elevation-dependent (vertically-stratified) ZWD and the long-wavelength signals of the ZWD derived from GNSS observations. Davis et al. [1993] and Elosegui et al. [1998] described the dependence of the wet refractivity, and hence the wet delay, on the surface elevation by an exponential law. Onn and Zebker [2006] suggested a more general model,

$$ZWD_{st}(z) = C \exp(-\alpha z) + z\alpha C \exp(-\alpha z) + \Delta L_{min} \tag{6.5}$$

where ZWD_{st} is the vertically-stratified wet delay, C, and α are the model parameters, and ΔL_{min} is the minimum wet delay. This model fits better to the ZWD measurements at low altitudes. By inspecting vertical profiles of the relative humidity and the water vapor pressure generated by WRF, we observed that the model in eq. (6.5) fits well to the profiles of water vapor partial pressure, as a measure for the wet delay, particularly at low altitudes. In **Figure 6.4**, we show the wet delay measurements from MERIS and two models fitted to the data, the first is the one in eq. (6.5) and the second is a regression line fitted using least squares. The linear regression and the model of eq. (6.5) show similar fit to the data due to the high density of observations, while in the case of GNSS data where the observations are spatially sparse, the model of eq. (6.5) is more representative, in particular at low altitudes. Nonetheless, the linear model can be used when no conversion towards a solution is possible using the model in (6.5).
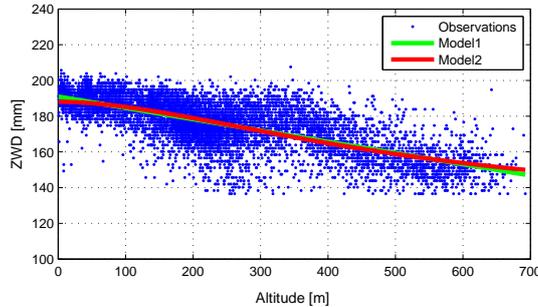
Figure 6.4: IWV observations from MERIS, converted into ZWD values, as a function of surface elevation. Model1 is a linear regression and Model2 is the model of eq. (6.5).

In fitting the model of eq. (6.5) to the ZWD measurements at the 10 GNSS sites, we need to estimate the model parameters C and α and ΔL_{min}. To measure the goodness of fit between the model and the observations, we calculated the reduced chi-squared statistic from the following formula [Onn and Zebker, 2006],

$$\chi^2_{gnss} = \frac{1}{\nu} \sum_i \frac{(ZWD^{gnss}(z_i) - ZWD_{st}(z_i))^2}{\sigma^2_{gnss,wet}} \tag{6.6}$$

ν is the degree of freedom, such that if there exist 10 GNSS observations and the model requires the estimation of three parameters, then we have 7 degrees of freedom. $\sigma_{gnss,wet}$ is the GNSS ZWD estimation error of 5.048 mm, which is computed using MERIS observations as an accepted reference due to the lack of a more accurate source of ground truth data.

If the GNSS sites are located close together at different altitudes, the variation in the observed ZWD values is most likely due to the stratification of water vapor. For the purpose of correcting the atmospheric phase, Onn and Zebker [2006] illustrated the ZWD as a function of the altitude of 29 GPS sites located within an area of 60 km $\times$ 60 km with a very good fit to the model of eq. (6.5). In our case study, however, the shortest baseline between two GNSS sites is greater than 20 km and the number of sites is limited to 10. Because of this relatively sparse distribution of the GNSS sites, the ZWD will show more variations due to local effects of turbulent mixing around the site and effects of neutrospheric anisotropy. In order to retrieve the non-turbulent signal of the ZWD, we estimated two components. First, we estimated the elevation-dependent signal by fitting the model of eq. (6.5) to the data. Second, we analyzed the residuals (the difference between the actual observations and the estimated model) to check if they contain any deterministic long-wavelength signal (ramp), which is considered as a correction term to account for the variations in pressure and temperature over large spatial scales. The ZWD non-turbulent signal is the sum of the elevation-dependent signal and the ramp. The parameters estimated for the model of eq. (6.5) and the ramp are used to compute a value for, respectively, the vertically-stratified ZWD and the long-wavelength component at any (x, y, z) location, hence the estimated ZWD component is written as:

$$ZWD_{nt}(x, y, z) = ZWD_{st}(x, y, z) + \Delta_{corr}(x, y) \tag{6.7}$$

$ZWD_{nt}(x, y, z)$: The modeled non-turbulent ZWD using GNSS observations [mm]

$ZWD_{st}(x, y, z)$: Vertically-stratified ZWD [mm]

$\Delta_{corr}(x, y, z)$: Correction term estimated from the residuals [mm]

We use the model of eq. (6.7) to obtain the complementary signals for the SWD-partial maps obtained from PSI, which are required to derive absolute SWD maps at each SAR acquisition time. In **Figure 6.5** (b, e), we show two maps of the non-turbulent SWD estimated at the SAR overpass times on 06.27.2005 and 09.05.2005. In **Figure 6.5** (a, d) are shown the GNSS SWD observations (ZWD mapped to the SAR LOS) at two SAR acquisition times, against sites altitudes. The vertically-stratified ZWD values at the PS locations are calculated using the parameter estimates given in **Table 6.1**. From the examples depicted in **Figure 6.5**, we observe that the SWD measurements on 06.27.2005 are less affected by variations due to local neutrospheric effects, which makes the stratified delay dominant. Therefore, the fitting of the model in eq. (6.5) achieves a goodness-of-fit much better than that for the example in (d). However, since we use the residuals to estimate a correction signal, the non-turbulent SWD can be accurately modeled. If we examine the maps in (c, f) for the ramps, we observe a larger

signal estimated for the example in (d) to compensate the worse goodness-of-fit.

Day	C [mm]	α [km^{-1}]	ΔL_{min} [mm]	χ^2
06.27.2005	48.7805	4.1342	151.9903	5.6705
09.05.2005	25.7566	6.0933	76.2429	21.4359

Table 6.1: Parameter estimates and goodness-of-fit (χ^2) for the elevation-dependent SWD from GNSS observations at SAR acquisition time.

(a) GNSS SWD   (b) Non-turbulent SWD   (c) Long-wavelength SWD

(d) GNSS SWD   (e) Non-turbulent SWD   (f) Long-wavelength SWD
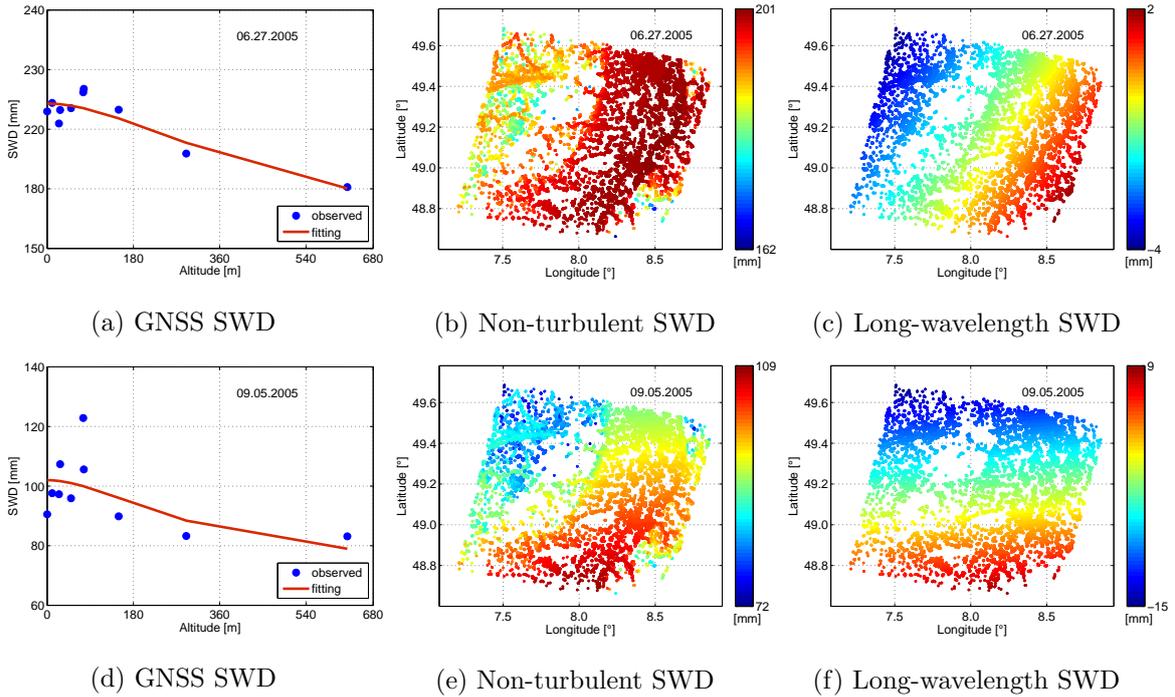
Figure 6.5: GNSS ZWD as a function of site altitudes at SAR overpass times on 06.27.2005 (a) and 09.05.2005 (d). The red curves represent the least squares fit of the model in (6.5) to the ZWD observations. In (b, e) are shown the non-turbulent ZWD (an elevation-dependent component and a ramp) calculated based on the fitting parameter estimates in **Table 6.1** at the PS locations. The ramps are estimated from the residuals are shown in (c, f).

From MERIS IWV maps and PSI SWD-difference maps we observed a special neutrospheric structure on 04.23.2007, where the neutrosphere was more active in the southern area and the water vapor content was larger due to turbulent mixing in the neutrosphere (see **Figure 2.3** (b)). The GNSS ZWD observations agree with MERIS and PSI and show also higher values at the sites located in the southern part of the SAR image (see the four observations above the red curve in **Figure 6.6** (a)). Because of the spatial variability, dominated by turbulent mixing of the ZWD from south to north, the best fit of model (6.5) to the observations underestimates the stratified delay mainly at low altitudes. The parameter estimates are summarized in **Table 6.2** and the model fitting is shown in **Figure 6.6** (a, c). In the first row of the table, where the model fitting was done based on observations from 10 sites, the estimate for α is close to zero, which is not realistic, and the χ^2 value is significantly large. In (b) is

Day	C [mm]	α [km^{-1}]	ΔL_{min} [mm]	χ^2
04.23.2007 (1)	83.1995	0.0036	61.2255	44.6014
04.23.2007 (2)	43.8800	3.0216	78.9312	4.8363

Table 6.2: Parameter estimates and goodness-of-fit (χ^2) for the elevation-dependent SWD from GNSS observations on 04.23.2007. (1) means that observations from 10 sites were used, while in (2) the observations from only 4 sites were used.

shown the non-turbulent SWD extrapolated at PS locations, where the spatial variations of the map follow the estimated ramp since the stratified SWD component has almost a constant value over the entire image. To improve the modeling of the vertically stratified ZWD, we made the model fitting based only on four observations (**Figure 6.6** (c)). The estimate of α is more realistic and the goodness-of-fit improved. This can be automated by removing one GNSS site each time and check the χ^2 value and keep iterating until the χ^2 resides. The ramp was estimated using the residuals from all ten sites. In this special case, the importance of analyzing the residuals after fitting the model in (6.5) can be clearly noticed. In the first case, where observations from all sites are included, the SWD will be almost constant over the entire map if we do not consider the residuals, which is intuitively not correct.

6.2.2 Combining PSI and GNSS SWD maps

The SWD-partial maps extracted from interferograms are combined with the non-turbulent SWD maps modeled using GNSS SWD observations to build maps of absolute SWD. **Figure 6.7** summarizes the method developed to derive absolute SWD maps by combining path delay from GNSS and PSI observations. These maps are compared with the IWV maps received from MERIS. IWV maps are first converted into the corresponding SWD values and mapped to the radar LOS using eq. (2.29). In **Figure 6.8**, the absolute SWD maps derived using our method, maps received from MERIS, and difference maps are shown. The spatial CC between both data and the RMS, mean, and standard deviation (SDEV) of the difference maps are summarized in **Table 6.3**. There exist strong spatial correlations between absolute SWD maps derived from PSI and GNSS and those from MERIS. The difference maps follow a Gaussian distribution with a mean close to zero and an SDEV of at most 6 mm. We show in **Table 6.3** also the parameters in case we did not analyze the GNSS residuals after fitting the model in (6.5). The spatial correlation on 06.27.2005 (master) is low compared to the other examples. We found that the ramp (gradient) estimated from the residuals does not agree in direction with that observed in MERIS. At this point, we do not have an independent source to conclude which of the maps is more accurate.

We should expect that maps of the absolute SWD derived using our method and those from MERIS do not show perfect correlation. This is because MERIS measures the content of IWV in vertical columns of 300×300 m^2 base, which is then mapped to the radar LOS. The radar, however, transmits signal at a look angle (off-nadir angle) of about 23°. Due to
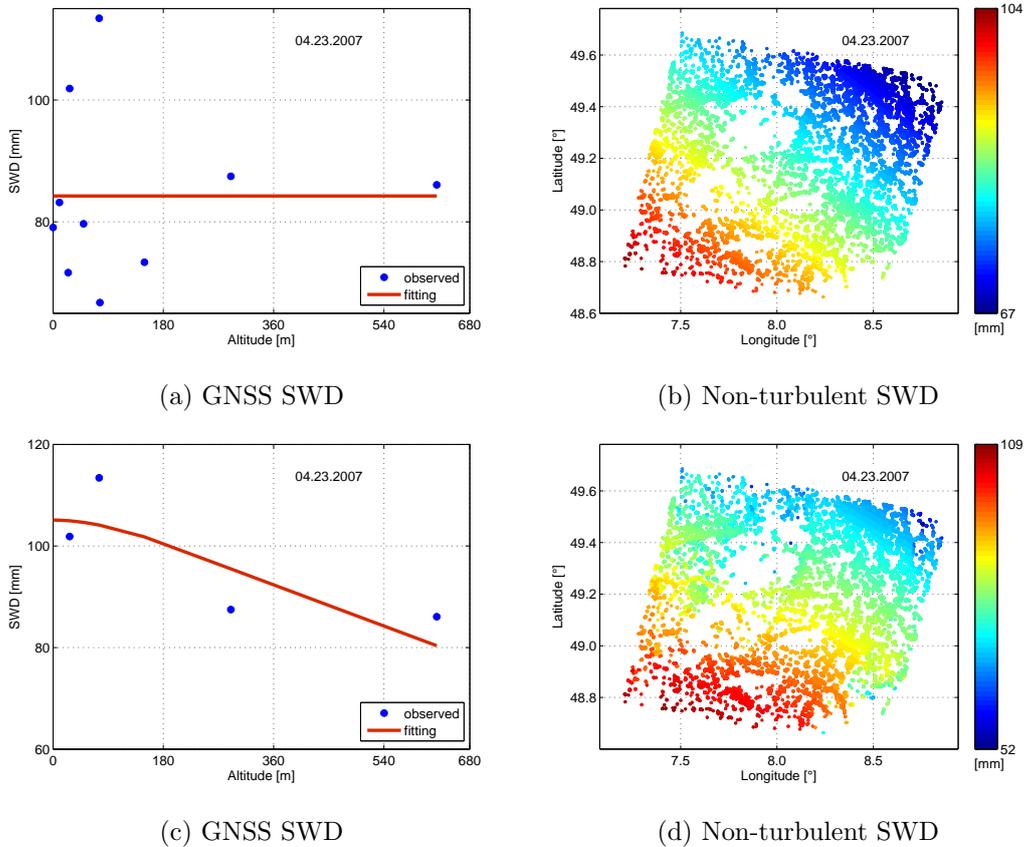
Figure 6.6: (a) GNSS ZWD as a function of site altitudes at SAR overpass times on 04.23.2007 and the least squares fit of the model in (6.5) to the observations. (c) The fit of eq. (6.5) (in red) is done using only four observations. In (b, d) are shown the non-turbulent ZWD (elevation-dependent component and a ramp) extrapolated based on the fitting parameters estimates at the PS locations.

the neutrospheric anisotropy, the IWV experienced along the signal path can differ from that measured in the vertical direction. Also, GNSS measures the average ZWD within conical sections of the neutrosphere and these measures are available at sparse sites. The modeling of the missing signals can be improved by increasing the density of the GNSS sites. All this results in differences between IWV maps measured by MERIS and those derived in this work.

6.3 Summary

In this chapter, we applied LSQ inversion to reconstruct SWD-partial maps corresponding to each SAR acquisition time. Since we had a rank-deficient system of equations, we added a constraint by assuming the weighted mean of the SWD residuals over time is zero. The SWD-partial maps derived from PSI contain the wet delay residuals corresponding to the short-scale SWD, while the largest part of the wet delay was reduced when building interferograms. To reconstruct the missing signals, we made a combination of SWD observations estimated from
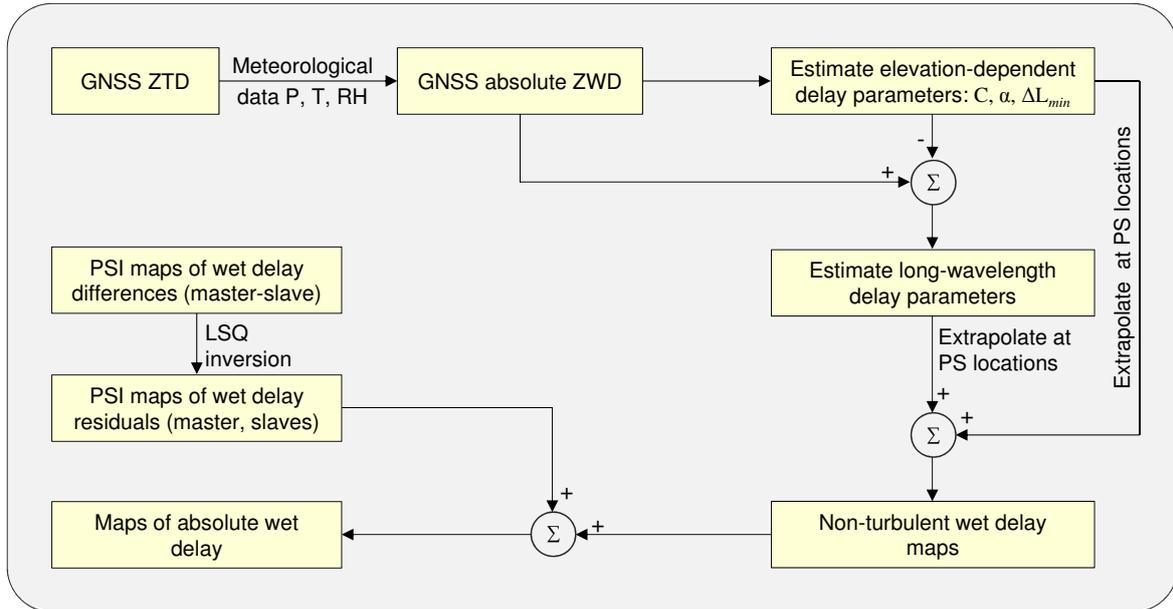
Figure 6.7: Method developed to derive absolute maps of the SWD by combining neutrospheric delay observations from GNSS and PSI.

GNSS and PSI. From GNSS observations we estimated the elevation-dependent wet delay (vertically-stratified SWD), and the parameter estimates were used to extrapolate values at the PS locations. Because the GNSS sites are relatively sparse in space, we used the residuals (SWD - vertically-stratified SWD) to estimate a ramp accounting for long-wavelength variations. We then combined the short-scale SWD component from PSI with the non-turbulent SWD (elevation-dependent and ramp) to build maps of the absolute SWD along the path from the radar to the ground. We compared absolute SWD maps derived from combining PSI and GNSS observations with absolute SWD maps from MERIS (5 maps available). The results show strong spatial correlation up to 92.2 %. The difference maps follow a Gaussian distribution with mean and standard deviations depicted in **Table 6.3**.

	Without residuals analysis				With residuals analysis			
Time	CC	RMS [mm]	MEAN [mm]	SDEV [mm]	CC	RMS [mm]	MEAN [mm]	SDEV [mm]
06.27.2005	0.7354	6.3761	1.2256	6.2575	0.7505	6.0997	0.4361	6.0837
09.05.2005	0.5630	10.7338	6.7192	8.3710	0.8854	5.3577	2.3377	4.8210
04.23.2007 (1)	0.7309	9.9048	-3.5767	9.3191	0.9135	6.5413	-3.3733	5.6041
04.23.2007 (2)	0.6885	20.9622	18.5578	9.7484	0.9216	5.7584	-2.6767	5.2092

Table 6.3: Spatial correlation coefficients between SWD maps obtained from combining PSI and GNSS observations and MERIS. The RMS, mean, and the STD of the difference maps are summarized. 04.23.2007 (1) uses the model fitting in **Figure 6.6** (a) and 04.23.2007 (2) uses the model in (c).

(a) Combined PSI-GNSS   (b) MERIS   (c) Differences
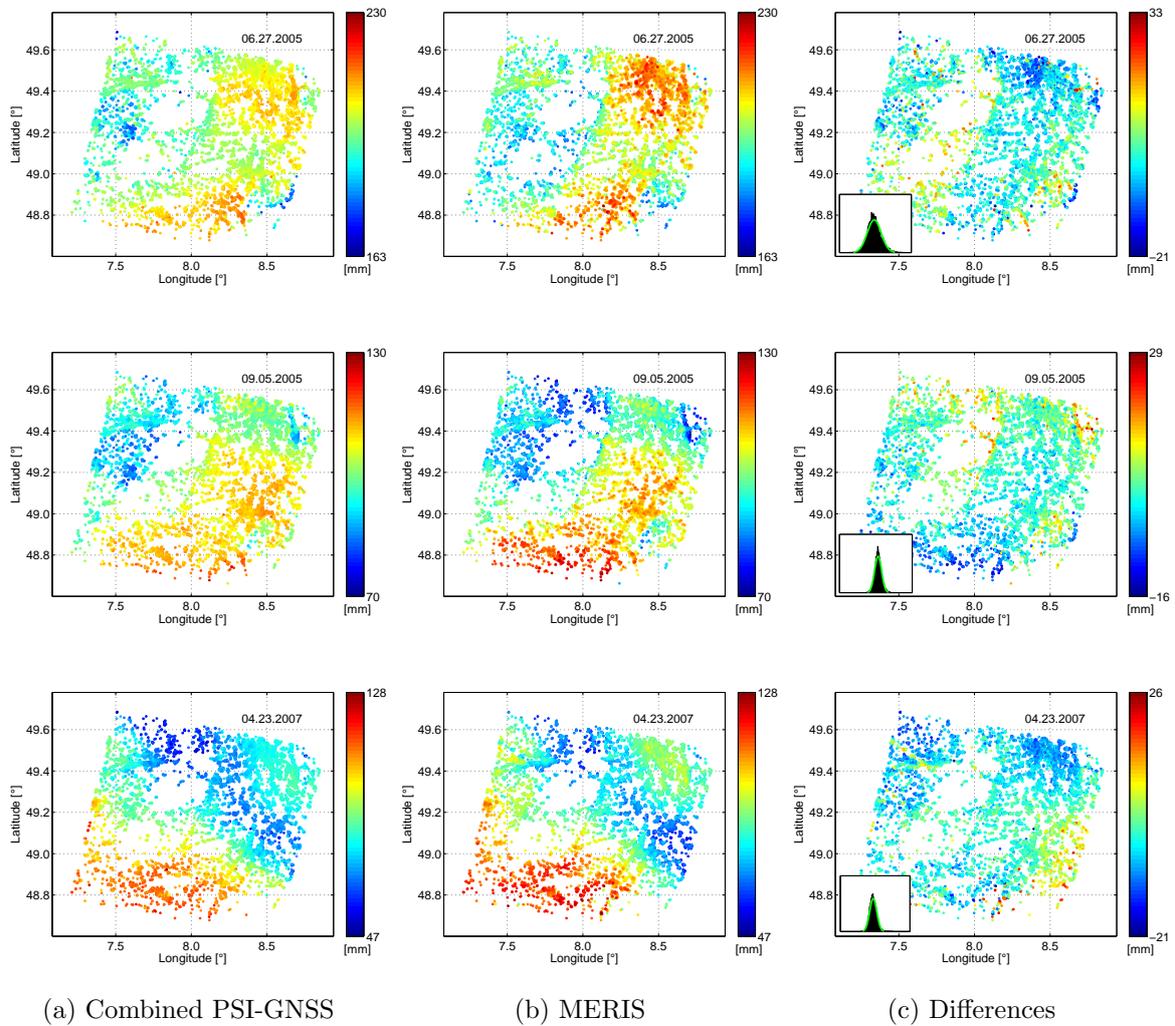
Figure 6.8: Absolute SWD maps from (a) combining GNSS and PSI observations and (b) MERIS at three SAR overpass times, 06.27.2005, 09.05.2005, and 04.23.2007. The difference maps between the observations in (a) and (b) are shown in (c).

7 Fusing remote sensing-based IWV maps and data from atmospheric models

The topic of data fusion has been insistent with the rapid increase of remote sensing data that are spatially, temporally, and spectrally inhomogeneous. Since the 1990's, the amount of data available for monitoring the Earth and its atmosphere by the satellites, such as GNSS and InSAR, is growing in a rapid and continuous way. NWP models have also been increasingly exploited to generate regular maps for quantifying different parameters of the atmosphere, for example air pressure, temperature and IWV content. These data sources could supply different representations of the geophysical quantity, which, if combined, can give a more complete representation of the ongoing physical process. Data fusion is a process that attempts to produce improved knowledge of the interesting object by integrating data from multiple sources. The process aims at synergetic exploitation of data that can be redundant or complementary to provide complete, accurate, and consistent information. The applied data fusion method should also be able to measure the quality of the output resulting from the fusion.

Data fusion is a wide topic of research in many scientific fields. In this review, we put the focus on the methods that are relevant for remote sensing data used for observing the Earth and the atmosphere. Geographic information systems packages contain several data fusion methods that combine data using weighted linear functions. The data sets are interpolated into a common grid by smoothing, moving-window averaging, IDW, or more complicated methods. The interpolation maps from all data sources are then weighted such that the weights are often related to the importance associated with each data source [Drobne and Lisec, 2009]. These methods define the weights for each output point by assuming that this point is independent from other points, which is usually not correct for geophysical phenomena. Also, the interpolation methods assume that the underlying process is spatially continuous although they may not be, which can result in a bias in the predicted value. Furthermore, making the interpolation and the combination independently might not provide the optimal solution. Therefore, data fusion approaches that are based on statistical analysis of the data have been developed [Braverman et al., 2009; Nguyen, 2009]. For spatial data (e.g., geophysical quantities) where close observations correlate more than distant ones, a covariance function is exploited as an important source of information to obtain predictions based either on single or multiple sets. For remote sensing data combination, the fusion methodology should consider the problems of huge size of the input data, probably incompatible supports, presence of noise and gaps. An approach that comprehensively solves these problems is the spatial statistical

data fusion (SSDF) strategy presented in [Braverman et al., 2009; Nguyen, 2009], which we use in this work for data fusion. This method is based on extending the FRK approach proposed by Cressie and Johannesson [2008] and described in Chapter 4 for data fusion of multiple data sets.

In this study, we are interested in deriving accurate and spatially-continuous SWD fields. We introduced methods to quantify the SWD (as a measure for IWV) from different remote sensing systems in the previous chapters. SWD fields obtained by combining GNSS and PSI show discontinuities in regions of low coherence as in forests, vegetated areas, and over water surfaces. IWV maps generated by WRF are available at continuous grids, but their spatial resolution and the ability of the model to detect small atmospheric patterns are still limited. SWD and WRF IWV maps are also different in their spatial support such that the SWD maps are at point-level support, while WRF provides block data. Also, WRF might generate maps that do not represent the real state of the atmosphere due to model limitations. When comparing WRF maps of IWV with the IWV maps from MERIS, there are examples that show good spatial correlation between both maps; however, there are scenarios where the simulated map demonstrates low correlation with the observed map. Two examples are depicted in **Figure 7.1**, the first (a, b) shows WRF and MERIS IWV maps that are strongly correlated with a CC of 0.8, while the second example (c, d) exhibits a lower spatial correlation (CC=0.71) between the maps. The IWV map from WRF has artifacts and an unexpected discontinuity in the IWV values in the lower left corner, which reduces the correlation between the maps. The goal in this chapter is to exploit the complementary properties of WRF and PSI-GNSS data to accurately estimate continuous 2D fields of the SWD. In the following, we present the applied SSDF method and explain the procedure required to obtain the predictions and the uncertainties. We also present comparisons of the prediction maps based on single data sets and those received from data fusion, and a validation of the results using IWV maps from MERIS.

In Section 7.1, we present the method of SSDF. We describe the FRK covariance model used for each data set depending on their support in Section 7.2. The matrix inversion method is described in Section 7.3. In Section 7.4, we describe the methods we used to estimate the covariance parameters required for the covariance model. We present examples for applying the SSDF in Section 7.5 and the conclusions are presented in Section 7.6.

7.1 Spatial statistical data fusion via fixed-rank kriging

Spatial statistical data fusion (SSDF) is a methodology for combining spatial data to optimally estimate the true values of the quantity of interest and the corresponding uncertainties at any predefined grid [Braverman et al., 2009; Nguyen, 2009]. The input data sets might have different characteristics, for example, spatial resolutions, continuity, and noise. Braverman et al. [2009] applied the SSDF method to estimate the amount of carbon dioxide (CO_2) in the

(a) WRF: 06.27.2005

(b) MERIS: 06.27.2005

(c) WRF: 09.05.2005
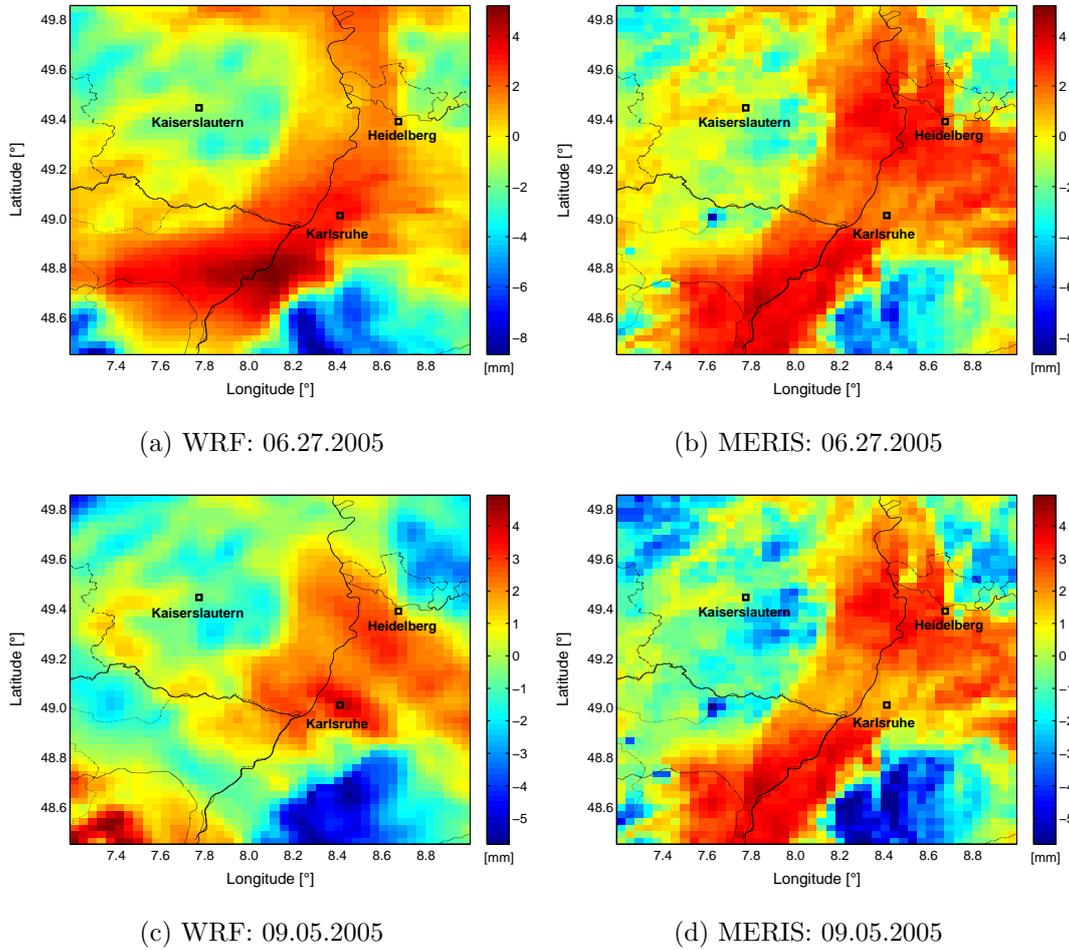
(d) MERIS: 09.05.2005

Figure 7.1: Maps of IWV content as received from MERIS and WRF, where a linear trend is subtracted from each map. The upper data are received on 06.27.2005 (SAR overpass), while the lower data on 09.05.2005. Gaussian averaging is applied to scale the MERIS data at WRF resolution, 3×3 km^2. The spatial CC between the upper maps is 0.8 and the CC is 0.71 for the lower.

lower layer of the Earth's atmosphere on a global scale. An advanced version of the algorithm that investigates the temporal and spatial characteristics to obtain the predictions is called the space-time data fusion (STDF) algorithm, which was developed by Braverman et al. [2011] and Katzfuss and Cressie [2011a]. In this work, we focus on SWD fields as a measure for water vapor concentration in the atmosphere. SWD maps derived in the previous chapter are available at SAR overpass time every couple of days or weeks. For time periods longer than one day, the SWD maps have no temporal correlation [Zebker et al., 1997]. Therefore, we are interested in combining the data spatially at the time they are available, thus the SSDF technique is appropriate.

SSDF is applied to statistically combine SWD fields from remote sensing data (PSI and GNSS) and WRF. Fusion is attainable, in spite of all differences in the input data, by constructing a statistical model that describes the structure of the input data and relates the

input data to the predictions. SSDF relies upon the assumptions that the underlying field has a constant mean and the covariance structure follows the FRK model presented in Section 4.2.2.3. For areal-level data, it is assumed that the value of the geophysical parameter within a certain cell represents the average of all values as if observed at many points within that cell, i.e.,

$$Z(B_i) = \frac{1}{|B_i|} \int_{B_i} Z(\mathbf{s})d\mathbf{s} \tag{7.1}$$

These assumptions hold for the water vapor content in the atmosphere. In Chapter 4, we presented how the method of OK finds the best estimator by minimizing the MSPE between the observations and the true process. We also presented the FRK method to solve for the lack of scalability of OK and to account for the data support. In this chapter, we present the method that extends the use of FRK for data fusion and we present the strategy to derive the true process from two data sets.

Let the underlying process $Y(\mathbf{s})$ to be estimated at the location $\mathbf{s}$ from the two data sets $\mathbf{Z}_1$ and $\mathbf{Z}_2$ with the size N_1 and N_2, respectively. $\mathbf{Z}_1$ contains the point-level SWD observations and $\mathbf{Z}_2$ contains the block-level WRF data, which have the structure in (7.1). The data in $\mathbf{Z}_1$ and $\mathbf{Z}_2$ are related to the true process as in eqs. (4.3) and (4.7), respectively. In Chapter 4, we demonstrated the use of the FRK algorithm to infer the true SWD fields from single data sets at a regular grid by minimizing the MSPE. In the same manner, the SSDF method attempts to estimate the true process from the combined data by minimizing the MSPE. The estimator $\hat{Y}(\mathbf{s})$ at the location $\mathbf{s}$ from a single data set is obtained from (4.16), similarly $\hat{Y}(\mathbf{s})$ is obtained from two data sets using the following formula:

$$\hat{Y}(\mathbf{s}) = \mathbf{a}_1'\tilde{\mathbf{Z}}_1 + \mathbf{a}_2'\tilde{\mathbf{Z}}_2 \tag{7.2}$$

where $\mathbf{a}_1$ and $\mathbf{a}_2$ are the fusion weighting coefficients, and $\tilde{\mathbf{Z}}_1$ and $\tilde{\mathbf{Z}}_2$ are detrended data sets. To ensure that the estimator is unbiased with respect to the true process, we search for the weights $(\mathbf{a}_1, \mathbf{a}_2)$ that solve the following constrained minimization problem:

$$\min_{\mathbf{a}_1,\mathbf{a}_2} \quad \mathrm{E}\left\{(\hat{Y}(\mathbf{s}) - Y(\mathbf{s}))^2\right\} \quad \text{subject to} \quad \mathrm{E}\{\hat{Y}(\mathbf{s})\} = \mathrm{E}\{Y(\mathbf{s})\} \tag{7.3}$$

under the constraint in (7.3), the MSPE $\mathrm{E}\left\{(\hat{Y}(\mathbf{s}) - Y(\mathbf{s}))^2\right\}$ equals the variance, i.e.,

$$
\begin{aligned}
\mathrm{E}\left\{(\hat{Y}(\mathbf{s}) - Y(\mathbf{s}))^2\right\} &= \mathrm{var}\left(\mathbf{a}_1'\tilde{\mathbf{Z}}_1 + \mathbf{a}_2'\tilde{\mathbf{Z}}_2 - Y(\mathbf{s})\right) \\
&= \mathbf{a}_1' \, \mathrm{var}(\tilde{\mathbf{Z}}_1)\mathbf{a}_1 + \mathbf{a}_2' \, \mathrm{var}(\tilde{\mathbf{Z}}_2)\mathbf{a}_2 + \mathrm{var}(Y(\mathbf{s})) \\
&+ 2\mathbf{a}_1' \, \mathrm{cov}(\tilde{\mathbf{Z}}_1, \tilde{\mathbf{Z}}_2)\mathbf{a}_2 - 2\mathbf{a}_1' \, \mathrm{cov}(\tilde{\mathbf{Z}}_1, Y(\mathbf{s})) \\
&- 2\mathbf{a}_2' \, \mathrm{cov}(\tilde{\mathbf{Z}}_2, Y(\mathbf{s}))
\end{aligned}
\tag{7.4}
$$

Let $\mathbf{\Sigma}_{ii} = \mathrm{var}(\tilde{\mathbf{Z}}_i)$, $\mathbf{\Sigma}_{ij} = \mathrm{cov}(\tilde{\mathbf{Z}}_i, \tilde{\mathbf{Z}}_j)$, and $\mathbf{c}_i = \mathrm{cov}(\tilde{\mathbf{Z}}_i, Y(\mathbf{s}))$ be the covariance functions required to solve eq. (7.3), which are specified in Section 7.2. Let the mean of the true process be μ, then the unbiasedness constraint in (7.3) implies that

$$
\begin{aligned}
\mathrm{E}\{Y(\mathbf{s})\} = \mu &= \mathrm{E}\left\{\mathbf{a}_1'\tilde{\mathbf{Z}}_1 + \mathbf{a}_2'\tilde{\mathbf{Z}}_2\right\} \\
\mu &= \mathbf{a}_1'\mathbf{1}_{N_1} \cdot \mu + \mathbf{a}_2'\mathbf{1}_{N_2} \cdot \mu \\
0 &= \mathbf{a}_1'\mathbf{1}_{N_1} + \mathbf{a}_2'\mathbf{1}_{N_2} - 1
\end{aligned}
\tag{7.5}
$$

Then the minimization problem can be solved using the Lagrange multipliers [Bertsekas, 1996]. The Lagrangian function L for (7.4) under the constraint in (7.5) is written as

$$
\begin{aligned}
L &= \mathbf{a}_1'\mathbf{\Sigma}_{11}\mathbf{a}_1 + \mathbf{a}_2'\mathbf{\Sigma}_{22}\mathbf{a}_2 + 2\mathbf{a}_1'\mathbf{\Sigma}_{12}\mathbf{a}_2' - 2\mathbf{a}_1'\mathbf{c}_1 - 2\mathbf{a}_2'\mathbf{c}_2 \\
&\quad + 2m(\mathbf{a}_1'\mathbf{1}_{N_1} + \mathbf{a}_2'\mathbf{1}_{N_2} - 1)
\end{aligned}
\tag{7.6}
$$

where $\mathbf{1}_{N_i}$ is a vector with all entries one and a length N_i, and the Lagrange multiplier is denoted by m. The last term of L accounts for the unbiasedness constraint in (7.3). Differentiating L with respect to $\mathbf{a}_1, \mathbf{a}_2, m$ and assigning the results to zero, we get in the following system of equations,

$$
\begin{bmatrix}
\mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} & \mathbf{1}_{N_1} \\
\mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22} & \mathbf{1}_{N_2} \\
\mathbf{1}_{N_1}' & \mathbf{1}_{N_2}' & 0
\end{bmatrix}
\begin{bmatrix}
\mathbf{a}_1 \\
\mathbf{a}_2 \\
m
\end{bmatrix}
=
\begin{bmatrix}
\mathbf{c}_1 \\
\mathbf{c}_2 \\
1
\end{bmatrix}
\tag{7.7}
$$

then, we solve for $\mathbf{a}_1, \mathbf{a}_2$, and m,

$$
\begin{bmatrix}
\mathbf{a}_1 \\
\mathbf{a}_2 \\
m
\end{bmatrix}
=
\begin{bmatrix}
\mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} & \mathbf{1}_{N_1} \\
\mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22} & \mathbf{1}_{N_2} \\
\mathbf{1}_{N_1}' & \mathbf{1}_{N_2}' & 0
\end{bmatrix}^{-1}
\begin{bmatrix}
\mathbf{c}_1 \\
\mathbf{c}_2 \\
1
\end{bmatrix}
\tag{7.8}
$$

This is a straight forward minimization problem and the solution can be found from eq. (7.8), but the open question is how to compute the solution if the matrices $\mathbf{\Sigma}_{ii}$ and $\mathbf{\Sigma}_{ij}$ are extremely huge, which is the case for most remote sensing data. Strictly speaking, two essential points should be considered in eq. (7.8): the first is the matrix inversion for massive data sets where the size of the covariance matrix is huge, and the other is the change of support problem between input data and the predictions. The covariance matrices should be modeled such that they would allow data prediction to any level of aggregation. In Chapter 4, we applied the FRK covariance model suggested by Cressie and Johannesson [2008] to solve the two problems; however, the model was applied for single data sets. In this chapter, we apply the generalized SSDF model presented in [Braverman et al., 2009; Nguyen, 2009] for multiple data sets. We will describe in Section 7.3 how the matrix inversion is achieved. In the next section, we

present the strategy to obtain the SSDF covariance terms apparent in eq. (7.8).

7.2 Covariance model for multiple data sets

In the previous section, we described the method of estimating the true process from multiple data sets and how to obtain the fusion weighting coefficients. In order to obtain the predictions, the covariance functions for different data sets and the covariance function between the observations and the true process at the prediction locations are essential. In the OK method, a semivariogram is estimated from the data and a semivariogram model is fit to the empirical one in order to predict a value at any location. In fitting the variogram model, it is assumed that the signal under study is isotropic. We show in **Figure 7.2** the 2D spatial autocorrelation functions computed directly from a detrended map of SWD derived from PSI-GNSS and the equivalent map from MERIS. Both maps show that the SWD signal is anisotropic. Therefore,

(a) PSI   (b) MERIS
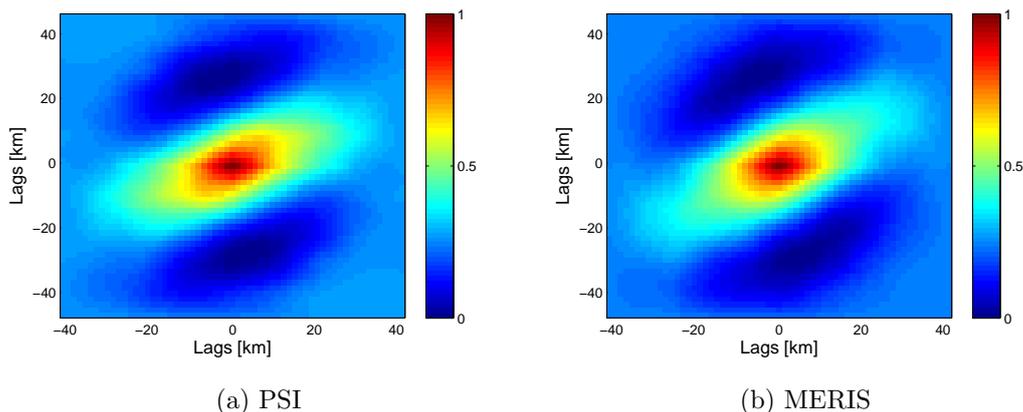
Figure 7.2: Spatial autocorrelation functions for SWD maps, with the long-wavelength component removed, derived from MERIS and PSI data on 09.05.2005.

we make no assumptions of isotropy and we use the FRK covariance model.

The FRK covariance model for single data sets was presented in Section 4.2.2.3, and in this section we present the generalized model for multiple data sets. This model solves the two problems discussed before. This means that it first makes the computations required to obtain the prediction and the MSPE maps scalable and linear in N, and second it addresses the change of support problem.

Following the SME model described in Section 4.2.2.3, the linear model of the data sets in $\mathbf{Z}_1$ can be written as follows

$$\mathbf{Z}_1 = \mathbf{T}_1\boldsymbol{\alpha} + \mathbf{S}_1\boldsymbol{\eta} + \boldsymbol{\zeta}_1 + \boldsymbol{\epsilon}_1 \tag{7.9}$$

and the linear model for $\mathbf{Z}_2$ is

$$\mathbf{Z}_2 = \mathbf{T}_2\boldsymbol{\alpha} + \tilde{\mathbf{S}}_2\boldsymbol{\eta} + \boldsymbol{\epsilon}_2 \tag{7.10}$$

where T is the design matrix that contains the longitude and latitude of the input data and a column of ones and $\boldsymbol{\alpha}$ contains regression coefficients. S_i is a matrix of weights and $\boldsymbol{\eta}$ is a hidden vector of random effects. $\boldsymbol{\zeta}_1$ accounts for the fine-scale variations and it is assumed to follow a Gaussian distribution of a mean zero and a variance σ_ζ^2. Because WRF data are smooth compared to the SWD maps, $\mathbf{Z}_2$ is modeled by a long-wavelength and medium- to short-scale signals $\tilde{\mathbf{S}}_2\boldsymbol{\eta}$, but no fine-scale signal. Note that the data of WRF are block data and the weighting matrix $\tilde{\mathbf{S}}_2$ is constructed by computing the average from the grid points superimposed to the grid of $\mathbf{Z}_2$. The model for both data sets can be rewritten in a more compact form

$$\mathbf{Z} = \mathbf{T}\boldsymbol{\alpha} + \mathbf{S}\boldsymbol{\eta} + \boldsymbol{\zeta} + \boldsymbol{\epsilon} \tag{7.11}$$

where $\mathbf{Z} = [\mathbf{Z}_1', \mathbf{Z}_2']'$, $\mathbf{T} = [\mathbf{T}_1', \mathbf{T}_2']'$, $\mathbf{S} = [\mathbf{S}_1', \tilde{\mathbf{S}}_2']'$, $\boldsymbol{\zeta} = [\boldsymbol{\zeta}_1', \mathbf{0}']'$, and $\boldsymbol{\epsilon} = [\boldsymbol{\epsilon}_1', \boldsymbol{\epsilon}_2']'$. The first term on the right-hand side of eq. (7.11) denotes the long-wavelength deterministic component of the delay. The second component accounts for the medium scale variations, and the third represents the fine-scale signal. $\boldsymbol{\epsilon}$ is a vector that contains the error affecting each system. The hidden state vector $\boldsymbol{\eta}$ has a fixed length r and the elements of this vector are random effects corresponding to a set of predefined locations at several levels of spatial resolutions, so that it can observe variations at several spatial scales. $\boldsymbol{\eta}$ is assumed to have a Gaussian distribution with a mean zero and a covariance matrix $\mathbf{K}$. $\boldsymbol{\zeta}$ and $\boldsymbol{\epsilon}$ are uncorrelated Gaussian distributed signals with variances σ_ζ^2 and σ_ϵ^2, respectively. Both $\boldsymbol{\zeta}$ and $\boldsymbol{\epsilon}$ are assumed independent of each other and of $\boldsymbol{\eta}$.

Recall the covariance model used in Section 4.2.2.3 for a single data set, the covariance model associated with the each data set in eqs. (7.9) and (7.10) has the following structure:

$$\boldsymbol{\Sigma}_{11} = \text{var}(\tilde{\mathbf{Z}}_1) = \mathbf{S}_1 \mathbf{K} \mathbf{S}_1' + \sigma_\zeta^2 \mathbf{V}_\zeta + \sigma_{\epsilon_1}^2 \mathbf{V}_{\epsilon_1} \tag{7.12}$$

$$\boldsymbol{\Sigma}_{22} = \text{var}(\tilde{\mathbf{Z}}_2) = \tilde{\mathbf{S}}_2 \mathbf{K} \tilde{\mathbf{S}}_2' + \sigma_{\epsilon_2}^2 \mathbf{V}_{\epsilon_2} \tag{7.13}$$

$$\boldsymbol{\Sigma}_{12} = \text{cov}(\tilde{\mathbf{Z}}_1, \tilde{\mathbf{Z}}_2) = \mathbf{S}_1 \mathbf{K} \tilde{\mathbf{S}}_2' = \boldsymbol{\Sigma}_{21}' \tag{7.14}$$

Note that in computing the cross covariance functions $\boldsymbol{\Sigma}_{12}$ and $\boldsymbol{\Sigma}_{21}$, the only part of the signals that is assumed correlated is $\boldsymbol{\eta}$. $\mathbf{S}_1(\cdot)$ is the weighting matrix of the size $N_1 \times r$ and $\tilde{\mathbf{S}}_2(\cdot)$ has the size $N_2 \times r$. Each row of the matrix represents an encoding vector for each spatial location where a data point in available. The weights are obtained by evaluating the basis functions given in eq. (7.34) at that location. In order to solve eq. (7.8), we need to specify the covariance matrices of the data and find the covariance between the observations and the true process at the prediction locations using the FRK covariance model. The covariance terms are

obtained from:

$$\mathbf{c}_1 = \mathrm{cov}(\tilde{\mathbf{Z}}_1, \mathbf{Y}(\mathbf{s})) \;\; = \;\; \mathbf{S}_p(\mathbf{s})\,\mathbf{K}\,\mathbf{S}_1' \tag{7.15}$$

$$\mathbf{c}_2 = \mathrm{cov}(\tilde{\mathbf{Z}}_2, \mathbf{Y}(\mathbf{s})) \;\; = \;\; \mathbf{S}_p(\mathbf{s})\,\mathbf{K}\,\tilde{\mathbf{S}}_2' \tag{7.16}$$

$\mathbf{S}_p(\cdot)$ is the weighting matrix for the prediction locations. We should pay high attention to the support of the data, so that the covariance matrix $\mathbf{K}$ is estimated despite having aggregated data. To obtain predictions at block-level, the matrix $\mathbf{S}_p(\mathbf{s})$ is aggregated to the required level using eq. (4.40). **Table 7.1** compares the data at point-level and footprint-level and the formulas to compute the components for each set.

	Point-level data	Areal-level data
True process	$Y(\mathbf{s})$	$Y(B_i) = \frac{1}{|B_i|} \sum\limits_{\mathbf{s} \subset B_i} Y(\mathbf{s})$
Trend	$\mathbf{T}(\mathbf{s})\boldsymbol{\alpha}$	$\left(\frac{1}{|B_i|} \sum\limits_{\mathbf{s} \subset B_i} \mathbf{T}(\mathbf{s}) \right) \boldsymbol{\alpha}$
Weighting matrix ($\mathbf{S}$)	$\mathbf{S}(\mathbf{s})$	$\frac{1}{|B_i|} \sum\limits_{\mathbf{s} \subset B_i} \mathbf{S}(\mathbf{s})$
Medium-scale variations	$\mathbf{S}(\mathbf{s})\boldsymbol{\eta}$	$\left(\frac{1}{|B_i|} \sum\limits_{\mathbf{s} \subset B_i} \mathbf{S}(\mathbf{s}) \right) \boldsymbol{\eta}$
Fine-scale variations	$\zeta(\mathbf{s})$	$\zeta(B_i) = \frac{1}{|B_i|} \sum\limits_{\mathbf{s} \subset B_i} \zeta(\mathbf{s})$
Error	$\epsilon(\mathbf{s})$	$\epsilon(B_i)$

Table 7.1: Point-level and areal-level data

Based on the SSDF covariance functions, the eqs. (7.2) and (7.8) can be rewritten such that the estimator $\hat{Y}(\mathbf{s})$ is

$$\hat{Y}(\mathbf{s}) = \left(\mathbf{S}_p(\mathbf{s})\,\mathbf{K} \begin{bmatrix} \mathbf{S}_1' & \tilde{\mathbf{S}}_2' \end{bmatrix}' + \sigma_\zeta^2 \mathbf{E}(\mathbf{s}) \right) \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}^{-1} \begin{bmatrix} \tilde{\mathbf{Z}}_1 \\ \tilde{\mathbf{Z}}_2 \end{bmatrix} \tag{7.17}$$

The matrix $\mathbf{E}(\mathbf{s})$ has the size $M \times N$, with M the number of prediction locations. It contains a value of one at $\mathbf{s} = \mathbf{s}_j$ and zeros, otherwise. Note that $\hat{Y}(\mathbf{s})$ does not contain the trend signal, which can be computed from $\mathbf{T}\hat{\boldsymbol{\alpha}}$ and added to result from (7.17) to obtain the absolute process. The MSPE corresponding to $\hat{Y}(\mathbf{s})$ can be obtained from

$$\mathrm{MSPE} = \mathbf{a}_1'\boldsymbol{\Sigma}_{11}\mathbf{a}_1 + \mathbf{a}_2'\boldsymbol{\Sigma}_{22}\mathbf{a}_2 + 2\mathbf{a}_1'\boldsymbol{\Sigma}_{12}\mathbf{a}_2 - 2\mathbf{a}_1'\mathbf{c}_1 - 2\mathbf{a}_2'\mathbf{c}_2 \tag{7.18}$$

Figure 7.3 summarizes the procedure followed to obtain the SWD prediction and MSPE maps by assimilating PSI-GNSS and WRF data by means of the SSDF strategy.
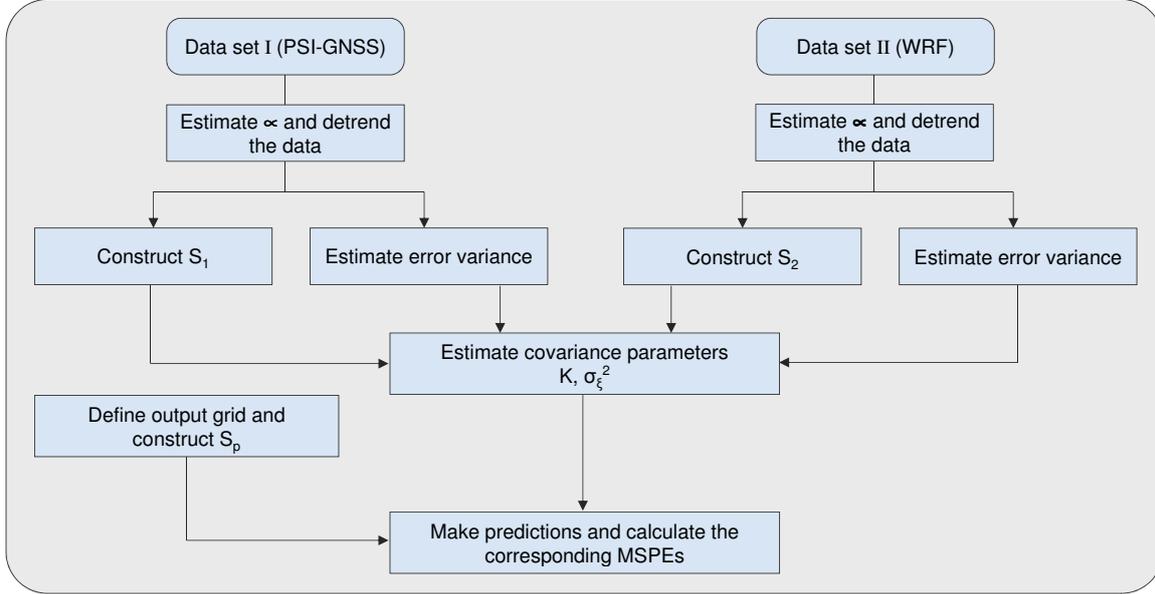
Figure 7.3: The procedure required for data inference based on multiple sets by applying the method of SSDF.

7.3 Matrix inversion

Using the FRK covariance model makes the matrix inversion of eq. (7.8) scalable. The matrix inversion can be achieved by applying a recursive block-wise inversion as follows [Petersen and Pedersen, 2006, pp. 46]:

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{A}^{-1} + \mathbf{A}^{-1}\mathbf{B}(\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{C}\mathbf{A}^{-1} & -\mathbf{A}^{-1}\mathbf{B}(\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1} \\ -(\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{C}\mathbf{A}^{-1} & (\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1} \end{bmatrix} \quad (7.19)$$

where $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}$ are matrices of any size, and $\mathbf{A}, \mathbf{D}$ must be square. For extremely large data sets the system in (7.8) can also be solved in pieces as presented in [Nguyen, 2009], see Appendix B.1. The inversion of individual matrices in (7.19) is achieved by applying the formula of Sherman-Morrison-Woodbury, which is made possible due to the FRK covariance structure,

$$\begin{aligned} \mathbf{\Sigma}_{ii}^{-1} &= (\mathbf{D}_i + \mathbf{S}_i \mathbf{K}_i \mathbf{S}_i')^{-1} \\ &= \mathbf{D}_i^{-1} - \mathbf{D}_i^{-1}\mathbf{S}_i(\mathbf{K}^{-1} + \mathbf{S}_i' \mathbf{D}_i^{-1} \mathbf{S}_i)^{-1}\mathbf{S}_i'\mathbf{D}_i^{-1} \end{aligned} \quad (7.20)$$

The computations require the inversion of the matrices $\mathbf{K}$ and $(\mathbf{K}^{-1} + \mathbf{S}_i' \mathbf{D}_i^{-1} \mathbf{S}_i)$, where each of them has the size $r \times r$. Note that $\mathbf{D}_i$ is a diagonal matrix, for which the inversion is achieved by inverting the diagonal elements. Using the FRK covariance model makes the computational burden for the matrix inversion linear with the data size [Cressie and Johannesson, 2008].

7.4 Estimating K

Since the SSDF method is based on the FRK covariance model, obtaining the predictions requires the covariance matrix $\mathbf{K}$, which should be estimated "offline" from the data. The estimation of $\mathbf{K}$ can be achieved either by the E-M method or the method-of-moments (MM).

7.4.1 Expectation-maximization method

We use the SSDF to estimate a single process (SWD maps); therefore, one covariance matrix $\mathbf{K}$ is estimated from all data sets [Braverman et al., 2009]. In Chapter 4, the matrix $\mathbf{K}$ was estimated from one data set using the E-M algorithm. In this chapter, we extend the E-M method to estimate one common matrix $\mathbf{K}$ based on the two data sets of WRF and PSI-GNSS. Recall eq. (4.45)

$$\mathbf{K}^{[t+1]} = \mathbf{K}^{[t]} + \mathbf{K}^{[t]} \left(\mathbf{S}' \mathbf{\Sigma}^{[t]-1} \left(\tilde{\mathbf{Z}} \tilde{\mathbf{Z}}' \mathbf{\Sigma}^{[t]-1} - \mathbf{I}_N \right) \mathbf{S} \right) \mathbf{K}^{[t]} \tag{7.21}$$

$$\sigma_\zeta^{2[t+1]} = \sigma_\zeta^{2[t]} + \sigma_\zeta^{2[t]} \operatorname{tr} \left(\frac{1}{N} \mathbf{\Sigma}_{11}^{[t]-1} \left(\tilde{\mathbf{Z}}_1 \tilde{\mathbf{Z}}_1' \mathbf{\Sigma}_{11}^{[t]-1} - \mathbf{I}_{N_1} \right) \mathbf{V}_\zeta \right) \sigma_\zeta^{2[t]} \tag{7.22}$$

where $\mathbf{K}$ is estimated from both data sets by setting $\tilde{\mathbf{Z}} = [\tilde{\mathbf{Z}}' \tilde{\mathbf{Z}}']'$, $\mathbf{S} = [\mathbf{S}_1' \mathbf{S}_2']'$, and

$$\mathbf{\Sigma} = \begin{bmatrix} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} \\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22} \end{bmatrix}$$

with $\mathbf{\Sigma}_{ii}$ and $\mathbf{\Sigma}_{ij}$ as given in eqs. (7.12) and (7.14), and the inversion is obtained using eq. (7.19). Note that the fine-scale variance σ_ζ^2 is estimated only from the first data set (SWD observations).

7.4.2 Method-of-moments

Another approach for estimating $\mathbf{K}$ based on the estimation of the empirical covariance $\mathbf{\Sigma}$ was proposed by Cressie and Johannesson [2008], which is called the method-of-moments (MM). The method is described here briefly, and for more details the reader is referred to [Cressie and Johannesson, 2008; Nguyen, 2009]. We describe the method first for a single data set and then

we generalize the approach for multiple sets. In the MM method, the matrix $\mathbf{K}$ is estimated from a single data set by minimizing the distance between the empirical covariance $\hat{\boldsymbol{\Sigma}}$ and the FRK covariance $\boldsymbol{\Sigma}(\mathbf{K}, \sigma_\zeta^2, \sigma_\epsilon^2)$ using the Frobenius norm, i.e.,

$$||\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}(\mathbf{K}, \sigma_\zeta^2, \sigma_\epsilon^2)||_F^2 \tag{7.23}$$

where $\hat{\boldsymbol{\Sigma}}$ is estimated directly from the data, and $\boldsymbol{\Sigma}$ is the theoretical covariance matrix from the FRK model. $||\cdot||_F$ denotes the Frobenius norm, where for a matrix A, $||A||_F = \sqrt{\text{tr}(AA')}$.

In order to estimate the empirical covariance $\hat{\boldsymbol{\Sigma}}$ for the MM method, it is conventional to bin the data to reduce the computational burden. The number of bins has to be greater than r, but less than the data size N. For good estimates of the covariance matrix, the number of bins is selected such that the number of observations within each bin is adequate for a good estimate of the covariance matrix. According to Shi and Cressie [2007], the number of observations should have an average of 40 per bin. The bin centers should offer a good coverage over the spatial domain of the data, and they do not have to be the same for different data sets [Nguyen, 2009]. Let $\{\mathbf{u}_j : j = 1, \ldots, M\}$ be the bin center locations, and $N(\mathbf{u}_j)$ define the neighborhood of the location $\mathbf{u}_j$, then the observations average within the bin is given by:

$$\bar{Z}_j = \frac{\mathbf{w}_j'\tilde{\mathbf{Z}}}{\mathbf{w}_j'\mathbf{1}_N} \tag{7.24}$$

where $\mathbf{w}_j = [w_{j1}, \ldots, w_{jN}]'$ is a vector of ones and zeros such that

$$w_{ji} = \begin{cases} 1 & \text{if } \mathbf{s}_i \in N(\mathbf{u}_j), \ i = 1, \ldots, N \\ 0 & \text{otherwise} \end{cases} \tag{7.25}$$

The binned version of the matrix $\mathbf{S}$ can be computed as follows

$$\bar{\mathbf{S}} = [\mathbf{S}_1', \ldots, \mathbf{S}_M']', \quad \mathbf{S}_j = \frac{\mathbf{w}_j'\mathbf{S}}{(\mathbf{w}_j'\mathbf{1}_N)}$$

The dimension of $\mathbf{S}_j$ is $1 \times r$ and the size of $\bar{\mathbf{S}}$ is $M \times r$. In a similar way the matrices $\mathbf{V}_\zeta$ and $\mathbf{V}_\epsilon$ can be binned such that:

$$\bar{\mathbf{V}}_\zeta = \text{diag}(V_{\zeta,11}, \ldots, V_{\zeta,MM}), \quad V_{jj} = \frac{\mathbf{w}_j'\mathbf{V}_\zeta\mathbf{w}_j'}{(\mathbf{w}_j'\mathbf{1}_N)^2} \tag{7.26}$$

$$\bar{\mathbf{V}}_\epsilon = \text{diag}(V_{\epsilon,11}, \ldots, V_{\epsilon,MM})', \quad V_{\epsilon,11} = \frac{\mathbf{w}_j'\mathbf{V}_\epsilon\mathbf{w}_j'}{(\mathbf{w}_j'\mathbf{1}_N)^2} \tag{7.27}$$

The empirical covariance matrix estimated from the binned data, denoted $\hat{\mathbf{\Sigma}}_B$, is used to estimate $\mathbf{K}$ by minimizing (7.23) [Cressie and Johannesson, 2008]. The FRK covariance $\mathbf{\Sigma}$ in (7.23) equals $\bar{\mathbf{S}}\mathbf{K}\bar{\mathbf{S}}' + \sigma_\zeta^2 \bar{\mathbf{V}}_\zeta + \sigma_\epsilon^2 \bar{\mathbf{V}}_\epsilon$. Then the estimate for $\mathbf{K}$, given that σ_ϵ^2 and σ_ζ^2 are known, is

$$\hat{\mathbf{K}} = \mathbf{R}^{-1}\mathbf{Q}'(\hat{\mathbf{\Sigma}}_B - \sigma_\zeta^2 \bar{V}_\zeta - \sigma_\epsilon^2 \bar{V}_\epsilon)\mathbf{Q}(\mathbf{R}^{-1})' \tag{7.28}$$

where $\hat{\mathbf{\Sigma}}_B$ is the empirical covariance matrix estimated from the binned data. $\bar{\mathbf{S}} = \mathbf{Q}\ \mathbf{R}$, with $\mathbf{Q}$ and $\mathbf{R}$ obtained via the Q-R decomposition of $\bar{\mathbf{S}}$ [Cressie and Johannesson, 2008].

Based on the above formulas for single data sets, we generalize the approach for two data sets. The SSDF covariance functions for binned data are

$$\bar{\mathbf{\Sigma}}_{11} = \bar{\mathbf{S}}_1\ \mathbf{K}\ \bar{\mathbf{S}}_1' + \sigma_\zeta^2 \bar{\mathbf{V}}_\zeta + \sigma_{\epsilon_1}^2 \bar{\mathbf{V}}_{\epsilon_1} \tag{7.29}$$

$$\bar{\mathbf{\Sigma}}_{22} = \bar{\mathbf{S}}_2\ \mathbf{K}\ \bar{\mathbf{S}}_2' + \sigma_{\epsilon_2}^2 \bar{\mathbf{V}}_{\epsilon_2} \tag{7.30}$$

$$\bar{\mathbf{\Sigma}}_{12} = \bar{\mathbf{S}}_1\ \mathbf{K}\ \bar{\mathbf{S}}_2' = \bar{\mathbf{\Sigma}}_{21}' \tag{7.31}$$

Minimizing the norm between the empirical covariance functions and the SSDF covariances in eq. (7.23) yields the following estimate for $\mathbf{K}$

$$\hat{\mathbf{K}} = \mathbf{R}^{-1}\mathbf{Q}' \begin{bmatrix} \hat{\mathbf{\Sigma}}_{11,B} - \sigma_{\zeta_1}^2 \bar{\mathbf{V}}_{\zeta_1} - \sigma_{\epsilon_1}^2 \bar{\mathbf{V}}_{\epsilon_1} & \hat{\mathbf{\Sigma}}_{12,B} \\ \hat{\mathbf{\Sigma}}_{21,B} & \hat{\mathbf{\Sigma}}_{22,B} - \sigma_{\epsilon_2}^2 \bar{\mathbf{V}}_{\epsilon_2} \end{bmatrix} \mathbf{Q}(\mathbf{R}^{-1})' \tag{7.32}$$

where $[\bar{\mathbf{S}}_1'\bar{\mathbf{S}}_2'] = \mathbf{Q}\ \mathbf{R}$. $\hat{\mathbf{\Sigma}}_{11,B}$ and $\hat{\mathbf{\Sigma}}_{22,B}$ is the empirical covariance matrices estimated from the first and the seconds data sets, respectively. $\hat{\mathbf{\Sigma}}_{12,B}$ is the empirical cross covariance matrix. The methods for estimating the empirical covariance matrices $\hat{\mathbf{\Sigma}}_{ii,B}$ and $\hat{\mathbf{\Sigma}}_{ij,B}$ are described in Appendix B.2.

The success of the FRK and SSDF methods depends on the proper estimation of the matrix $\mathbf{K}$ and the variances of the fine-scale component and the noise. For the sake of evaluation and comparison, we estimate $\mathbf{K}$ based on the two presented methods, E-M and MM. The E-M algorithm is based on the assumption that the data follow a joint Gaussian distribution, while the MM does not depend on any statistical assumptions. As a consequence, the MM method should be more robust [Katzfuss and Cressie, 2011b]. If the assumption of Gaussianity is justified, which is most likely for the SWD data, the E-M algorithm shows more efficient results, particularly when the observations are spatially sparse [Katzfuss and Cressie, 2009]. The fine-scale variance, σ_ζ^2 is estimated independently for the MM from the robust semivariogram as described in Section 4.2.2.3; however, it can be estimated together with $\mathbf{K}$ in the E-M algorithm.

7.4.3 Comparison of EM and MM estimations

In this section, we present an example for estimating the matrix $\mathbf{K}$ using the SWD maps derived from PSI and GNSS in the previous chapter. We use the detrended observations to estimate $\mathbf{K}$ and the associated variances σ_ζ^2 and σ_ϵ^2, first by the E-M method and second by the MM. The variances σ_ζ^2 and σ_ϵ^2 are estimated from the point-level observations based on the robust semivariogram method. In the E-M method, we obtain the estimate for $\mathbf{K}$ (denoted $\mathbf{K}_{EM}$) iteratively as described in Section 4.3. Estimating $\mathbf{K}$ based on the MM (denoted $\mathbf{K}_{MM}$) requires binning the observations to block-level; therefore, the bin center locations are created on a regular grid of 1 km. The SWD observations (available at the PS locations) are aggregated within each bin and the bins utilized for estimating $\mathbf{K}$ should contain more than 25 observations while the others are rejected.

It remains to build the matrix $\mathbf{S}$, so we construct a set of spatial nodes within the spatial domain of the data, to which the locations of the measurements should be related. The nodes are established in three groups such that the spacing between them is 40 km, 20 km, and 10 km for the first, the second and the third group, respectively. We used 93 nodes with 4 nodes in the first group, 17 in the second, and 72 in the third. The selection of the nodes was adjusted until the FRK covariance fits the empirical covariance. The nodes setup is depicted in **Figure 7.4**. These nodes define the centers of the BFs used to encode the locations of the available data. This means that each location of the data is represented by a vector of r weights. Each weight is computed by evaluating the BFs at the distance of the observation point $\mathbf{s}$ to the nodes. These vectors are stored in the matrix $\mathbf{S}$, which has the size of $N \times r$ for N observations, such that $\mathbf{S}$ has the following form:

$$
\mathbf{S} = \begin{bmatrix} S^{(1)}(\mathbf{s}_1) & S^{(2)}(\mathbf{s}_1) & \cdots & S^{(r)}(\mathbf{s}_1) \\ S^{(1)}(\mathbf{s}_2) & S^{(2)}(\mathbf{s}_2) & \cdots & S^{(r)}(\mathbf{s}_2) \\ \vdots & \vdots & \ddots & \vdots \\ S^{(1)}(\mathbf{s}_N) & S^{(2)}(\mathbf{s}_N) & \cdots & S^{(r)}(\mathbf{s}_N) \end{bmatrix} \tag{7.33}
$$

where the $S^{(l)}$ are selected from the bisquare BFs family, which has the following form

$$
S^{(l)}(\mathbf{s}) = \begin{cases} \left[1 - (\|\mathbf{s} - \mathbf{m}_l\| / r_l)^2 \right]^2, & \text{for } \|\mathbf{s} - \mathbf{m}_l\| \leq r_l, \\ 0 & \text{otherwise} \end{cases} \tag{7.34}
$$

$\mathbf{m}_l$ is the geographic location of the node and r_l is the effective radius of the BF, which is defined as 1.5 times the distance separating two adjacent nodes in a certain group of BFs [Nguyen, 2009].

The results for the estimates of the covariance matrices $\mathbf{K}_{EM}$ and $\mathbf{K}_{MM}$ are shown in
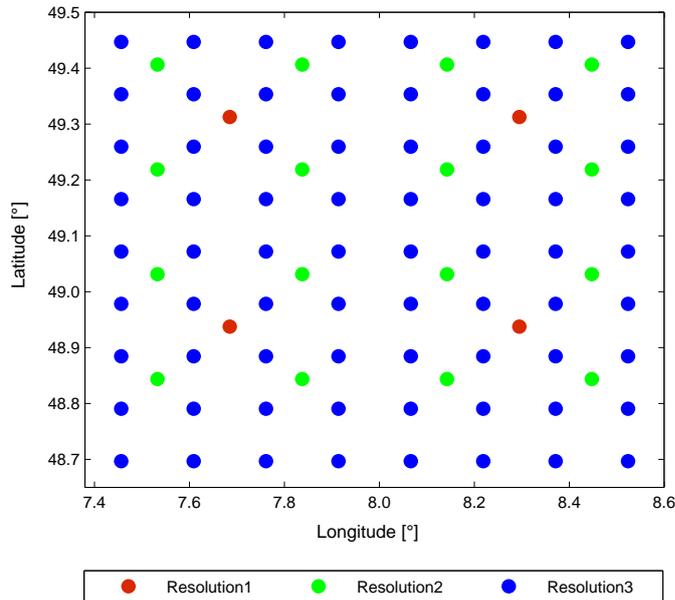
Figure 7.4: FRK nodes which also define the center locations of 93 basis functions at three spatial resolutions. The first resolution is 40 km (4 nodes), the second resolution is 20 km (17 nodes), and the third resolution is 10 km (72 nodes).

Figure 7.5. The estimate of $\mathbf{K}_{EM}$ is slightly smoother than the estimate of $\mathbf{K}_{MM}$, but the difference between the two matrices is small. We found that the difference between the estimates of $\mathbf{K}_{MM}$ and $\mathbf{K}_{EM}$ increases when decreasing the number of bins, i.e., increasing the bin size and hence aggregating more observations. From the matrices $\mathbf{K}_{EM}$ and $\mathbf{K}_{MM}$, a maximum value is observed at the element (29,29), which is equivalent to the node in the lower right corner at the location (8.524°E, 48.69°N), see **Figure 7.4**. This can be explained by the sparseness of observations close to this node, and some of them have inaccurate values. This large variance in $\mathbf{K}$ will affect only a small number of points in the covariance matrix, and its effect will be insignificant if we infer at block-level. In **Figure 7.5** are also shown the corresponding covariance matrices of the detrended observations computed from the estimated parameters using eq. (4.28). The covariance matrix is computed for the observations binned into 7×7 km^2 blocks to demonstrate covariance structure, and the covariance for 1×1 km^2 grid is shown in Appendix B.3 in **Figure B.2**. The covariance matrices computed from both estimates of $\mathbf{K}$ show almost identical values with marginal differences as observed from the figure. We observe from the covariance matrices that the variances, on the main diagonal, increase in areas of sparse observations. The reader should note that the observations do not exist on a regular grid (due to the spatial distribution of PS points); hence, the covariance values in the off-diagonal cells can be negative and then again positive.

The estimates of the covariance matrix $\mathbf{K}$ are exploited to obtain the predictions of the SWD on a regular grid of 3×3 km^2. The prediction maps and the corresponding MSPE maps are illustrated in **Figure 7.6**. The prediction maps show almost the same spatial correlation
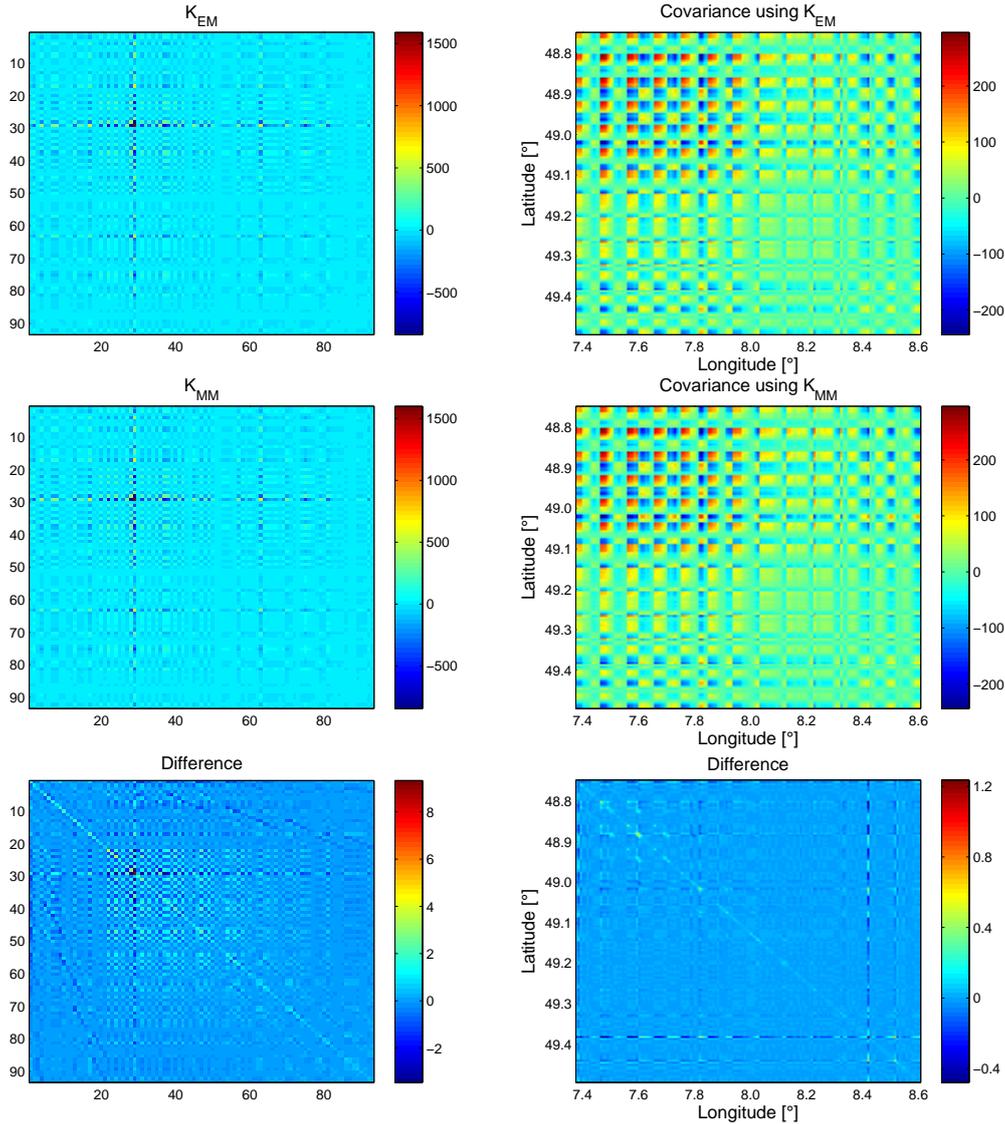
Figure 7.5: Estimates of the covariance matrix **K** using the E-M and the MM algorithms and the corresponding covariance matrices for the SWD maps from PSI-GNSS. The SWD observations are aggregated into maps of 7×7 km^2 cells before their covariance matrices are computed.

with the equivalent map from MERIS. When building the differences between the prediction map and MERIS map, the RMS is about 5 mm in the case of using $\mathbf{K}_{EM}$ and it is slightly larger for the map predicted using $\mathbf{K}_{MM}$. The MSPE values tend to increase in regions of sparse observations. We can also observe that the MSPE values are much higher when using $\mathbf{K}_{MM}$ than those associated with the matrix $\mathbf{K}_{EM}$. This might be explained by the lifting of the eigenvalues for non positive-definite covariance matrices, which is used in the MM method [Kang et al., 2010].

The surface altitude starts to increase when getting close to the lower right corner, which means the SWD values should decrease. This is, however, not observed in either prediction

maps, since the observations in this area are sparse and their values in the original SWD map are not accurate (biased). Using the SSDF method for prediction will overcome this problem by taking the benefit of the data available from WRF, as will be presented in the next section.
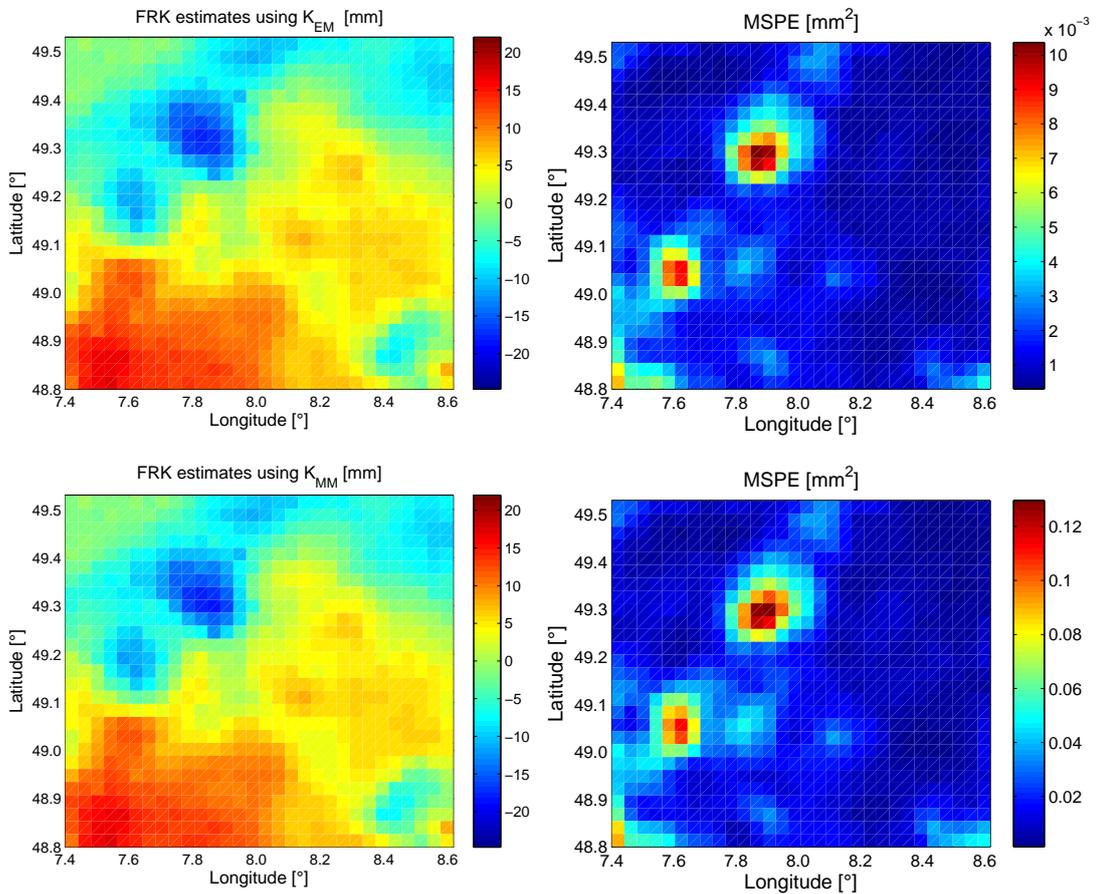
Figure 7.6: SWD prediction maps and the corresponding MSPE values using data from PSI-GNSS combination on 09.05.2005 (09:51 UTC). The grids have a block size of $3{\times}3$ km^2. The upper map is obtained using the covariance matrix $\mathbf{K}_{EM}$, while the lower is predicted based on $\mathbf{K}_{MM}$.

7.5 Fusion of SWD maps

In this section, we apply the SSDF strategy to combine IWV maps, converted into wet delay, from WRF and the absolute SWD maps derived from the combination of PSI and GNSS data as explained in Chapter 6. The SWD maps from PSI-GNSS have point-level observations available at 76841 discrete points. WRF provides data on a regular grid of $3{\times}3$ km^2, which have a block-level support following the structure in eq. (7.1). The WRF map over the region covered by the SAR image contains 1296 data samples. The data to be fused have huge size, different spatial support, gaps in the SWD maps, and noise. WRF simulates data on regular grids, but the quality of the IWV maps and the ability of WRF to model the fine-

to small-scale atmospheric structures should be improved. These reasons motivate the fusion of SWD measurements from PSI-GNSS and WRF simulated maps to study the effect of data fusion on generating accurate prediction maps on regular grids. Also, we test the benefit of the observations in improving the capabilities of WRF to model spatial variations at short scales. We present two examples of applying SSDF to produce regular SWD maps using data on 06.27.2005 and 09.05.2005 (at SAR overpass time), where PSI-GNSS, WRF, and MERIS data are available.

First, the trends are removed from the data by constructing the matrix $\mathbf{T}$ with entries of the longitude and latitude, and estimating the value of $\boldsymbol{\alpha}$ using eq. (4.10). The centered maps are shown in **Figure 7.7**. Then, the covariance parameters $(\mathbf{K}, \sigma_\zeta^2, \sigma_\epsilon^2)$ are estimated using the centered data $\tilde{\mathbf{Z}}_1, \tilde{\mathbf{Z}}_2$ either individually or combined. The error variances of both data sets are estimated as described in Chapter 4, while $\mathbf{K}$ and σ_ζ^2 for multiple data sets are estimated as described in Section 7.4. Note that when the two data sets are combined to infer a single process (SWD fields in this work), one $\mathbf{K}$ is estimated for all data sets.

| (a) WRF: 09.05.2005 | (b) PSI-GNSS: 09.05.2005 |
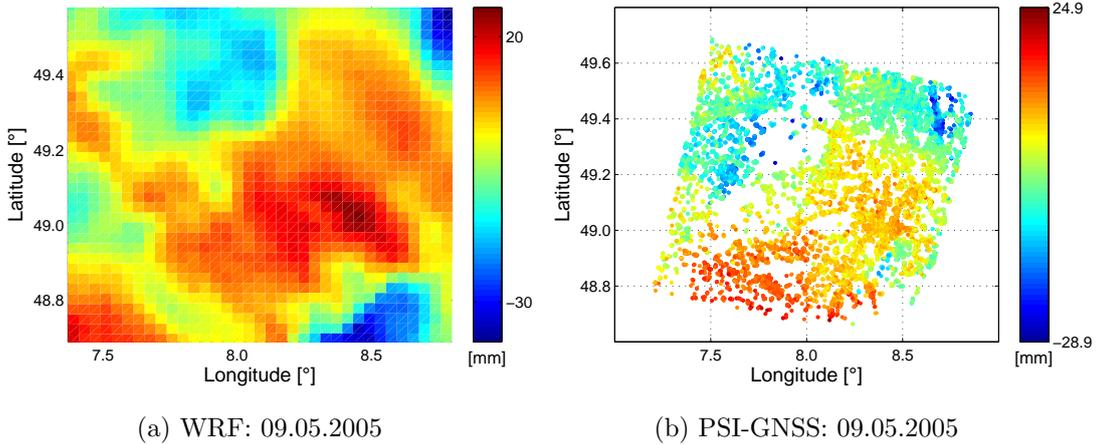
Figure 7.7: SWD maps from PSI-GNSS combination and IWV maps from WRF on 09.05.2005, with a linear trend subtracted from each map. PSI-GNSS are point-level observations, while WRF generates block data with a block size of 3×3 km^2.

We use the FRK method described in Section 4.2.2.3 to interpolate each data set individually on a regular grid and then perform a fusion of them using the SSDF method. Since we fuse two data sets, there exist a matrix $\mathbf{S}_1$ for the first data set (SWD maps) and a matrix $\mathbf{S}_2$ for the second data set (WRF IWV maps). The number of nodes and hence the BFs must be the same for both data sets. The nodes are selected such that $\mathbf{S}$ does not contain columns of zeros, otherwise the corresponding node has to be removed. For the point-level observations from PSI-GNSS, a weighting value is calculated for each data point with respect to all basis functions. However, WRF simulates block-level data, hence we superimpose the grid of WRF with a lattice of regular points such that each cell in the WRF grid contains 9 points. A weighting value is calculated for each point and the values are averaged to get a weighing value for each WRF cell. We form the matrix $\tilde{\mathbf{S}}_2$ at the block-level support of WRF. Building

the matrix $\mathbf{S}_p$ for the prediction locations is done in a similar way, either at point-level or block-level supports, depending on the output grid.

The output grid is defined at 3×3 km^2 (block-level support). In the first example, we show the results of applying the SSDF to the data on 09.05.2005. We compare the prediction maps in three ways: first, we compare the interpolations obtained by applying FRK to single data sets with those obtained by SSDF. Second, we compute the predictions based on $\mathbf{K}$ estimated by the E-M and the MM methods. Third, we validate the results against IWV maps from MERIS interpolated on the same prediction grid by applying the block FRK method.

In **Figure 7.8**, we show the prediction maps obtained by applying FRK to individual data sets and the map obtained by SSDF. The figure also contains the MSPE maps corresponding to each prediction map. The results show that the map obtained from SSDF correlate better with the map predicted from PSI-GNSS observations. In the IWV map generated by WRF, shown in **Figure 7.7** (a), the area in the lower left corner shows artifacts that do not reflect the real map of IWV as observed from its comparison with MERIS, **Figure 7.1** (c, d). Applying FRK to WRF does not remove these artifacts from the prediction map. However, in the map obtained by the fusion of both data sets, the artifacts in the lower corner disappeared, but the corresponding MSPE values are large for this region. The MSPE values corresponding to the SSDF predictions are generally smaller, and we should note that in the regions of sparse observations, the corresponding MSPE values tend to increase. In **Figure 7.9**, we show the SWD profiles over a line drawn horizontally at the latitude 49.37°N. It is observed from the plots that the predictions made by applying the SSDF method are affected more by the data from WRF in the region A, where the there is a lack in the remote sensing observations. However, in region B, the WRF data are overestimated and they have a lower effect in the predictions than those received from the remote sensing data. The results received by applying the data fusion (SSDF) show a better correlation with the data from MERIS.

For regions of sparse observations in the SWD map (**Figure 7.7** (b)), i.e., the areas in the west of the Rhine valley or in the lower right corner, the map from WRF contributes to improve the estimation of the SWD values in the prediction map. The region in the lower right corner has a higher topography and the SWD values are expected to decrease as we observe from the map of WRF. In the prediction map obtained by applying FRK to SWD observations, the predicted values tend to increase since the observations in this area are sparse and partially biased. By applying the SSDF approach, the data available from WRF influence the predictions such that their values are more reasonable and they decrease by moving to the lower right corner. In a similar way, the data from WRF improve the predictions in the region around (7.8°E, 49.25°N), where only sparse SWD observations exist. The data from the model, however, affects the prediction in the lower left corner such that they are smaller than those observed in the MERIS map.

For the sake of comparison, we also estimated the matrix $\mathbf{K}$ once using the MM method ($\mathbf{K}_{MM}$) and second by applying the E-M method ($\mathbf{K}_{EM}$) and data inference by SSDF was done

using either estimates. We found that SSDF prediction map using $\mathbf{K}_{EM}$ shows better spatial correlation with MERIS and a lower RMS value. The more important point here is that the corresponding MSPE values using $\mathbf{K}_{MM}$ are larger than the other MSPE values by more than an order of magnitude. This can be related to the estimate of $\mathbf{K}$ itself, where the MM method requires data binning. According to Shi and Cressie [2007], the number of observations per bin should be large enough (40 or more) to allow good estimates of the covariance between bins. Since WRF data are at block-level and of limited data points, we made the binning with 9 points per bin. This might have influenced the computing the empirical covariance and hence the estimation of $\mathbf{K}$. One more point that can affect the estimation of $\mathbf{K}$ is the lifting of the eigenvalues in case of non positive-definite covariance, which is used in the MM method [Kang et al., 2010].

In order to validate the SSDF output maps, we exploit IWV maps from MERIS and block FRK is applied to predict values on the same grid. In **Table 7.2** are summarized the spatial CCs and the RMS values of the difference maps. The map resulting from the SSDF offers the largest spatial correlation and the smallest mean-squared difference to MERIS maps. From the table, it becomes clear that the WRF map (on 09.05.2005) has a much smaller correlation with the MERIS map and a large RMS value. However, in the synergistic approach, the SSDF map shows better correlation and a lower RMS value compared to the maps from individual data sets.
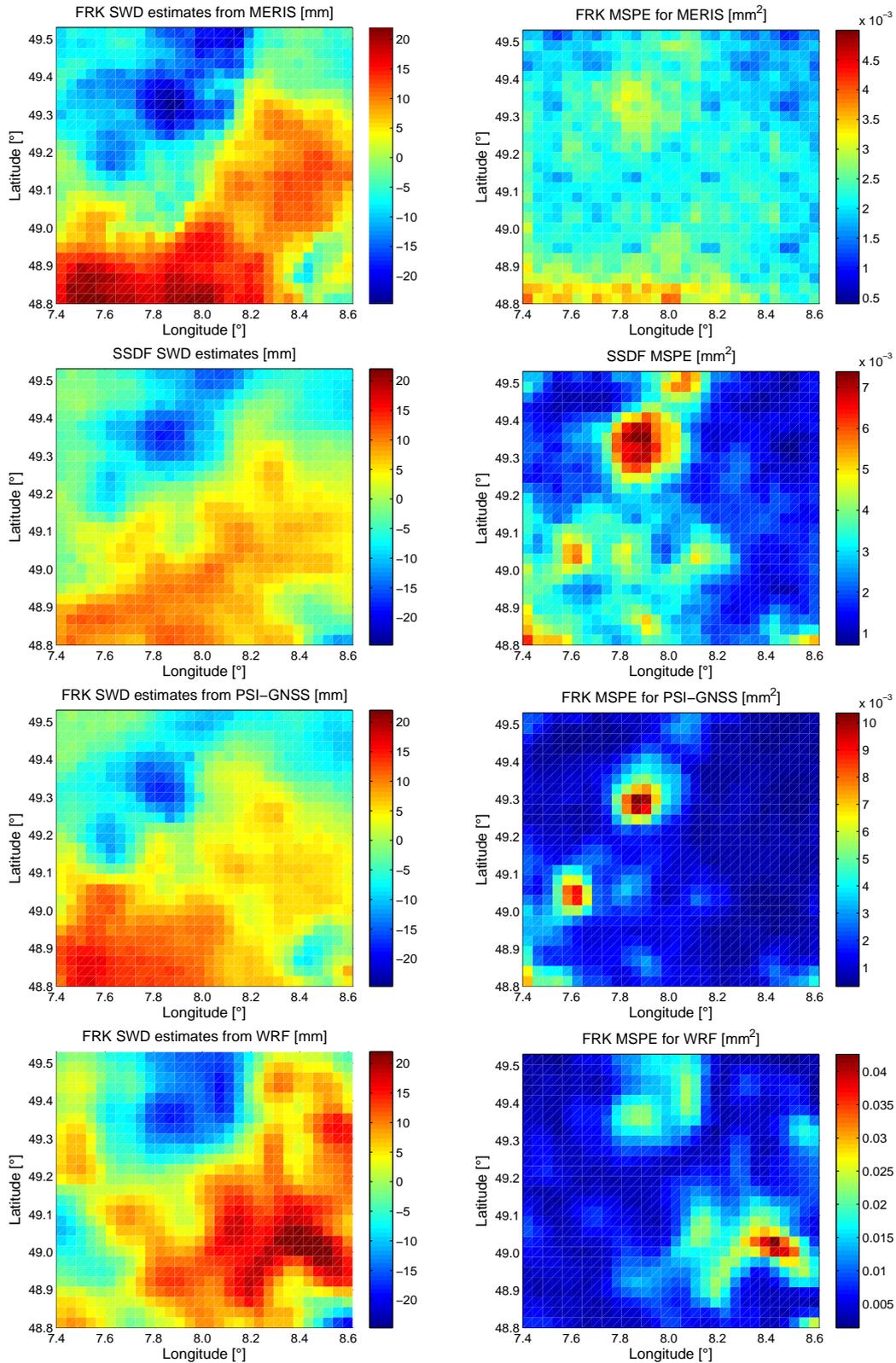
Figure 7.8: SWD prediction maps and MSPE maps obtained by applying SSDF to SWD observations and IWV maps from WRF. Predictions are obtained also by applying FRK is applied to individual maps. The data are available on 09.05.2005 at SAR overpass time. The output grid has a block size of 3×3 km^2.
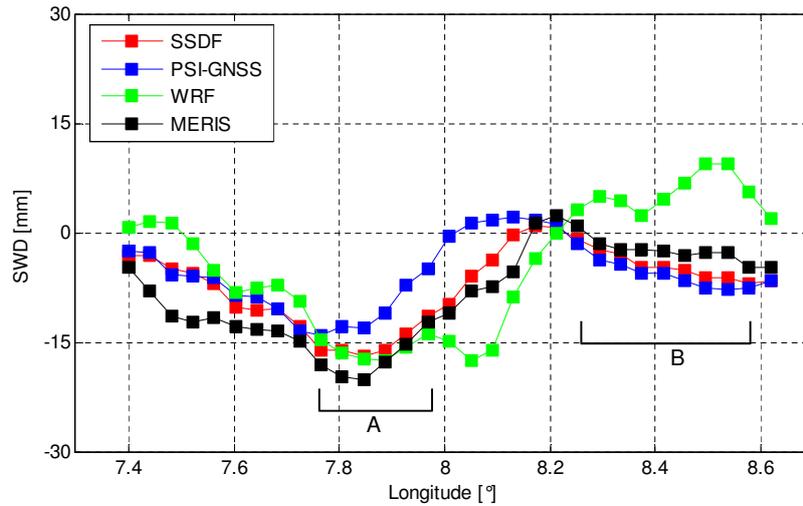
Figure 7.9: SWD profiles following a horizontal line on the latitude 49.37°N in the SWD prediction maps shown in **Figure 7.8**.

	09.05.2005		06.27.2005	
Method	CC	RMS [mm]	CC	RMS [mm]
FRK using PSI-GNSS data	0.8742	5.5427	0.7199	6.8437
FRK using GNSS data	0.7698	8.0081	0.3898	8.8171
FRK using WRF data	0.7022	8.1284	0.8555	5.2743
SSDF ($\mathbf{K}_{MM}$)	0.9042	5.1334	0.7980	6.2011
SSDF ($\mathbf{K}_{EM}$)	0.9111	5.0466	0.8585	5.6036

Table 7.2: Spatial correlations coefficients and the RMS values between SWD maps from MERIS and SWD maps obtained either by SSDF or by FRK of single data sets.

We applied the SSDF approach again to show other maps (on 06.27.2005). The centered data are shown in **Figure 7.10**, and **Figure 7.11** depicts the prediction maps using the SSDF and FRK methods. In this example, we have the opposite case where the WRF IWV map exhibit better spatial correlation with the map from MERIS than that of PSI-GNSS, see **Table 7.2**. Similar to the previous example, the SSDF map demonstrates better results than predictions based on single data sets and the corresponding MSPE values are generally smaller. However, the improvement is not significant and the RMS value of the difference between MERIS and SSDF maps is larger than that for WRF and MERIS. This is related to the support of the input data, where more data samples from the point-level observations in the PSI-GNSS map can have a significant effect on the output value within the predicted cell. The predictions made based on the matrix $\mathbf{K}_{EM}$ are better than those achieved using the matrix $\mathbf{K}_{MM}$. The CC and the RMS values are given in **Table 7.2**.

(a) WRF: 06.27.2005
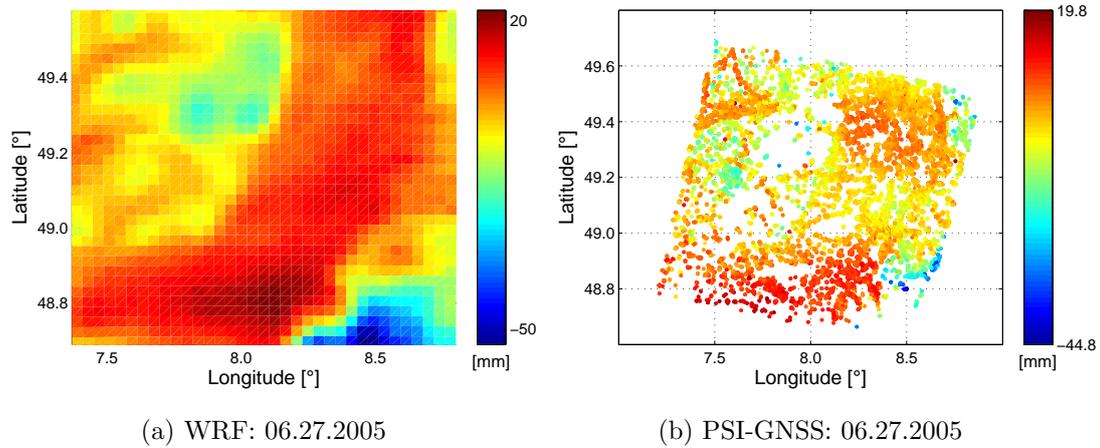
(b) PSI-GNSS: 06.27.2005

Figure 7.10: IWV maps from WRF and SWD maps from PSI-GNSS combination, with a linear trend subtracted from each map. The data are received on 06.27.2005 (SAR overpass). PSI-GNSS are point-level observations, while WRF IWV maps are block data with block size of 3×3 km^2.
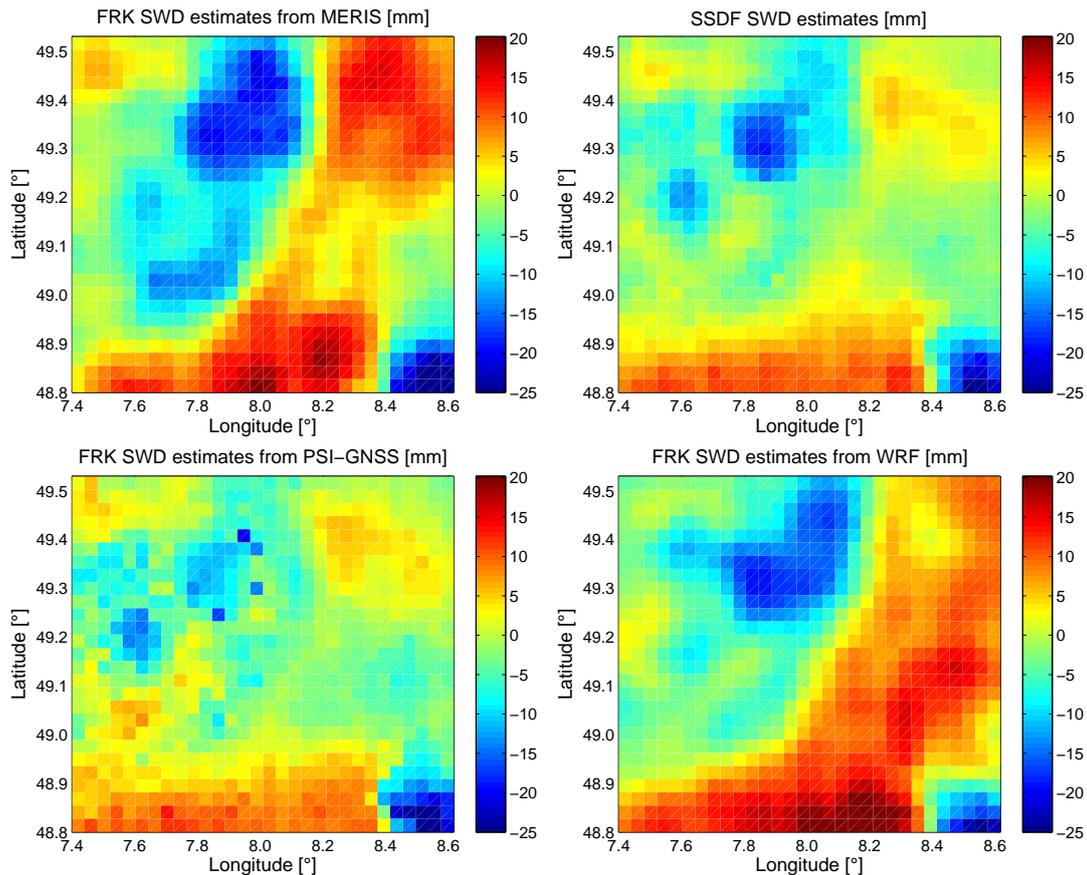
Figure 7.11: SWD prediction maps obtained by applying SSDF to SWD observations and IWV maps from WRF. Predictions are obtained also by applying FRK is applied to individual maps. The data are avilable on 06.27.2005 at SAR overpass time. The output grid has a block size of 3×3 km^2.

113

7.6 Summary and discussion

Maps of absolute SWD were derived in Chapter 6 by combining observations from PSI and GNSS. These maps are available at discrete points and the observations are absent in regions of low coherence. In some areas where the observations are not dense, the value of SWD might be biased. On the other hand, the WRF model provides simulations of the IWV in the atmosphere on regular grids at a coarse spatial resolution, in which modeling the small-scale structures in the atmosphere should be improved. Also, the quality of the maps is still to be improved. In this chapter, we presented a method for spatial fusion of the SWD maps and the IWV maps to infer maps on regular grids of any resolution. For that purpose, the SSDF method, first presented in [Nguyen, 2009], was exploited. This method is based on the FRK approach and it attempts to solve the problems of huge data sets, change of support, and errors.

- We inferred SWD data on a grid of 3×3 km^2 based on estimates from E-M and MM methods and also compared the results with maps from MERIS inferred on the same grid. There exist a strong correlation between SSDF maps and those maps from MERIS. The difference between both maps has RMS values below 6.2 mm, which is lower than that obtained from inferring data based on single sets. The CC and the RMS values for the examples we presented are given in **Table 7.2**.

- In the SSDF, we construct a matrix $\mathbf{S}_i$ for each data source by defining a set of spatial nodes. The number of the nodes is adjusted such that we can estimate a matrix $\mathbf{K}$ that is used to compute a covariance function for the data sets, which should approximate the empirical covariance. In the future work, the size and the locations of nodes should be optimized by minimizing the difference between the true covariance and the estimated one.

- The estimation of the covariance matrix $\mathbf{K}$ was achieved by the methods of E-M and MM. A comparison of the matrices shows a good correspondence of the estimates from both methods, the difference may increase for the off-diagonal elements. We found that increasing the size of bins, i.e., reducing their number, makes the difference between the estimates severe and the relation is not anymore linear.

- The SSDF approach can be extended such that more than two data sets are used. We can test the algorithm by including MERIS maps in the fusion. Hence, maps of IWV based on spectral imaging systems, such as EnMAP (Environmental Mapping and Analysis Program), can be investigated for improving the quality of the NWP models.

- For modeling the fine-scale signal, we assumed that it is spatially independent and identically distributed with a Gaussian distribution of mean zero and a variance σ_ζ^2, such that the covariance matrix has the form $\sigma_\zeta^2 \cdot \mathbf{I}$, with $\mathbf{I}$ the identity matrix. If we assume that the fine-scale signal is correlated within a certain radius, say approximately 2-3 km

for atmospheric turbulence patterns, the model of this signal should be adapted, and at the same time we should keep the computational costs feasible. If we follow the same approach of setting a number of nodes, we should ensure that the matrix **S** does not contain columns of zeros. This might mean that the nodes will be close together in the areas of dense observations and sparse elsewhere.

8 Conclusions and Outlook

Atmospheric water vapor is responsible for the high temporally and spatially variable distortions observed in GNSS and InSAR measurements. Numerous methods have been developed to mitigate these artifacts in order to improve the quality of the methods developed for high precision positioning based on the GNSS and Earth's surface monitoring using InSAR observations. In this work, we investigated this "error" signal to obtain time series and maps of the slant wet delay (SWD), as a measure for IWV, from GNSS and Persistent Scatterer InSAR (PSI) observations, respectively. Using the method of Precise Point Positioning (PPP), we derived absolute measurements of the wet delay from GNSS at a coarse spatial density. PSI, however, was applied to derive maps of the wet delay difference maps at a high spatial density. We combined the wet delay measurements, of complementary properties, from GNSS and PSI to build absolute maps of the wet delay at the density of PSI. Then, we applied data fusion of the derived maps and the IWV maps simulated by the WRF model to overcome the deficits of each set and to improve the quality of the output maps.

8.1 Summary

In this section, we briefly summarize the methods presented in the previous chapters. The algorithm developed for combining PSI and GNSS observations addressed the following points:

- GNSS

  – We used meteorological observations of air pressure, temperature, and relative humidity to determine the zenith dry delay at the GNSS site based on the model of Saastamoinen. The zenith dry delay was subtracted from the neutrospheric zenith total delay, estimated from the phase measurements using the PPP strategy, to obtain a value for the zenith wet delay at each site.

  – The zenith wet delay observations were mapped to the radar line-of-sight to get values of the SWD, which are used to model the elevation-dependent wet delay by fitting the model in Eq. (6.5). The parameter estimates were exploited to extrapolate a value for the wet delay at each PS location depending on its altitude that is obtained from a laser digital elevation model model.

116

– We analyzed the residuals (after subtracting the elevation-dependent component from the GNSS SWD observations) to estimate the parameters of a plane that models the long-wavelength wet delay component, which are used to extrapolate a value at each persistent scatterer location.

We compared the zenith wet delay estimated at each GNSS site with the ZWD observations from MERIS. The results show strong correlation between them and the RMS values are significantly small.

- PSI

    – We built 16 interferograms from the SAR images acquired by the Envisat satellite. We subtracted the topographic phase and the flat Earth phase from each interferogram. Then, we used the StaMPS software for PSI analysis.

    – A series of spatial and temporal filtering was applied to extract the SWD maps from the interferograms, which are then inverted by the method of least squares inversion to derive maps at each SAR acquisition time.

The SWD-difference maps extracted from the interferograms show strong correlation with the the SWD-difference maps from MERIS. We interpolated PSI observations at the locations of MERIS by the method of inverse distance weighting, moving-window averaging and kriging. All of the methods produce similar output maps with insignificant variations in the correlation coefficient and the RMS values when compared with MERIS.

- Combination

    – The maps derived using the GNSS and PSI data are combined to build absolute maps of the SWD.

The comparison of the maps derived using our method with IWV maps from MERIS, converted into wet delay and mapped to the radar line-of-sight, demonstrates strong spatial correlations between 75% and 92% at 300×300 m^2 spatial resolution. The difference maps show RMS values of less than 1 mm (IWV).

- SSDF

We combined the derived absolute SWD maps together with the IWV maps generated by the WRF model by a rigorous data fusion method. Data fusion is applied to test the effect of combining multiple data sources on improving the quality of the output maps, and to use data available from the model for predicting values of the SWD in areas where no measurements are available. For data fusion, we exploited the geostatistical method of SSDF that extends the method of fixed-rank-kriging (FRK) for combining multiple data sets. Based on the FRK covariance model, we addressed the problems of massive data sets, gaps, change of support, and non-stationary and non-isotropy assumptions. In applying the SSDF method, the following points were processed:

– We subdivided the SWD into three components: the first is a long-wavelength signal modeled by a 2D linear surface, the second is a component of small- to medium-scale variations modeled by the spatial random effect model, and the third is a fine-scale component modeled as an uncorrelated random signal with a Gaussian distribution of a mean zero and a diagonal covariance matrix.

– We estimated the error variance for each data set using the robust semivariogram method. The empirical covariance matrix for each data set and the cross covariance of both sets were estimated directly from the data, and they were modeled by the FRK covariance structure.

– We defined the output grid at a block-level support, which is considered in the covariance model, and obtained the predictions of the SWD and the corresponding mean squared prediction error values.

The results show that applying the SSDF approach produce prediction maps with a better spatial correlation and a smaller distance to those from MERIS compared with the maps predicted by applying the FRK method to each data set individually. The artifacts that might exist in any of the data sets are compensated in the output map of the fusion. In the maps predicted from the PSI-GNSS and WRF data on 09.05.2005, for example, the prediction map obtained from applying the FRK method to the WRF map correlates with the map from MERIS to 70%, and the SWD prediction map correlates with MERIS to 87%; however, the prediction map obtained by applying the SSDF approach shows correlations of 91%.

8.2 Recommendations and future work

There might be possible improvements to consider:

In the estimation of the neutrospheric phase from PSI, (i) we assumed that the contribution of the surface displacement is negligible due to the stable crust in the region of URG. However, we should consider the case where the interferogram also contains a displacement signal and the possible ways to split its phase from the neutrospheric phase. (ii) The separation of the neutrospheric signal from other phase components is attained by a series of temporal and spatial filtering. Tuning the filter parameters, in particular for the temporal filters, can result in removing a useful signal or adding a a non-neutrospheric signal. For more accurate results, the extraction of the neutrospheric phase from the interferogram should be achieved in a more efficient way. (iii) We used the StaMPS functions for PS processing where only PS are selected, which results in empty regions in the forest and vegetated areas. To increase the density of stable points, we suggest the use of the SqueeSAR approach, which considers not only PS but also distributed scatterers. This can effectively increase the density of stable points and reduces the probability to have regions with no observations.

In modeling the elevation-dependent wet delay from the GNSS SWD observations, (iv) we assumed that the SWS-residual maps derived from the interferograms contain no altitude-dependent signals since the region of research has a rather flat topography. If we consider regions of strong topography, we should expect that the elevation-dependent signal will be partially observed in the maps derived from PSI. Hence, only the missing component should be modeled based on the GNSS observations. (v) Modeling the elevation-dependent component can be improved by increasing the density of the GNSS sites such that they properly cover the entire topography range.

In the topic of data fusion, we have solved the change of support problem by using the extended FRK covariance model. The observations are related to a fixed set of nodes, which are adjusted manually. (vi) The number of nodes and the spacing between them should be optimized such that the norm of the difference between the empirical covariance and the FRK covariance is minimized. (vii) For modeling the fine-scale signal, we assumed that it is spatially independent and identically distributed with a Gaussian distribution of a mean zero and a covariance matrix $\sigma_\zeta^2 \cdot \mathbf{I}$. If we assume that the fine-scale signal is correlated within a certain radius, say approximately 2-3 km for atmospheric turbulence patterns, the model of this signal should be adapted and at the same time we should pay attention to the computational costs. Following the same approach of setting a number of nodes, we should confirm that the matrix $\mathbf{S}$ does not contain columns of zeros. This might mean that the nodes will be close together in the areas of dense observations and sparse elsewhere.

A General formulas

Let $\mathbf{z}$ be an observed random variable that has a signal component $\boldsymbol{\eta}$ and a noise component. Then, the expectation of the signal, say $\boldsymbol{\eta}$, conditional upon the observation $\mathbf{z}$ is [Klebaner, 2005]:

$$\mathrm{E}\{\boldsymbol{\eta}|\mathbf{z}\} = \mathrm{E}\{\boldsymbol{\eta}\} + \frac{\mathrm{cov}(\boldsymbol{\eta}, \mathbf{z})}{\mathrm{var}(\mathbf{z})}\left(\mathbf{z} - \mathrm{E}\{\mathbf{z}\}\right) \tag{A.1}$$

$$\mathrm{E}\{\boldsymbol{\eta}\boldsymbol{\eta}'|\mathbf{z}\} = \mathrm{cov}(\boldsymbol{\eta}|\mathbf{z}) + \left(\mathrm{E}\{\boldsymbol{\eta}|\mathbf{z}\}\right)^2 \tag{A.2}$$

$$\mathrm{cov}(\boldsymbol{\eta}|\mathbf{z}) = \mathrm{var}(\boldsymbol{\eta}) - \frac{\mathrm{cov}(\boldsymbol{\eta}, \mathbf{z})}{\mathrm{var}(\mathbf{z})} \tag{A.3}$$

B Data fusion

B.1 SSDF for massive data sets

For large data sets, we can find the solution for Eq. (7.7) by partition:

$$
\begin{aligned}
\mathbf{a}_1 &= -\Sigma_{11}^{-1}\Sigma_{12}\mathbf{a}_2 + \Sigma_{11}^{-1}\boldsymbol{c}_1 - \Sigma_{11}^{-1}\mathbf{1}_{N_1}m \\
\mathbf{a}_2 &= -\Sigma_{22}^{-1}\Sigma_{21}\mathbf{a}_1 + \Sigma_{22}^{-1}\boldsymbol{c}_2 - \Sigma_{22}^{-1}\mathbf{1}_{N_2}m
\end{aligned}
\tag{B.1}
$$

substituting the second equation into the first and vice versa, then define the following:

$$
\begin{aligned}
A_1 &= I_{N_1} - \Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} \\
A_2 &= I_{N_2} - \Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} \\
B_1 &= \Sigma_{11}^{-1}(\boldsymbol{c}_1 - \Sigma_{12}\Sigma_{22}^{-1}\boldsymbol{c}_2) \\
B_2 &= \Sigma_{22}^{-1}(\boldsymbol{c}_2 - \Sigma_{21}\Sigma_{11}^{-1}\boldsymbol{c}_1) \\
C_1 &= -\Sigma_{11}^{-1}(\mathbf{1}_{N_1} - \Sigma_{12}\Sigma_{22}^{-1}\mathbf{1}_{N_2}) \\
C_2 &= -\Sigma_{22}^{-1}(\mathbf{1}_{N_2} - \Sigma_{21}\Sigma_{11}^{-1}\mathbf{1}_{N_1})
\end{aligned}
$$

where I_{N_i} is an identity matrix of $N_i \times N_i$. Substituting in Eq. (B.1) then,

$$
\begin{aligned}
\mathbf{a}_1 &= A_1^{-1}(B_1 + C_1 m) \\
\mathbf{a}_2 &= A_2^{-1}(B_2 + C_2 m)
\end{aligned}
\tag{B.2}
$$

Thus, we solve for m by substituting (B.2) into the constraint

$$
\mathbf{a}_1'\mathbf{1}_{N_1} + \mathbf{a}_2'\mathbf{1}_{N_2} = 1
$$

Then,

$$
m = \frac{1 - \mathbf{1}_{N_1}'A_1^{-1}B_1 + \mathbf{1}_{N_2}'A_2^{-1}B_2}{\mathbf{1}_{N_1}'A_1^{-1}C_1 + \mathbf{1}_{N_2}'A_2^{-1}C_2}
\tag{B.3}
$$

The value of m can be substituted back in Eq. (B.2) to obtain the optimal fusion coefficients.

B.2 Estimation of the empirical covariance

In Section 7.4.2, we presented the method for estimating the covariance matrix $\mathbf{K}$ using the MM. We assumed that the empirical covariance matrices $\hat{\boldsymbol{\Sigma}}_{ii,B}$ and $\hat{\boldsymbol{\Sigma}}_{ij,B}$ are known. We now describe the methods for estimating these matrices directly from the data as presented in [Cressie and Johannesson, 2008].

For a data sets $\tilde{\mathbf{Z}}$ of the dimension N, the covariance, denoted C, of any two bins centered at the locations $\mathbf{u}_j$ and $\mathbf{u}_k$ is given by:

$$C(\mathbf{u}_j, \mathbf{u}_k) = \bar{Z}(\mathbf{u}_j) \cdot \bar{Z}(\mathbf{u}_k), \quad j, k = 1, \ldots, M \tag{B.4}$$

where M is the number of bins and

$$\bar{Z}(\mathbf{u}_j) = \frac{\mathbf{w}_j' \tilde{\mathbf{Z}}}{\mathbf{w}_j' \mathbf{1}_N} \tag{B.5}$$

where $\mathbf{w}_j = [w_{j1}, \ldots, w_{jN}]'$ is a vector of ones and zeros such that

$$w_{ji} = \begin{cases} 1 & \text{if } \mathbf{s}_i \in N(\mathbf{u}_j), \quad i = 1, \ldots, N \\ 0 & \text{otherwise} \end{cases} \tag{B.6}$$

The variance, V, for any bin centered at $\mathbf{u}_j$ is obtained from

$$V(\mathbf{u}_j) = \frac{\sum_{i=1}^{N} w_{ji}(Z(\mathbf{s}_i) - \bar{Z}(\mathbf{u}_j))^2}{\mathbf{w}_j' \mathbf{1}_N}, \quad j, k = 1, \ldots, M \tag{B.7}$$

Hence, the empirical covariance $\hat{\boldsymbol{\Sigma}}_B$ is determined from

$$\hat{\boldsymbol{\Sigma}}_B = \mathbf{C} + \text{diag} \{V(\mathbf{u}_1), \ldots, V(\mathbf{u}_M)\} \tag{B.8}$$

Where the matrix $\mathbf{C}$ contains the covariances calculated for all bins from (B.4).

The method presented before is exploited to estimate the covariance matrix for each data set. The cross-covariance between any two bins within different data sets is computed from:

$$C(\mathbf{u}_{1j}, \mathbf{u}_{2k}) = \bar{Z}_1(\mathbf{u}_j) \cdot \bar{Z}_2(\mathbf{u}_k) \tag{B.9}$$

where $\mathbf{u}_{1j}$ is a bin center in the first data set and $\mathbf{u}_{2k}$ is a bin center in the second set.

$$\bar{Z}_1(\mathbf{u}_j) = \frac{\mathbf{w}'_{1j}\tilde{\mathbf{Z}}_1}{\mathbf{w}'_{1j}\mathbf{1}_{N_1}}, \quad j = 1, \ldots, M_1 \tag{B.10}$$

$$\bar{Z}_2(\mathbf{u}_k) = \frac{\mathbf{w}'_{2k}\tilde{\mathbf{Z}}_2}{\mathbf{w}'_{2k}\mathbf{1}_{N_2}}, \quad j = 1, \ldots, M_2 \tag{B.11}$$

where N_1 is the length of the first data set and N_2 is the length of the second. M_1 and M_2 are the number of bins of the first and the second data set, respectively. $\mathbf{w}_{1j}$ and $\mathbf{w}_{2k}$ are defined using (B.6) for the first and the second data set, respectively. Then the cross-covariance matrix is constructed for all bins, i.e.,

$$\hat{\boldsymbol{\Sigma}}_{12,B} \equiv [C(\mathbf{u}_{1j}, \mathbf{u}_{2k})] \tag{B.12}$$

B.3 Additional graphs
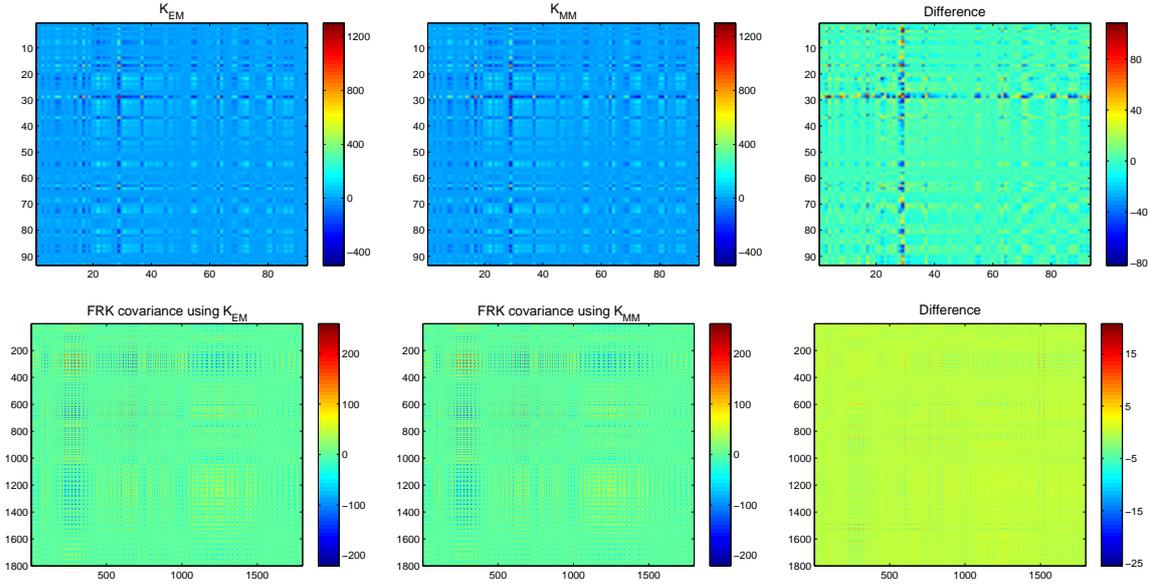
Figure B.1: Estimates of the covariance matrix **K** using the EM and the MM algorithms and the corresponding FRK covariance matrices for the SWD maps from PSI-GNSS on 06.27.2005. Note that the SWD observations are aggregated into maps of 1 km bins before their covariance matrices are computed.
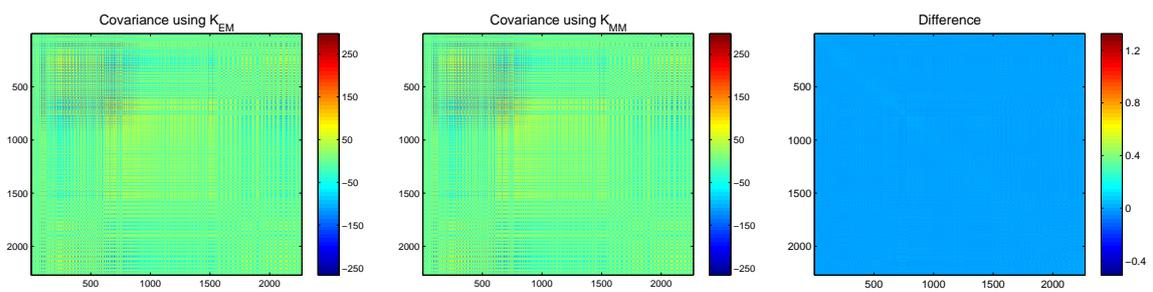
Figure B.2: FRK covariance matrices for the SWD maps from PSI-GNSS on 09.05.2005. Note that the SWD observations are aggregated into maps of 1 km bins before their covariance matrices are computed.

References

Adam, N., Kampes, B., and Eineder, M. (2003). The development of a scientific permanent scatterer system. In *Proceedings of ISPRS workshop. Inst. for Photogramm. and Geoinf., Hannover, Germany*.

Adam, N., Kampes, B., and Eineder, M. (2005). Development of a scientific permanent scatterer system: Modifications for mixed ERS/ENVISAT time series. In *Proceedings of the 2004 Envisat/ERS Symposium*. European Space Agency, ESA SP-572, 457465.

Agram, P. (2010). *Persistent Scatterer Interferometry in Natural Terrain*. PhD thesis, Stanford University, Stanford, California, USA.

Askne, J. and Nordius, H. (1987). Estimation of tropospheric delay for microwaves from surface weather data. *Radio Science*, 22(3):379–386.

Bai, Z. and Feng, Y. (2003). GPS water vapor estimation using interpolated surface meteorological data from Australian automatic weather stations. *Journal of Global Positioning Systems*, 2(2):83–89.

Bamler, R. and Hartl, P. (1998). Synthetic aperture radar interferometry. *Inverse problems*, 14(4):R1.

Barry, R. G. and Chorley, R. J. (1987). *Atmosphere, weather and climate*. Methuen, New York, NY, 5th edition.

Bender, M., Dick, G., Wickert, J., Schmidt, T., Song, S., Gendt, G., Ge, M., and Rothacher, M. (2008). Validation of GPS slant delays using water vapour radiometers and weather models. *Meteorologische Zeitschrift*, 17(6):807–812.

Bertsekas, D. P. (1996). *Constrained Optimization and Lagrange Multiplier Methods*. Academic Press, Inc.

Bevis, M., Businger, S., Chiswell, S., Herring, T. A., Anthes, R. A., Rocken, C., and Ware, R. H. (1994). GPS Meteorology: Mapping Zenith Wet Delays onto Precipitable Water. *Journal of Applied Meteorology*, 33(3):379–386.

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.

Boehm, J. and Schuh, H. (2004). Vienna mapping functions in VLBI analyses. *Geophysical Research Letters*, 31(1).

Braverman, A., Nguyen, H., and Cressie, N. (2011). Space-time data fusion for remote sensing applications. In *34th International Symposium on Remote Sensing of Environment, Sydney, Australia, 10-15 April 2011*. Pasadena, CA: Jet Propulsion Laboratory, National Aeronautics and Space Administration, 2011.

Braverman, A. J., Cressie, N., Katzfuss, M., Michalak, A. M., Miller, C. E., Nguyen, H., Olsen, E. T., and Wang, R. (2009). Geostatistical Data Fusion for Remote Sensing Applications. *AGU Fall Meeting Abstracts*, page C1014.

Byers, H. R. (1974). *General Meteorology*. New York (McGraw-Hill).

Cressie, N. and Hawkins, D. M. (1980). Robust estimation of the variogram: I. *Journal of the International Association for Mathematical Geology*, 12(2):115–125.

Cressie, N. and Johannesson, G. (2008). Fixed rank kriging for very large spatial data sets. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(1):209–226.

Dach, R., Hugentobler, U., Fridez, P., and Meindl, M. (2007). *Bernese GPS Software Version 5.0*. Astronomical Institute, University of Bern, Switzerland.

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. (1986). Atmospheric propagation effects on radio interferometry. *NASA STI/Recon Technical Report N*, 87:23855.

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(6):1003–1018.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, et al. (2011). The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 137(656):553–597.

Delacourt, C., Briole, P., and Achache, J. A. (1998). Tropospheric corrections of SAR interferograms with strong topography. Application to Etna. *Geophysical Research Letters*, 25(15):2849–2852.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the royal statistical society, series B*, 39(1):1–38.

Deutsch, C. and Journel, A. (1998). *GSLIB Geostatistical software library and user's guide*. Oxford University Press.

Doin, M.-P., Lasserre, C., Peltzer, G., Cavalié, O., and Doubre, C. (2009). Corrections of stratified tropospheric delays in SAR interferometry: Validation with global atmospheric models. *Journal of Applied Geophysics*, 69(1):35 – 50.

Drobne, S. and Lisec, A. (2009). Multi-attribute decision analysis in GIS: weighted linear combination and ordered weighted averaging. *Informatica (03505596)*, 33(4).

Eineder, M., Fritz, T., Abdel Jaber, W., Rossi, C., and Breit, H. (2012). Decadal Earth topography dynamics measured with TanDEM-X and SRTM. In *Geoscience and Remote Sensing Symposium (IGARSS), 2012 IEEE International*, pages 1916–1919. IEEE.

Elgered, G., Davis, J. L., Herring, T. A., and Shapiro, I. I. (1991). Geodesy by radio interferometry: Water vapor radiometry for estimation of the wet delay. *Journal of Geophysical Research: Solid Earth*, 96(B4):6541–6555.

126

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.

Emardson, T., Simons, M., and Webb, F. (2003). Neutral atmospheric delay in interferometric synthetic aperture radar applications: Statistical description and mitigation. *Journal of Geophysical Research: Solid Earth (1978–2012)*, 108(B5).

Emardson, T. R. and Derks, H. J. P. (2000). On the relation between the wet delay and the integrated precipitable water vapor in the European atmosphere. *Meteorological Applications*, 7:61–68.

ESA (2006). *ESA, MERIS Product Handbook*.

Ferretti, A., Fumagalli, A., Novali, F., Prati, C., Rocca, F., and Rucci, A. (2011). A New Algorithm for Processing Interferometric Data-Stacks: SqueeSAR. *Geoscience and Remote Sensing, IEEE Transactions on*, 49(9):3460–3470.

Ferretti, A., Prati, C., and Rocca, F. (2001). Permanent scatterers in SAR interferometry. *Geoscience and Remote Sensing, IEEE Transactions on*, 39(1):8–20.

Fischer, J. and Bennartz, R. (1997). *Retrieval of Total Water Vapour Content from MERIS Measurements: Algorithm Theoretical Basis Document ; ATBD 2.4*. Freie Universität Berlin, Berlin, Germany.

Foster, J., Brooks, B., Cherubini, T., Shacat, C., Businger, S., and Werner, C. L. (2006). Mitigating atmospheric noise for InSAR using a high resolution weather model. *Geophysical Research Letters*, 33(16).

Fuhrmann, T., Heck, B., Knöpfler, A., Masson, F., Mayer, M., Ulrich, P., Westerhaus, M., and Zippelt, K. (2012). Recent surface displacements in the Upper Rhine Graben - Preliminary results from geodetic networks. *Tectonophysics*, 602(0):300–315.

Fuhrmann, T., Knöpfler, A., Luo, X., Mayer, M., and Heck, B. (2010). *Zur GNSS-basierten Bestimmung des atmosphrischen Wasserdampfgehalts mittels Precise Point Positioning*. Schriftenreihe des Studiengangs Geodsie und Geoinformatik / Karlsruher Institut fr Technologie, Studiengang Geodsie und Geoinformatik ; 2010,2 ISSN: 1612-9733. KIT Scientific Publishing, Karlsruhe.

Gelfand, A. E., Zhu, L., and Carlin, B. P. (2001). On the change of support problem for spatio-temporal data. *Biostatistics*, 2(1):31–45.

Goldstein, R. (1995). Atmospheric limitations to repeat-track radar interferometry. *Geophysical Research Letters*, 22(18):2517–2520.

Gong, W., Meyer, F., Webley, P., Morton, D., and Liu, S. (2010). Performance analysis of atmospheric correction in InSAR data based on the Weather Research and Forecasting Model (WRF). In *Geoscience and Remote Sensing Symposium (IGARSS), 2010 IEEE International*, pages 2900–2903.

González, P. J., Tiampo, K. F., Camacho, A. G., and Fernández, J. (2010). Shallow flank deformation at Cumbre Vieja volcano (Canary Islands): Implications on the stability of steep-sided volcano flanks at oceanic islands. *Earth and Planetary Science Letters*, 297(3):545–557.

Gotway, C. A. and Young, L. J. (2002). Combining Incompatible Spatial Data. *Journal of the American Statistical Association*, 97(458):632–648.

Gray, A. L., Mattar, K. E., and Sofko, G. (2000). Influence of ionospheric electron density fluctuations on satellite radar interferometry. *Geophysical Research Letters*, 27(10):1451–1454.

Gringarten, E. and Deutsch, C. (2001). Teacher's Aide Variogram Interpretation and Modeling. *Mathematical Geology*, 33(4):507–534.

Hager, W. W. (1989). Updating the inverse of a matrix. *SIAM review*, 31(2):221–239.

Hanssen, R. (2001). *Radar interferometry: data interpretation and error analysis*. Kluwer Academic Publishers.

Herring, T. A. (1992). Modelling atmospheric delays in the analysis of space geodetic data. In *Symposium on Refraction of Transatmospheric Signals in Geodesy, Netherlands Geod. Commis. Ser 36*, 36, pages 157–164, Ned. Comm. voor Geod., Delft.

Hetland, E., Musé, P., Simons, M., Lin, Y., Agram, P., and DiCaprio, C. (2012). Multiscale InSAR time series (MInTS) analysis of surface deformation. *Journal of Geophysical Research: Solid Earth (1978–2012)*, 117(B2).

Hoffmann, J., Galloway, D. L., and Zebker, H. A. (2003). Inverse modeling of interbed storage parameters using land subsidence observations, Antelope Valley, California. *Water Resources Research*, 39(2):1031.

Hofmann-Wellenhof, B., Litchtenegger, H., and Wasle, E. (2008). *GNSS– global navigation satellite systems: GPS, GLONASS, Galileo and more*. Springer-Verlag, Vienna.

Hooper, A. (2008). A multi-temporal InSAR method incorporating both persistent scatterer and small baseline approaches. *Geophysical Research Letters*, 35(16).

Hooper, A., Segall, P., and Zebker, H. (2007). Persistent scatterer interferometric synthetic aperture radar for crustal deformation analysis, with application to Volcán Alcedo, Galápagos. *Journal of Geophysical Research: Solid Earth*, 112(B7).

Hooper, A., Zebker, H., Segall, P., and Kampes, B. (2004). A new method for measuring deformation on volcanoes and other natural terrains using InSAR persistent scatterers. *Geophysical research letters*, 31(23).

Hopfield, H. (1969). Two-quartic tropospheric refractivity profile for correcting satellite data. *Journal of Geophysical research*, 74(18):4487–4499.

Ishimaru, A. (1978). *Wave propagation and scattering in random media*, volume 2. Academic press New York.

Jade, S. and Vijayan, M. (2008). GPS-based atmospheric precipitable water vapor estimation using meteorological parameters interpolated from NCEP global reanalysis data. *Journal of Geophysical Research: Atmospheres (1984–2012)*, 113(D3).

Jin, S. and Luo, O. (2009). Variability and climatology of PWV from global 13-year GPS observations. *Geoscience and Remote Sensing, IEEE Transactions on*, 47(7):1918–1924.

Kampes, B. M. (2005). *Displacement parameter estimation using permanent scatterer interferometry.* PhD thesis, Delft Univ. of Technol., Delft, Netherlands.

Kang, E. L. and Cressie, N. (2011). Bayesian Inference for the Spatial Random Effects Model. *Journal of the American Statistical Association*, 106(495):972–983.

Kang, E. L., Cressie, N., and Shi, T. (2010). Using temporal variability to improve spatial mapping with application to satellite data. *Canadian Journal of Statistics-revue Canadienne De Statistique*, 38:271–289.

Karabatić, A., Weber, R., and Haiden, T. (2011). Near real-time estimation of tropospheric water vapour content from ground based GNSS data and its potential contribution to weather now-casting in Austria. *Advances in Space Research*, 47(10):1691 – 1703.

Katzfuss, M. and Cressie, N. (2009). Maximum likelihood estimation of covariance parameters in the spatial-random-effects model. In *Proceedings of Joint Statistical Meetings Alexandria: American Statistical Association*, pages 3378–90.

Katzfuss, M. and Cressie, N. (2011a). Spatio-temporal smoothing and EM estimation for massive remote-sensing data sets. *Journal of Time Series Analysis*, 32(4):430–446.

Katzfuss, M. and Cressie, N. (2011b). Tutorial on Fixed Rank Kriging (FRK) of CO2 data.

Klebaner, F. C. (2005). *Introduction to Stochastic Calculus with Applications.* Imperial College Press.

Lanyi, G. (1984). Tropospheric delay effects in radio interferometry. *TDA Prog. Rep. 42-78, vol. April*, pages 152–159.

Lark, R. (2000). Estimating variograms of soil properties by the method-of-moments and maximum likelihood. *European Journal of Soil Science*, 51(4):717–728.

Lawrence, M. G. (2005). The Relationship between Relative Humidity and the Dewpoint Temperature in Moist Air: A Simple Conversion and Applications. *Bull. of the American Meteorological Society*, 86(2):225–233.

Leick, A. (2004). *GPS Satellite Surveying.* Wiley & Sons, New Jersey.

Li, F. and Goldstein, R. M. (1990). Studies of multi-baseline spaceborne interferometric synthetic aperture radars. *IEEE Trans. Geoscience and Remote Sensing*, 28:88–97.

Li, J. and Heap, A. D. (2008). *A review of spatial interpolation methods for environmental scientists.* Geoscience Australia, Canberra.

Li, Z., Fielding, E. J., Cross, P., and Muller, J.-P. (2006a). Interferometric synthetic aperture radar atmospheric correction: GPS topography-dependent turbulence model. *Journal of Geophysical Research: Solid Earth*, 111(B2):n/a–n/a.

Li, Z., Muller, J., Cross, P., Albert, P., Fischer, J., and Bennartz, R. (2006b). Assessment of the potential of MERIS near-infrared water vapour products to correct ASAR interferometric measurements. *International Journal of Remote Sensing*, 27(2):349–365.

Li, Z., Muller, J.-P., Cross, P., and Fielding, E. J. (2005). Interferometric synthetic aperture radar (InSAR) atmospheric correction: GPS, Moderate Resolution Imaging Spectroradiometer (MODIS), and InSAR integration. *Journal of Geophysical Research: Solid Earth (1978–2012)*, 110(B3).

Liu, S., Hanssen, R. F., and Mika, A. (2009). Feasibility of retrieving spatial variations of atmospheric phase screen at epochs of SAR acquisitions from SAR interferometry. In *Proceedings of the 2009 Envisat/ERS Symposium*.

Lu, Z. (2007). InSAR imaging of volcanic deformation over cloud-prone areas-Aleutian islands. *Photogrammetric engineering and remote sensing*, 73(3):245.

Luo, X., Mayer, M., and Heck, H. (2008). EXTENDED NEUTROSPHERIC MODELLING FOR THE GNSS-BASED DETERMINATION OF HIGH-RESOLUTION ATMOSPHERIC WATER VAPOUR FIELDS. *Boletim de Ciencias Geodesicas*, 14(2):149–170.

Lyons, S. and Sandwell, D. (2003). Fault creep along the southern San Andreas from interferometric synthetic aperture radar, permanent scatterers, and stacking. *Journal of Geophysical Research: Solid Earth (1978–2012)*, 108(B1).

MacMillan, D. (1995). Atmospheric gradients from very long baseline interferometry observations. *Geophysical Research Letters*, 22(9):1041–1044.

Marini, J. W. (1972). Correction of satellite tracking data for an arbitrary tropospheric profile. *Radio Science*, 7(2):223–231.

Massonnet, D., Rossi, M., Carmona, C., Adragna, F., Peltzer, G., Feigl, K., and Rabaute, T. (1993). The displacement field of the Landers earthquake mapped by radar interferometry. *Nature*, 364(6433):138–142.

Mateus, P., Nico, G., Tomé, R., Catalão, J., and Miranda, P. (2010). Comparison of precipitable water vapor (PWV) maps derived by GPS, SAR interferometry, and numerical forecasting models. In *SPIE, 2008*, pages 782714–782714–7.

Mayer, M., Knöpfler, A., Heck, B., Masson, F., Ulrich, P., and Ferhat, G. (2012). GURN (GNSS Upper Rhine Graben Network): Research Goals and First Results of a Transnational Geo-scientific Network. In Kenyon, S., Pacino, M. C., and Marti, U., editors, *Geodesy for Planet Earth*, volume 136 of *International Association of Geodesy Symposia*, pages 673–681. Springer Verlag series.

Meyer, F., Bamler, R., Jakowski, N., and Fritz, T. (2006). The Potential of Low-Frequency SAR Systems for Mapping Ionospheric TEC Distributions. *Geoscience and Remote Sensing Letters, IEEE*, 3(4):560–564.

Meyer, F., Bamler, R., Leinweber, R., and Fischer, J. (2008). A Comparative Analysis of Tropospheric Water Vapor Measurements from MERIS and SAR. In *Geoscience and Remote Sensing Symposium, 2008. IGARSS 2008. IEEE International*, volume 4, pages IV – 228–IV – 231.

Meyer, F., Gernhardt, S., and Adam, N. (2007). Long-term and seasonal subsidence rates in Urban Areas from Persistent Scatterer Interferometry. In *Urban Remote Sensing Joint Event, 2007*, pages 1–6. IEEE.

Misra, P. and Enge, P. (2001). *Global Positioning System: Signals, Measurements and Performance*. Ganga-Jumuna Press.

Nguyen, H. (2009). *spatial statistical data fusion for remote sensing applications*. PhD thesis, University of California, Los Angeles, USA.

Niell, A. (1996). Global mapping functions for the atmosphere delay at radio wavelengths. *Journal of Geophysical Research: Solid Earth (1978–2012)*, 101(B1):3227–3246.

Onn, F. (2006). *Modeling Water Vapor Using GPS with Application to Mitigating InSAR Atmospheric Distortions*. PhD thesis, Stanford University, Stanford, California, USA.

Onn, F. and Zebker, H. A. (2006). Correction for interferometric synthetic aperture radar atmospheric phase artifacts using time series of zenith wet delay observations from a GPS network. *Journal of Geophysical Research: Solid Earth*, 111(B9).

Osmanoglu, B., Dixon, T. H., Wdowinski, S., Cabral-Cano, E., and Jiang, Y. (2011). Mexico City subsidence observed with persistent scatterer InSAR. *International Journal of Applied Earth Observation and Geoinformation*, 13(1):1–12.

Peltzer, G., Crampé, F., Hensley, S., and Rosen, P. (2001). Transient strain accumulation and fault interaction in the Eastern California shear zone. *Geology*, 29(11):975–978.

Petersen, K. B. and Pedersen, M. S. (2006). The matrix cookbook.

Physorg (2006). Greenhouse theory smashed by biggest stone. http://phys.org/news11710.html.

Pichelli, E., Ferretti, R., Cimini, D., Perissin, D., Montopoli, M., Marzano, F. S., and Pierdicca, N. (2010). Water vapor distribution at urban scale using high-resolution numerical weather model and spaceborne SAR interferometric data. *Natural Hazards and Earth System Science*, 10(1):121–132.

Plattner, C., Wdowinski, S., Dixon, T., and Biggs, J. (2010). Surface subsidence induced by the Crandall Canyon Mine (Utah) collapse: InSAR observations and elasto-plastic modelling. *Geophysical Journal International*, 183(3):1089–1096.

Rocken, C., Hove, T. V., Johnson, J., Solheim, F., Ware, R., Bevis, M., Chiswell, S., and Businger, S. (1995). GPS/STORM-GPS sensing of atmospheric water vapor for meteorology. *Journal of Atmospheric and Oceanic Technology*, 12(3):468–478.

Saastamoinen, J. (1973). Contributions to the theory of atmospheric refraction. *Bulletin Géodésique (1946-1975)*, 107(1):13–34.

Sandwell, D. T. and Price, E. J. (1998). Phase gradient approach to stacking interferograms. *Journal of Geophysical Research: Solid Earth (1978–2012)*, 103(B12):30183–30204.

Schmidt, D., Bürgmann, R., Nadeau, R., and d'Alessio, M. (2005). Distribution of aseismic slip rate on the Hayward fault inferred from seismic and geodetic data. *Journal of Geophysical Research: Solid Earth (1978–2012)*, 110(B8).

Schüler, T. (2001). *On ground-based GPS tropospheric delay estimation*.

Seidel, D. J. (2002). Water vapor: distribution and trends. *Encyclopedia of Global Environmental Change, John Wiley & Sons, Ltd, Chichester*.

Shi, T. and Cressie, N. (2007). Global statistical analysis of MISR aerosol data: a massive data product from NASA's Terra satellite. *Environmetrics*, 18(7):665–680.

Smith, E. and Weintraub, S. (1953). The Constants in the Equation for Atmospheric Refractive Index at Radio Frequencies. *Proceedings of the IRE*, 41(8):1035–1037.

Tatarskii, V. (1961). *Wave propagation in a turbulent medium*. McGraw-Hill, New York.

Thayer, G. D. (1974). An improved equation for the radio refractive index of air. *Radio Science*, 9(10):803–807.

Treuhaft, R. N. and Lanyi, G. E. (1987). The effect of the dynamic wet troposphere on radio interferometric measurements. *Radio Science*, 22(2):251–265.

Wadge, G., Webley, P., James, I., Bingley, R., Dodson, A., Waugh, S., Veneboer, T., Puglisi, G., Mattia, M., Baker, D., et al. (2002). Atmospheric models, GPS and InSAR measurements of the tropospheric water vapor field over Mount Etna. *Geophysical Research Letters*, 29(19):1905.

Wald, L. (1983). Some examples of the use of structure functions in the analysis of satellite images of the ocean. *Photogrammetric Engineering and Remote Sensing*, 55:1487–1490.

Webley, P., Bingley, R., Dodson, A., Wadge, G., S.J., W., and James, I. (2002). Atmospheric water vapour correction to InSAR surface motion measurements on mountains: results from a dense GPS network on Mount Etna. *Physics and Chemistry of the Earth, Parts A/B/C*, 27(4).

Williams, S., Bock, Y., and Fang, P. (1998). Integrated satellite interferometry: Tropospheric noise, GPS estimates and implications for interferometric synthetic aperture radar products. *Journal of Geophysical Research: Solid Earth*, 103(B11):27051–27067.

Wright, T., Parsons, B., and Fielding, E. (2001). Measurement of interseismic strain accumulation across the North Anatolian Fault by satellite radar interferometry. *Geophysical Research Letters*, 28(10):2117–2120.

Wright, T. J., Lu, Z., and Wicks, C. (2003). Source model for the M w 6.7, 23 October 2002, Nenana Mountain earthquake (Alaska) from InSAR. *Geophysical Research Letters*, 30(18):1974.

Zebker, H. and Villasenor, J. (1992). Decorrelation in interferometric radar echoes. *Geoscience and Remote Sensing, IEEE Transactions on*, 30(5):950–959.

Zebker, H. A. and Goldstein, R. M. (1986). Topographic mapping from interferometric synthetic aperture radar observations. *Journal of Geophysical Research: Solid Earth (1978–2012)*, 91(B5):4993–4999.

Zebker, H. A., Rosen, P. A., Goldstein, R. M., Gabriel, A., and Werner, C. L. (1994). On the derivation of coseismic displacement fields using differential radar interferometry: The Landers earthquake. *Journal of Geophysical Research: Solid Earth (1978–2012)*, 99(B10):19617–19634.

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.

Zumberge, J., Heflin, M., Jefferson, D., Watkins, M., and Webb, F. (1997). Precise point positioning for the efficient and robust analysis of GPS data from large networks. *Journal of Geophysical Research*, 102(B3):5005–5017.

Acknowledgements

This dissertation would not have been possible without the support of many people. I would like to take the opportunity to express my sincere gratitude to all those who have helped me in the past four years.

First and foremost, I would like to express my deep thanks to Professor Stefan Hinz, who served as my principal supervisor, for his guidance, support and mentorship. My gratitude also goes to Professor Bernhard Heck for serving as my second reviewer. I would like to express my heartfelt gratitude to Dr. Franz Meyer for his support, useful discussions and insightful suggestions.

I would like to extend my sincere thanks to all my colleagues at the Institute of Photogrammetry and Remote Sensing and the Geodetic Institute for their support, help and precious advices.

Also, many thanks go to all data providers, without their support this work would not have been be possible.

I am deeply thankful and gratitude to my family for their love and sacrifices. To them, I owe everything. I would also like to thank all my friends for continuing to encourage me throughout.