Bender, M., Dick, G., Ge, M., Deng, Z., Wickert, J., Kahle, H.-G., Raabe, A., Tetzlaff, G. (2011): Development of a GNSS water vapour tomography system using algebraic reconstruction techniques. - Advances in Space Research, 47, 10, 1704-1720

DOI: 10.1016/j.asr.2010.05.034

Development of a GNSS Water Vapor Tomography System Using Algebraic Reconstruction Techniques

Michael Bender*[,a], Galina Dick[a], Maorong Ge[a], Zhiguo Deng[a], Jens Wickert[a], Hans-Gert Kahle[a], Armin Raabe[b], Gerd Tetzlaff[b]

[a] *Helmholtz-Zentrum Potsdam, Deutsches GeoForschungsZentrum, Telegraphenberg, 14473 Potsdam, Germany*
[b] *Institut für Meteorologie, Universität Leipzig, Stephanstraße 3, 04103 Leipzig, Germany*

Abstract

A GNSS water vapour tomography system developed to reconstruct spatially resolved humidity fields in the troposphere is described. The tomography system was designed to process the slant path delays of about 270 German GNSS stations in near real-time with a temporal resolution of 30 minutes, a horizontal resolution of 40 km and a vertical resolution of 500 m or better. After a short introduction to the GPS slant delay processing the framework of the GNSS tomography is described in detail. Different implementations of the iterative algebraic reconstruction techniques (ART) used to invert the linear inverse problem are discussed. It was found that the multiplicative techniques (MART) provide the best results with least processing time, i. e. a tomographic reconstruction of about 26000 slant delays on a 8280 cell grid can be obtained in less than 10 minutes. Different iterative reconstruction techniques are compared with respect to their convergence behaviour and some numerical parameters. The inversion can be considerably stabilized by using additional non-GNSS observations and implementing various constraints. Different strategies for initialising the tomography and utilizing extra information are discussed. At last an example of a reconstructed field of the wet refractivity is presented and compared to the corresponding distribution of the integrated water vapour, an analysis of a numerical weather model (COSMO-DE) and some radiosonde profiles.

Key words: GPS, tomography, water vapour, algebraic reconstruction techniques (ART)

1. Introduction

GNSS based atmosphere sounding has recently become a widely used technique (Jin et al., 2007). Zenith total delays (ZTDs) and integrated water vapour (IWV) data as provided by GPS analysis centres are commonly accepted atmospheric observations (Jin and Luo, 2009; Jin et al., 2009). Several European weather services make use of these data to improve their operational weather forecasts by assimilating near real-time ZTDs provided by different GNSS analysis centres (http://egvap.dmi.dk). The German Research Center for Geosciences (GFZ) in Potsdam is a large European analysis centre which provides IGS products and atmospheric products in near real-time and high quality postprocessed products as well. ZTD and IWV data of an increasing number of German and European GPS stations are available from the GFZ since 2002. Currently, about 270 German stations are analysed operationally. Since 2007 the slant total delays, i. e. the GPS signal delays along each single satellite receiver path are available with a sampling rate of 2.5 minutes.

The IWV is a measure of the total amount of water vapour above a certain station and IWV maps based on a dense network of GPS stations provide detailed information on the water vapour distribution. Unfortunately, no information about the vertical profiles is available. The GFZ made therefore attempts to extend the established ZTD/IWV products by making use of the STD data which can be combined to a well resolved three dimensional representation of the humidity above Germany. A GPS water vapour tomography system has been developed in cooperation with the Leipzig Institute for Meteorology (LIM) which can complement the already available atmospheric GPS products in near real-time.

In the last decade several campaigns demonstrated the potential of the GPS tomography to provide spatially resolved humidity fields (Seko et al., 2000; Flores et al., 2001; Gradinarsky and Jarlemark, 2004; Champollion et al., 2005; de Haan and van der Marel, 2008). But also some problems connected to this technique became apparent. Especially the sub optimal spatial distribution of the existing GPS stations and the noise of the STDs have a negative impact on the quality of the tomographically reconstructed humidity fields. The high temporal and spatial variability of the information contained in the GPS slant data sets (Bender et al., 2009) is another fundamental problem which complicates the continuous reconstruction of humidity fields with a rather uniform quality. However, a nation-

*Corresponding author
Email address: bender@gfz-potsdam.de (Michael Bender)

wide GPS tomography system was established in Switzerland and provides 3D water vapour fields for assimilation into the aLMo numerical weather model (Troller et al., 2006a,b). The rapidly increasing number of GPS/GNSS applications will be further enhanced by new navigation satellite systems like Galileo, Compass and the renewed GLONASS and advanced processing strategies will provide more reliable STD data. Attempts are made to densify existing geodetic GNSS networks with inexpensive single-frequency receivers and to customise the receiver distribution for meteorological applications (Rocken et al., 2000; Deng et al., 2009).

Information about the atmospheric state as provided by GNSS techniques is not only important for meteorological applications but could in near future also become interesting for GNSS positioning. The uncertain atmospheric state, especially the water vapour distribution, is currently one of the most important limitations for high precision positioning. If the refractivity of the neutral atmosphere could be estimated with sufficient accuracy by means of the GNSS tomography it would be possible to make much more realistic atmosphere corrections. Instead of using mapping functions a raytracer could estimate the atmospheric delay for each single slant path. Ray-tracing techniques using refractivity fields from numerical weather models are already successfully used in precise positioning (Hobiger et al., 2008a). GNSS atmosphere sounding and GNSS positioning could improve each other iteratively.

The GNSS water vapour tomography package described here was designed to provide spatially resolved humidity fields for Germany in near real-time. The software runs with the data already available at the GFZ and provides reconstructed humidity fields with a temporal resolution of less than 1 hour and a computing time of less than 15 minutes in order to bring all atmospheric products (ZTD, IWV, STD, SWD and humidity fields) to completion with a maximum delay of less than 30 minutes, i. e. in near real-time. The quality of the near real-time products is comparable to the quality of post-processed data with a bias well below 0.1 mm in the IWV corresponding to about 0.5 mm in the ZTD (Dick et al., 2001). An iterative approach for solving the basic tomographic equations has been chosen to avoid time consuming matrix operations. Several algorithms of the algebraic reconstruction techniques (ART) were implemented and tested. The operationally available STD data and several meteorological observations as well are used to reconstruct the humidity field. Different weights can be applied to the data and diverse options can be varied to tune the results. An example of a reconstructed field of the wet refractivity is given and compared to a COSMO-DE field and some radiosonde profiles.

2. GPS tomography

The total delay of the GPS signal due to the neutral atmosphere is given by (Bevis et al., 1992)

$$STD = 10^{-6} \cdot \int_S N \, ds \tag{1}$$

where N is the atmospheric refractivity and S is the GPS signal path. The small contribution of the geometric path delay is neglected. In case of the GPS atmosphere sounding the STDs are the observations and the unknown refractivity field N has to be reconstructed from a large number of such observations. This defines an inverse problem which turns out to be ill-posed, i. e. solutions are in general not unique and not stable. In case of the GPS tomography one has also to deal with incomplete input data as the spatial coverage of the atmosphere by GPS slant paths is not optimal for a tomographic reconstruction and highly variable in space and time (Bender et al., 2009; Bender and Raabe, 2007).

Equ. 1 can be solved by discretizing the atmosphere using a spatial grid. Inside each grid cell or voxel the refractivity is assumed to be constant. In discrete form equ. 1 becomes

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{m} \, , \tag{2}$$

where the vector $\boldsymbol{m}$ represents the observations, i. e. the slant path delays, $\boldsymbol{x}$ describes the current state of the atmosphere, i. e. the refractivity N_j in each voxel j, and the kernel matrix $\boldsymbol{A}$ defines the mapping of the state $\boldsymbol{x}$ on the observations $\boldsymbol{m}$. The matrix elements a_{ij} are the subsections of the ith slant path in the jth grid cell.

Depending on the observations different atmospheric quantities can be reconstructed: Equ. 1 which correlates the STDs with the refractivity N is also true for the dry and wet part, i. e. the slant wet delays (SWDs) are correlated with the wet refractivity N_w and the slant dry delays with the dry refractivity N_d. The Smith and Weintraub formula (Smith and Weintraub, 1953) or the Thayer formula (Thayer, 1974) relate N_w to the atmospheric humidity:

$$N_\mathrm{w} = Z_\mathrm{w}^{-1} \left(k_2 \frac{e}{T} + k_3 \frac{e}{T^2} \right) \tag{3}$$

where e is the partial pressure of water vapour, T is the temperature and Z_w^{-1} is the inverse compressibility factor of water vapour. The constants $k_2 = 70.4$ K hPa^{-1} and $k_3 = 3.739 \cdot 10^5$ K^2 hPa^{-1} can, e. g., be found in Bevis et al. (1994) and no real gas corrections are made by Smith and Weintraub ($Z_\mathrm{w}^{-1} = 1$). With extra information about the atmospheric temperature field it is therefore possible to obtain the spatial distribution of the absolute humidity.

Another approach is to estimate the slant water vapour (SWV), i. e. the integrated water vapour along the slant path, prior to the tomography. Water vapour radiometer or specific GPS processing techniques provide this kind of information. Using SWV input data the tomography

results in the absolute humidity H_{abs} without further assumptions:

$$SWV = \int_S H_{\mathrm{abs}} \; ds \tag{4}$$

The ill-posed character of the inverse problem depends significantly on the exact form of the kernel matrix which represents the geometric aspects of the problem. $\boldsymbol{A}$ is defined by the voxel structure and the slant paths inside each voxel. The voxel structure can -within reasonable limits- freely be chosen while the slant paths are part of the observations. Usually, the ray bending is neglected and the slant path S is assumed to be a straight line (de Brito Mendes, 1999). Under this assumption equ. 2 becomes a linear inverse problem. As a good spatial coverage of the atmosphere by the slant paths is required, it is in most cases necessary to combine observations from a certain period, e.g. 30 minutes. In this case the temporal resolution has also a considerable impact on $\boldsymbol{A}$. The kernel matrix is therefore not only determined by the inverse problem itself but also by several assumptions made to simplify the problem. Solving equ. 2 is only one aspect of the problem, efforts should also be made to modify the specification of the basic problem in order to reduce its ill-posed character.

The dimension of the solution space of equ. 2 can most easily be reduced by appending extra information and constraints. Synoptic observations, radiosonde profiles, radio occultation profiles or other independent observations can be used to improve the tomographic reconstruction (see sec. 4.4, Foelsche and Kirchengast (2001)).

3. GPS data processing

The tropospheric delay is one of the major error sources in the precise positioning with GNSS data. Its impact on the vertical components is still a crucial issue especially for (real-time) kinematic positioning because of their high correlation. Usually, the tropospheric delay includes the hydrostatic and the wet components. Most of the models are derived under the assumption of horizontal isotropy with zenith delays and mapping functions for both hydrostatic and wet components. 'The former one can be modelled from climatological data or more precisely using surface pressure observations but not the latter one because of its temporal and spatial changes. In the geodetic positioning, the tropospheric delay is modelled with an a priori model, e. g. Saastamoinen, and the remaining zenith delay is parameterised as a stochastic process and to be estimated in the data processing. The estimated delays demonstrated to be very useful information for meteorological studies and weather forecasts (Bevis et al., 1992; Rocken et al., 1993). However, the troposphere is very often not isotopic. To account for the anisotropy of refractivity in the troposphere, horizontal gradients are introduced in the processing (Chen and Herring, 1997). The anisotropy is also considered by using more realistic mapping functions derived from numerical weather models (Rocken et al., 2001; Boehm et al., 2006, 2008).

For retrieving tropospheric delays in real-time, mapping functions from a predicted numerical model cannot reflect the true anisotropic distributions. It was proven that the inaccurately modelled tropospheric delays remain in the residuals of the double-differenced observations (Ware et al., 1997) and the slant delays along the ray paths can be obtained from the estimated zenith delay and gradients and the double-differenced residuals after removing the multi-path effects (Alber et al., 2000). The slant delays derived from the double-differenced observations are validated for example by Braun et al. (2001) by comparing the GPS derived slant water vapour with pointed radiometer observations.

There are two major schemes for precise GPS data processing, the network solution mode and the precise point positioning mode (PPP). In the network solution mode, at least two stations are involved and satellite and receiver clock biases must be estimated if zero-differenced observations are used. Alternatively, the biases can be removed by forming double-differenced observations. Satellite orbits can be estimated together with proper weights if they are not precisely known (Ge et al., 2002). Theoretically there is no difference using zero- or double-differenced observations. The former one will provide directly the ray path residuals while for the latter one the double-differenced residuals must be transformed to the zero-differenced residuals. In the PPP mode, precise orbits and satellite clock corrections are used and fixed in the data analysis (Zumberge et al., 1997). As the satellite orbits and clocks as well as Earth Rotation Parameters (ERP) are fixed to known values, there are no common parameters between stations and the data can be processed station by station. This is also the most important advantage of the PPP mode, for networks with hundreds of stations the data can be processed distributedly in different processes (distributed computing). However, the determination of precise orbits and clock corrections, especially in (near) real-time, from a global network is a tough work and can only be carried out by institutes like IGS analysis centres. Precise orbits and clocks are required to estimate the rather small anisotropic contributions to the slant delays. In principle, these data are available from IGS but not in real-time and the products from the IGS real-time pilot project (http://www.rtigs.net) might be available soon but its accuracy is still to be improved.

At GFZ the EPOS software (Gendt et al., 2001a, 2004) is used for GPS meteorological applications in both network solution and PPP modes, in near real-time and post processing. Zero-differenced phase and range observations are used with proper weights scaled according to elevations. For network solutions, satellite orbits and ERPs are fixed to the ultra rapid products from the GFZ IGS analysis centre, satellite and receiver clock biases are estimated epoch by epoch. Tropospheric delays are corrected using the Saastamoinen model for the ZTD and the Global Mapping Function (GMF) (Boehm et al., 2006) or Niell mapping function (Niell, 1996) and the remaining tropo-

spheric impact is parameterised with zenith delays every 15 minutes at each station and gradients in east and north directions every hour, respectively. The data are processed using the sliding window approach with a window width of 12 hours and a forward step-size of one hour. For the operational data processing of the SAPOS network including about 270 stations within and around Germany, the PPP mode is used. For the (near) real-time clock estimation the IGS ultra-rapid products are used as base and are improved by making use of some permanent stations around Germany. The stations are scheduled in different tasks or computers and processed parallel. The other variables are kept the same as for the network solution. With the estimated zenith delays, gradients and the mapping functions, and of course the ray path residuals, the slant total delays (STDs) are retrieved according to

$$STD = m_{\mathrm{h}} \cdot ZHD + m_{\mathrm{w}} \cdot [ZWD + \cot\varepsilon\,(G_{\mathrm{N}}\cos\phi + G_{\mathrm{E}}\sin\phi)] + \delta \quad (5)$$

where ZHD and ZWD are the hydrostatic and the wet zenith delay, respectively, m_{h} and m_{w} are the hydrostatic and the wet mapping functions, G_{N} and G_{E} are the delay gradient parameters in the northern and eathern direction, ε is the elevation, ϕ is the geographic latitude and δ is the postfit phase residual.

3.1. Separation of the wet delay

As described in the previous section precise STD data can be obtained by GPS processing techniques but the GPS signals provide no information to separate the wet part from the hydrostatic part of the delay. Meteorological observations and some assumptions about the vertical atmospheric structure are required to separate both terms. One widely used model to estimate the slant hydrostatic delay from surface meteorological observations was developed by Saastamoinen (1972, 1973). The slant hydrostatic delays can either be obtained using a formula given by Saastamoinen (1973)

$$SHD = \frac{0.002277}{\cos z} \cdot \left(p_0 - 0.155471\, e_0 - B\tan^2 z\right) + \delta_{\mathrm{R}} \quad (6)$$

or from the zenith hydrostatic delay (Saastamoinen, 1972; Davis et al., 1985; Bevis et al., 1992)

$$ZHD = \frac{(0.0022768 \pm 0,0000005)\, p_0}{f(\phi, H)} \quad (7)$$

which can later be mapped onto the slant path by using more advanced mapping functions, e. g. the dry Niell mapping function (Niell, 1996) or the Vienna mapping function (Boehm et al., 2006, 2008). p_0 and e_0 are the total atmospheric pressure and the partial pressure of the water vapour at the antenna reference point both given in hPa, z is the zenith distance, B and δ_{R} are correction terms given, e. g., in Saastamoinen (1973) or Hofmann-Wellenhof et al. (1993). The function $f(\phi, H)$ is given by

$$f(\phi, H) = 1 - 0.00266\cos(2\phi) - 0.00028\, H \quad (8)$$

where ϕ is the geodetic latitude and H is the height above geoid in km. Using equ. 7 has the advantage that pressure observations are sufficient while the e_0 term in equ. 6 requires also temperature and humidity information.

Precise meteorological observations are available for some GPS stations equipped with a meteosensor. In other cases the values of p, T and e must be interpolated from data provided by a dense network of synoptic stations. In this study the hourly synoptic data of the German Meteorological Service (DWD) were used together with the data of the GPS meteo stations to feed a least-squares collocation model (Moritz, 1973; Hirter, 1996). The pressure and temperature can be interpolated with an error of ± 1 hPa and ± 1 K, respectively. The humidity is much more variable and the error of the partial pressure of water vapour is about ± 5 hPa, considerably depending on the distances to the surrounding meteorological stations. All available data within a 2 hours period are used to interpolate the synoptic observations at the GPS stations at the right time.

After the SHD was estimated the slant wet delay (SWD) can easily be obtained from the difference to the STD: SWD = STD - SHD. The SWD can now be used by the tomography to reconstruct the wet refractivity field N_{w}.

The reconstruction of the absolute humidity field requires the estimation of the slant water vapour (SWV). Bevis et al. (1994) developed a widely used method to obtain the SWV from the SWD and the weighted mean temperature T_{m}. However, T_{m} must be estimated from the surface temperature and an average vertical temperature profile which introduces even more assumptions about the atmospheric state into the tomography. The tomography is very sensitive to small variations in the input data and SWDs are used in this work to stay as close as possible to the GPS observations. The humidity field can in principle also be obtained from the wet refractivity and an extrapolated temperature field.

4. GNSS tomography at the GFZ

The GPS processing centre of the GFZ provides several GPS products for meteorological applications in near real-time (Gendt et al., 2004). ZTD and IWV observations are processed with a temporal resolution of 15 minutes and give a good impression of the horizontal water vapour distribution above Germany (Gendt et al., 2001b; Dick et al., 2001; Tomassini et al., 2002). The ZTD data are assimilated by several European weather services to improve the weather forecasts and especially the precipitation forecasts (Poli et al., 2007; Bengtsson and Hodges, 2005). Recently, the processing of STDs became operational and slant delays observed along a large number of different views through the atmosphere are now available. About 8 GPS satellites can be tracked simultaneously by each station which leads to 150 – 200 STDs per station and per hour assuming a sampling rate of 2.5 minutes. Approximately 270 German stations provide in total between 40000 and 60000 observations per hour. This large

data set yields spatially and temporally resolved information about the atmospheric state, e. g. the water vapour distribution. To exploit these data the GFZ developed in cooperation with the LIM the GNSS water vapour tomography system PORTOS (Potsdam Real-Time Tomography System). This tomography system was designed to extend the 2D ZTD/IWV data sets to the third dimension in near real-time and to provide spatially resolved humidity fields with virtually the same delay of max. 30 minutes as the IWV and STD products.

4.1. Realisation of the tomography system

The basic task of the GNSS tomography is to invert the linear equation defined by equ. 2, i. e. to estimate the refractivity field from a set of integral observations. However, before the inversion can be started the specific problem must be discretised, linearised and optionally some extra information must be converted to fit into the linear model of equ. 2.

The atmospheric refractivity is discretised to a voxel structure using a regular WGS84 grid with arbitrarily definable horizontal and vertical resolutions. The ellipsoidal horizontal layers of this grid account for a smooth alignment to Earth's surface. The refractivity N within each voxel is assumed to be constant. The horizontal dimensions of the grid can be customised to the desired region. The vertical extension of the grid should always cover the whole troposphere up to ∼10 km in order to represent the total IWV. Otherwise, the grid would contain only a fraction of the atmospheric water vapour and correction terms would be necessary to estimate the water vapour between the upper grid level and the tropopause.

The chosen grid and the observed slant paths define the kernel matrix $\boldsymbol{A}$, where the a_{ij} are given by the slant subpaths in each voxel. As the GPS signal propagates only through a limited number of grid cells most $a_{ij} = 0$ and $\boldsymbol{A}$ is a large sparse matrix. Discretising the problem in this way does not lead to a linear form as the signal propagates in general on a curved path. The ray bending is therefore neglected which is a good approximation for elevations above 10° (de Brito Mendes, 1999). Under this assumption the slant paths and with it the matrix $\boldsymbol{A}$ are independent from the refractivity and must not be adjusted by the inversion process. Slant paths leaving the grid at the horizontal boundaries are rejected to avoid any assumptions about the humidity distribution outside the grid.

Integrated observations of other remote sensing systems like water vapour radiometers, sun photometers or Lidars can be used in exactly the same way while point observations from, e. g. synoptic stations or radio sondes must be handled differently (see section 4.4).

Combining all available observations and constraints the linear tomographic problem has the form

$$\begin{pmatrix} \boldsymbol{m}_{\mathrm{GPS}} \\ \boldsymbol{m}_{\mathrm{ZWD}} \\ \boldsymbol{m}_{\mathrm{int}} \\ \boldsymbol{m}_{\mathrm{pt}} \\ \boldsymbol{c} \end{pmatrix} = \begin{pmatrix} \boldsymbol{A}_{\mathrm{GPS}} \\ \boldsymbol{A}_{\mathrm{ZWD}} \\ \boldsymbol{A}_{\mathrm{int}} \\ \boldsymbol{A}_{\mathrm{pt}} \\ \boldsymbol{A}_{\mathrm{c}} \end{pmatrix} \cdot \boldsymbol{x} \tag{9}$$

where $\boldsymbol{m}_{\mathrm{GPS}}$, $\boldsymbol{A}_{\mathrm{GPS}}$ belong to the GPS slant delay observations, $\boldsymbol{m}_{\mathrm{ZWD}}$, $\boldsymbol{A}_{\mathrm{ZWD}}$ to the zenith wet delays computed together with the integrated water vapour (IWV), $\boldsymbol{m}_{\mathrm{int}}$, $\boldsymbol{A}_{\mathrm{int}}$ to other path integrated observations, $\boldsymbol{m}_{\mathrm{pt}}$, $\boldsymbol{A}_{\mathrm{pt}}$ to point observations and $\boldsymbol{c}$, $\boldsymbol{A}_{\mathrm{c}}$ to any constraints formulated in matrix notation. The dimensions of the vectors $\boldsymbol{m}$ and $\boldsymbol{x}$ are given by the number of observations p and the number of grid cells q. $\boldsymbol{A}$ is therefore a $p \times q$ matrix. Regarding a horizontal resolution of about 40 km, a vertical resolution of 500 m and a temporal resolution of 30 minutes a grid with approx. 11000 cells would be required for Germany which leads to $\sim 200 \cdot 10^6 - 400 \cdot 10^6$ matrix elements. Future applications which will use Galileo and GLONASS observations and permit an improved spatial resolution can easily lead to matrices of $> 5 \cdot 10^9$ elements.

4.2. Tomographic reconstruction using algorithms of the ART family

Algebraic reconstruction techniques (ART) have successfully been used to reconstruct the total electron content (TEC) of the ionosphere (Stolle, 2004; Stolle et al., 2006; Zhai and Cummer, 2005; Jin and Park, 2007; Jin et al., 2008). These techniques combine high numerical stability even under bad conditions with computational efficiency and can also be applied to the tropospheric water vapour tomography. Several reconstruction algorithms of the ART family have been implemented (Subbarao et al., 1997; Kunitsyn and Tereshchenko, 2003; Raymund, 1995), i. e. the original additive ART algorithm (Gordon et al., 1970; Censor, 1983), several multiplicative algebraic reconstruction techniques (MART) (Gordon et al., 1970) and the simultaneous iterations reconstruction technique (SIRT) (Gilbert, 1972).

The ART algorithms are iterative techniques which process observation by observation. No matrix inversion has to be performed and it is not even necessary to create the large sparse matrix $\boldsymbol{A}$ in computer memory. Only the two vectors $\boldsymbol{x}$, $\boldsymbol{m}$ and a data structure containing the slant subpaths in each voxel are required to solve the equations given below. This leads to an efficient usage of computer memory and computing time.

The algorithms consist basically of two loops. The inner loop (index i, $i = 1 \ldots p$) processes observation by observation and applies an adequate correction to each grid cell (index j, $j = 1 \ldots q$). After all observations have been executed the next iteration (index k) is started in the outer loop. The iteration is stopped as soon as a good result has been obtained (see section 4.3). The number of necessary iterations depends on the quality of the initial

field, the data quality and other parameters. Usually 100 – 200 iterations are required.

The relaxation parameter λ gives the weight of the correction term computed for each voxel with respect to the initial voxel value. λ must be carefully adjusted for each algorithm (see section 5). In general it is possible to define individual relaxation parameters λ_i for each observation (Lu and Yin, 2004) but in this work only one constant value of λ is chosen. Large values of λ lead in most cases to faster convergence but also to more pronounced artefacts in the resulting field and can sometimes initiate oscillations.

The result of the reconstruction depends on the order of the observations within the vector $\boldsymbol{m}$. The simultaneous iterative reconstruction technique (SIRT) algorithm was developed to avoid this effect. A comparative study with different ART algorithms and reordered observations showed that this effect is negligible compared to the impact of other parameters like the relaxation parameter λ, the initial field or the number of iterations. Besides, SIRT does not lead to better results than the other algorithms (see sec. 5).

The following members of the ART family were so far implemented (Kunitsyn and Tereshchenko, 2003). $\boldsymbol{A}^i$ indicates the ith row of the matrix $\boldsymbol{A}$ and the inner product $\left\langle \boldsymbol{A}^i, \boldsymbol{x}^k \right\rangle$ is the backprojection of the ith observation after the kth iteration.

Original ART. The original ART algorithm estimates a correction term for each grid cell and adds this term to the preceding value. It is therefore an additive technique. A geometric interpretation of the original ART algorithm can be found in Kak and Slaney (1999).

$$\boldsymbol{x}^{k+1} = \boldsymbol{x}^k + \lambda \frac{m_i - \left\langle \boldsymbol{A}^i, \boldsymbol{x}^k \right\rangle}{\left\langle \boldsymbol{A}^i, \boldsymbol{A}^i \right\rangle} \boldsymbol{A}^i \tag{10}$$

MART – multiplicative techniques. The multiplicative techniques also estimate a correction term but this term is multiplied to the preceding value of the corresponding grid cell. This leads in general to a faster convergence compared to additive techniques. Several MART algorithms with different exponents have been suggested:

MART1:

$$x_j^{k+1} = x_j^k \cdot \left(\frac{m_i}{\left\langle \boldsymbol{A}^i, \boldsymbol{x}^k \right\rangle} \right)^{\frac{\lambda A_j^i}{\sqrt{\left\langle \boldsymbol{A}^i, \boldsymbol{A}^i \right\rangle}}} \tag{11}$$

MART2:

$$x_j^{k+1} = x_j^k \cdot \left(\frac{m_i}{\left\langle \boldsymbol{A}^i, \boldsymbol{x}^k \right\rangle} \right)^{\frac{\lambda A_j^i}{\left\langle \boldsymbol{A}^i, \boldsymbol{A}^i \right\rangle}} \tag{12}$$

MART3:

$$x_j^{k+1} = x_j^k \cdot \left(\frac{m_i}{\left\langle \boldsymbol{A}^i, \boldsymbol{x}^k \right\rangle} \right)^{\lambda A_j^i \frac{\left\langle \boldsymbol{A}^i, \boldsymbol{x}^k \right\rangle}{\left\langle \boldsymbol{A}^i, \boldsymbol{A}^i \right\rangle}} \tag{13}$$

DART. DART is the result of a Taylor expansion of MART2. It consists of an additive and a multiplicative part.

$$x^{k+1} = x^k \cdot \left(1 + \lambda \frac{m_i - \left\langle \boldsymbol{A}^i, \boldsymbol{x}^k \right\rangle}{\left\langle \boldsymbol{A}^i, \boldsymbol{A}^i \right\rangle} \boldsymbol{A}^i \right) \tag{14}$$

SIRT. All algorithms discussed so far iterate through the given vector of experimental data $\boldsymbol{m}$ and the result of the reconstruction depends on the order of the data within this vector. To avoid such a behaviour the SIRT algorithm has been developed and further refined, e. g., by Hobiger et al. (2008b). It evaluates only one correction term for each grid cell which considers all experimental data m_i. The correction term is therefore independent of the order within $\boldsymbol{m}$.

$$x_j^{k+1} = x_j^k + \sum_i \lambda \, a_{ij} \, \frac{m_i - \left\langle \boldsymbol{A}^i, \boldsymbol{x}^k \right\rangle}{\left\langle \boldsymbol{A}^i, \boldsymbol{A}^i \right\rangle} \tag{15}$$

4.3. Estimation of the reconstruction error, stop criteria

The ART algorithms require several iterations to reach good correspondence with the experimental data. Therefore, criteria have to be defined which describe the convergence behaviour and which can be used to stop the iteration in an optimal way. One possibility is to compare the state $\boldsymbol{x}^k$ of consecutive iterations k, $k+1$, ... The procedure can be stopped if the system converges to a stable state and the difference between the fields is below some threshold. Another possibility is based on the back projection of the current state on the experimental data. A good result has been obtained if the backprojection $A\boldsymbol{x}^k = \boldsymbol{m}^k$ is close to the experimental data $\boldsymbol{m} = \boldsymbol{m}^0$, i. e. $\left| \boldsymbol{m}_0 - \boldsymbol{A}\boldsymbol{x}^k \right| = \min$.

However, due to the inverse character of the problem, these criteria alone cannot describe the quality of the result and it is in general a severe problem to define suitable stop criteria. A set of parameters must be defined which describes the quality of the reconstruction during the iteration and requires less computing time than the iteration itself. It must also be considered that the iteration might at a certain point lead to worse results or to oscillations in some parameters.

The four parameters $\Delta_1, \Delta_2, \delta, \sigma$ are currently computed between consecutive iterations and used to stop the iteration at a certain state. They are also useful to describe the convergence behaviour and to estimate the quality of the reconstruction.

Relative deviation. The relative deviation Δ between consecutive iterations gives a first impression of the convergence behaviour but cannot be used to describe the quality

of the reconstruction. Δ can be defined in different ways depending on the norm $\frac{||\boldsymbol{x}^{k+1}-\boldsymbol{x}^k||}{||\boldsymbol{x}^k||}$, e. g.:

$$\Delta_1 = \frac{\sqrt{\sum_i \left(x_i^{k+1} - x_i^k\right)^2}}{\sqrt{\sum_i {x_i^k}^2}} \tag{16}$$

$$\Delta_2 = \frac{\max_i \left|x_i^{k+1} - x_i^k\right|}{\max_i \left|x_i^k\right|} \tag{17}$$

Backprojection. The backprojection $\boldsymbol{m}^k$ can be used to define the mean deviation between the reconstruction and the experimental data $\boldsymbol{m}^0$:

$$\delta = \frac{1}{I}\sum_{i=1}^{I}\left(m_i^k - m_i^0\right) \tag{18}$$

Variance. More meaningful is the variance of this quantity:

$$\sigma = \sqrt{\frac{1}{I-1}\sum_{i=1}^{I}\left(\left\{m_i^k - m_i^0\right\} - \delta\right)^2} \tag{19}$$

4.4. Constraints

The solution of an ill-posed problem can lead to completely unreliable results. Very small variations in the input data can under certain conditions lead to a completely different result and the application of the basic equation (equ. 2) alone will therefore in general not lead to stable solutions. The situation can be considerably improved by applying several constraints which introduce a priory information of the system and stabilise the reconstruction. So far the following set of constraints was implemented:

- Initial field, relaxation parameter λ used to weight the result of the reconstruction against the initial value.
- Surface observations combined to the wet refractivity at the grid ground level, included to the set of observations.
- IWV/ZWD observations to adjust the total amount of the atmospheric humidity.
- At the upper level of the grid the humidity can be assumed to be near zero at least if the tomography grid reaches the tropopause (> 10 km).
- Inter-voxel constraints limiting the difference between neighboured voxels.

4.4.1. Initialisation

The initial field which serves as the start of the iterative reconstruction has a large impact on the quality of the result. As shown in Bender et al. (2009) there are situations where more than 50 % of the voxel are not touched by any slant path and will therefore retain their initial value. This can somewhat be smoothed by applying inter voxel constraints but wrong initial assumptions will considerably reduce the quality of the result even if they appear only in a limited region.

Several methods have been implemented to provide an initial state for the reconstruction: An easy way is to use the standard atmosphere as a first guess. A much more realistic state can be obtained by using the synoptic observations and to extrapolate a three dimensional field of the wet refractivity. Another possibility is to take an analysis field or a forecast of a numerical weather model.

If numerical weather forecasts are available the model fields can be interpolated on the tomography grid and used to initialise the tomographic reconstruction. The GPS tomography applies in this case corrections to the forecast. The quality of the tomographic reconstruction depends considerably on the quality of the model forecast. However, this approach requires in time access to model data and it might be preferable to assimilate the slant data directly to the model (Zus et al., 2008; de Haan and van der Marel, 2008).

For near real-time applications only observations should be used to estimate the current state of the atmosphere. The least-squares collocation (Troller et al., 2006a; Hirter, 1996; Moritz, 1973) is a powerful tool to interpolate and extrapolate observations in space and time. The collocation is used to interpolate the synoptic observations p, T and the humidity horizontally at a given altitude, e. g. at sea level. These data are extrapolated using vertical profiles and are at each grid point combined to the wet refractivity.

4.4.2. Surface observations

The synoptic observations can also be used to constrain the refractivity at the lowest grid level. While the extrapolated vertical profiles used to initialise the refractivity field are only a first guess, the refractivity computed at the synoptic stations should be reliable. This kind of extra information is especially important as the GNSS observations at the lower levels are very inhomogeneously distributed.

The refractivity N_{j_0} computed from the observations of the synoptic station in voxel j_0 are used in the same way as the SWD observations except that only one voxel x_{j_0} is affected:

$$a_{ij} \cdot x_{j_0} = N_{j_0} \;\text{ with }\; \begin{cases} a_{ij} = 1\ ,\ j = j_0 \\ a_{ij} = 0\ ,\ j \neq j_0 \end{cases} \tag{20}$$

For each synoptic station one line is appended to the matrix $\boldsymbol{A}$.

The minimum refractivity at the upper grid level can be applied in the same way by replacing N_{j_0} with a certain

minimum refractivity $N_{\min}$ at the given altitude. If the grid spans the whole troposphere up to the tropopause ($\sim$ 10000 m) $N_{\min} = 0$ could be used.

4.4.3. Zenith wet delays

The horizontal humidity distribution as reconstructed by the GPS tomography should be very close to the IWV distribution. This can most easily be achieved by using the corresponding ZTDs as extra observations. The ZTDs computed together with the IWV are different from the STDs mapped back to the zenith. These ZTDs have been obtained by an least-squares solution of a large number of GPS observations and are of much higher quality than a single STD mapped to zenith (Dick et al., 2001). As in case of the STDs the ZTDs must be converted to ZWDs. Fortunately, this is already done by EPOS and the ZWDs can directly be used as an input to the tomography: Beginning with the station altitude, the vertical slant paths in each grid cell are computed which is simply the vertical grid spacing. Only in the lowest layer containing the station the vertical distance to the next layer must be computed:

$$\sum_j a_{ij} x_j = ZWD_i \tag{21}$$

4.4.4. Inter-voxel constraints

The reconstruction techniques of the ART family make all use of the same basic idea: The back projection $\langle \boldsymbol{A}^i, \boldsymbol{x}^k \rangle$ is compared to the observed quantity m_i either by computing the difference or the quotient of both terms and a correction proportional to this value is applied to the voxel. The relaxation parameter λ is used to tune the weight given to the current value with respect to the correction term. It is obvious that no corrections are applied to voxels which are not touched by any slant. This can lead to rather inhomogeneous and unrealistic distributions of the refractivity. Inter-voxel constraints can therefore be applied during the reconstruction to restrict the refractivity gradients between neighboured voxels. Such constraints can be appended to the basic linear system (Hobiger et al., 2008b) but in case of iterative techniques they can be more easily realised as filters which are applied to the refractivity field between certain loops of the iteration procedure.

A Gaussian filter was implemented to the PORTOS software which can be started several times: Prior to the reconstruction, e. g. to smooth the initial filed, during the iteration and after the reconstruction. The filtered refractivity N' in the voxel (i_0, j_0, k_0) is given by the weighted sum over the current refractivity N within the domain $i_0 \pm \Delta i$, $j_0 \pm \Delta j$, $k_0 \pm \Delta k$

$$N'_{i_0 j_0 k_0} = \frac{\sum_{i=i_0-\Delta i}^{i_0+\Delta i} \sum_{j=j_0-\Delta j}^{j_0+\Delta j} \sum_{k=k_0-\Delta k}^{k_0+\Delta k} \alpha_{i,j,k}\, N_{i,j,k}}{\sum_{i=i_0-\Delta i}^{i_0+\Delta i} \sum_{j=j_0-\Delta j}^{j_0+\Delta j} \sum_{k=k_0-\Delta k}^{k_0+\Delta k} \alpha_{i,j,k}} \tag{22}$$

where the Gaussian weighting factor is given by

$$\alpha_{ijk} = e^{-\frac{(x-x_0)^2}{2\sigma_1^2}} \cdot e^{-\frac{(y-y_0)^2}{2\sigma_2^2}} \cdot e^{-\frac{(z-z_0)^2}{2\sigma_3^2}} \tag{23}$$

To apply the filter the 3D refractivity field must be mapped to the state vector $\boldsymbol{x}$ and vice versa. The Gaussian filter leads to a spreading of the information through the grid. The information from voxels which have considerably be changed by the iterative application of the slant data as well as the initial information from voxels which have not been changed spread through the grid. By applying the filter several times the information provided by the slants will affect the whole grid and the weight of the initial state will decrease. The number of neighbored voxels $\Delta i, \Delta j$, and Δk used by the filter as well as the σ's giving the width of the Gaussian function can be chosen in order to obtain the best result. Usually one ore two neighbored voxels and $\sigma = 0.5 \ldots 2$ leads to good results.

4.4.5. Weighting matrix

Using observations from different sources leads usually to an imbalance in their impact on the result. A large number of slant observations is often accompanied by a rather limited number of other observations with different error characteristics. The surface observations and ZWD data are available with sampling periods of one hour or 15 minutes. Compared with about 8 slant delay observations per station and 2.5 minutes these data are outnumbered and have little impact on the result. It is therefore necessary to weight the different observations with respect to their quantity and their measuring error.

This can be achieved by introducing a weighting vector $\boldsymbol{\omega}$ which defines a weighting factor ω_i for each observation m_i. The weighted observations $\boldsymbol{m}'$ are then given by

$$\boldsymbol{m}' = \boldsymbol{\omega}\, \boldsymbol{I}\, \boldsymbol{m} = \boldsymbol{W}\, \boldsymbol{m} \tag{24}$$

where the weighting matrix $\boldsymbol{W}$ is a diagonal matrix with the weighting factors ω_i being the diagonal elements and $\boldsymbol{I}$ is the unit matrix. Equ. 2 becomes

$$\boldsymbol{W}(\boldsymbol{A}\, \boldsymbol{x}) = (\boldsymbol{W}\boldsymbol{A})\boldsymbol{x} = \boldsymbol{W}\, \boldsymbol{m} = \boldsymbol{m}' \tag{25}$$

which has the same structure as the original equation equ. 2 and can be solved in the same way.

In general, the weighting matrix should be replaced by the data covariance matrix which describes not only the error of each single observation but also the correlation of these errors. Unfortunately, this information is currently not available and only the weighting factors defined above are used.

5. Comparison of different ART algorithms

The different reconstruction algorithms from section 4.2 can be compared with respect to several criteria, e. g. quality of the resulting field, convergence behaviour, stability of the reconstruction process and computing time. A test data set was simulated for these comparisons which allowed the systematic variation of all relevant parameters. SWV data of a model atmosphere were simulated

using a raytracer and several hypothetical GNSS networks with dimensions of 700 km in EW direction and 1000 km in NS direction which would, e. g., cover Germany. The GNSS stations were placed on a regular latitude/longitude grid with inter station distances between 10 km and 40 km. The GPS satellite constellation was estimated using ephemeris data. For a 12 hours period the constellation was computed every 30 minutes leading to 25 different scenarios. In combination with the hypothetical networks described above more than 100 data sets providing between 4700 SWDs (438 stations with 40 km distances) and 74000 SWDs (7600 stations with 10 km distances) were investigated. Ideal noise-free SWD data were simulated. For testing the sensitivity on imperfect data white noise as provided by a random number generator was applied to these data. Only the SWV data but no extra information or constraints were used for the intercomparison study which gave the following results.

The most important parameter is the relaxation parameter λ used for the reconstruction. λ was kept constant within the reconstruction process and the same λ was applied to all observations (see section 4.2). As λ strongly influences both the quality of the reconstructed field and the convergence behaviour, the optimal λ-value should be estimated prior to further comparative studies. Tab. 1 shows the results obtained for different ART algorithms. The valid λ-range as well as the "best" λ-value vary considerably between different algorithms. The data given in tab. 1 depend on the actually used SWV data and the resolution of the tomographic grid and can only be regarded as approximate values. Nevertheless, the optimal λ-values seem to represent more the algorithm than the data as, e. g. $\lambda = 0.2$ has been found for several applications of the MART1 algorithm (Stolle et al., 2006). The results for δ and σ are in contrary closely related to the chosen data set and are comparable only if all other parameters are kept constant. The "best" λ has been chosen by comparing the resulting 3D fields with the original field and by observing Δ, δ and σ and their convergence behaviour. The station density within the GPS networks and GPS satellite constellation have also a strong impact on the quality of the reconstructed fields. These effects are not discussed here as they have little influence on the best choice for λ and on the convergence behaviour.

Fig. 1 shows the variation of δ and σ in case of a MART1 algorithm. The quality of the reconstruction and the convergence behaviour changes significantly as λ is varied. If λ is too small very little weight is given to the observations and the initial field with a large variance σ remains almost unchanged. If the weight of the observations is increased artifacts are evolving which lead to an increasing variance. The convergence of different ART algorithms with their "optimal" λ is shown in fig. 2. SIRT and the MART algorithms lead to comparable results but DART shows a more slowly convergence behaviour and leads to results of reduced quality. The small difference between MART1 and SIRT indicates that the order of the

ART	λ-range	best λ	δ	σ
ART1	0.05 – 1,0	~ 0.175	$-1.62 \cdot 10^{-1}$	2.588
MART1	0.05 – 2.0	~ 0.2	$-9.34 \cdot 10^{-2}$	2.71
MART2	100 – 10000	~ 5500	-1.27	5.73
MART3	0.5 – 500	~ 25.0	$-7.64 \cdot 10^{-2}$	2.493
DART	0.1 – 350	~ 30.0	$-3.31 \cdot 10^{-2}$	2.49
SIRT	0.01 – 0.5	~ 0.5	$-4.67 \cdot 10^{-1}$	2.97

Table 1: Relaxation parameter λ for different ART algorithms. λ's outside the given range lead to instable or oscillating results and cannot be used. The best λ shows in general fast convergence behaviour and leads to good results. The given δ and σ values have been obtained after 1000 iterations.

SWV data within $\boldsymbol{m}$ is of little importance. It can also be seen that optimal results are obtained after 100 – 200 iterations. Too many iterations result in numerical artefacts and can completely destroy former good results. A small variance σ does not always indicate a good reconstruction. In several cases the result became worse with an increasing number of iterations even though σ decreased.

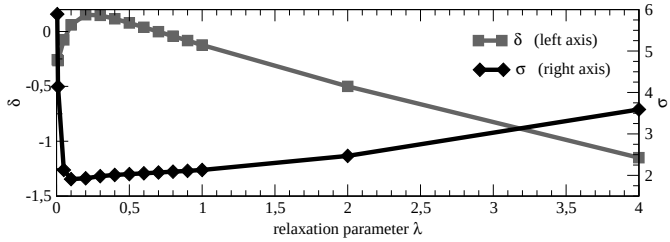

Figure 1: Quality of the reconstruction for different values of the relaxation parameter λ. A MART1 algorithm has been used to carry out several reconstructions which differ only in λ. Visual inspection of the reconstructed 3D humidity fields confirms that the best results are obtained in the vicinity of the σ-minimum. In case of MART1 $\lambda = 0.2$ has been chosen (tab. 1) because this field was closer to the original field than the result obtained for $\lambda = 0.1$. In all cases a constant number of 100 iterations was passed.

Summarising, all algorithms of the ART family tested here lead to comparable results and all of them might be used for further studies. It turned out that the optimisation of λ is more critical than the chosen algorithm. However, MART1 seems to be the fastest algorithm which gives the best results, while DART is less favourable.

5.1. Impact of SWV errors

Most inverse problems are very sensitive to small variations of the input data. It is therefore important to estimate the impact of experimental errors on the stability of the tomographic inversion. The error of SWV data has many contributions which arise not only from the experimental error but from several assumptions and approximations made by the GPS data analysis. The complex interrelation of the different GPS error sources was not simulated within this work. Instead, white noise was applied to the simulated SWV data.

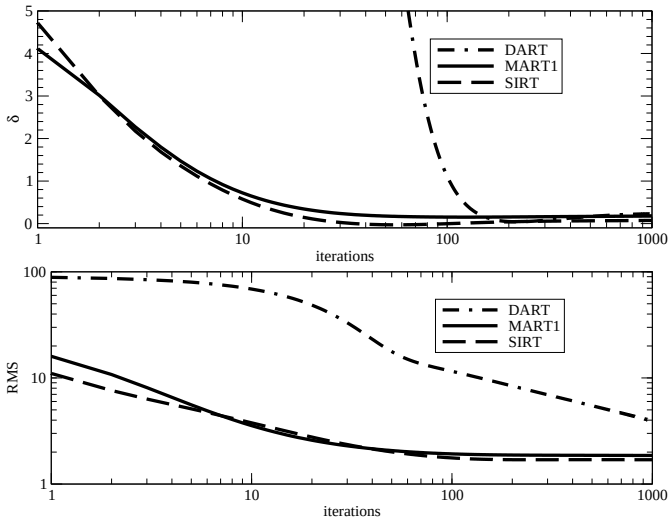

Figure 2: Convergence of three different algorithms: MART1, SIRT and DART. The convergence behaviour differs depending on the parameter used to quantify the reconstruction quality: Backprojection (δ, top) and variance (σ, bottom).

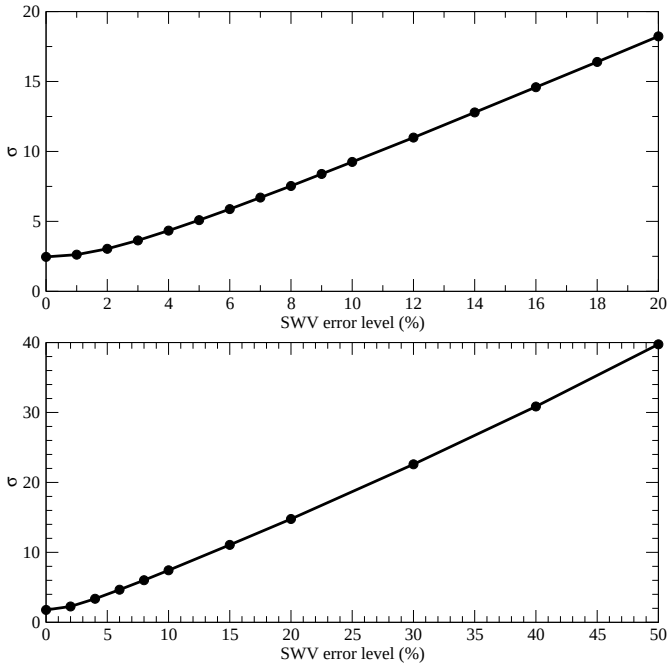

Figure 3: Quality of the reconstructed field as indicated by σ for different error levels. Horizontal resolutions of 40 km (top) and 30 km (bottom) have been used for the reconstruction of a 20 km network .

Several tomographic reconstructions were carried out using white noise up to 50 % of the "true" SWV value. The variation of the constellation, the humidity field and the resolution gave rather consistent results. The quality of the results decreases almost linearly with the error level as can be seen in fig. 3. Visual inspection of the reconstructed humidity fields shows granular structures which increase with an increasing error. An error of up to 20-30 % could be tolerated, higher values result in artefacts in all parts of the grid.

To get rid of the granular structures intervoxel smoothing was applied between consecutive iteration steps. The result is disappointing, the granularity reappears after some iterations. Nevertheless, the results of this analysis are very promising. The ART inversion algorithms are stable and converge to the same final state even if the input data have substantial errors.

6. GPS water vapour tomography for Germany

In summer 2009 the data of about 270 German GPS stations were processed operationally by the GFZ (see fig. 4). The data are supplied in near real-time by several network providers, e. g, the German satellite positioning service (SAPOS), the Federal Agency for Cartography and geodesy (BKG) or the GFZ itself. Some stations from worldwide or European networks are also available, e. g. from the International GNSS Service (IGS) or the IAG Reference Frame Sub-Commission for Europe (EUREF). Slant delay data of all stations are processed with elevation cutoff ange of 7°, a temporal resolution of 2.5 minutes and are provided in hourly batches. A description of the spatiotemporal distribution of these data can, e. g., be found in Bender et al. (2009). This section can only give an example of a tomographic reconstruction to illustrate the capabilities of the PORTOS software. A detailed study which quantifies the impact of the different options provided by PORTOS is subject to further work.

A special weather situation from the Convective and Orographically-induced Precipitation Study (COPS) (Wulfmeyer et al., 2008) in summer 2007 was reconstructed to demonstrate the potential of the GNSS tomography. In August 2007 about 150 German stations (fig. 4, blue stations) were processed by the GFZ and provided between 20000 and 35000 STD data per hour. These data were used together with hourly synoptic observations and the COSMO-DE numerical weather model analyses to reconstruct the tropospheric water vapour distribution. A spatial grid consisting of $21 \times 26 \times 31 = 16926$ cells covering a region of $700 \times 930 \times 10$ km with a horizontal resolution of about 35 km and a vertical resolution of 330 m was used to discretise the refractivity field. The grid was initialised by interpolating the corresponding COSMO-DE analysis fields onto the grid nodes. Combining the 16926 grid cells with the observations leads to a kernel matrix with more than $4.5 \cdot 10^8$ elements. Inverting equ. 2 with such a large sparse matrix by means of least-squares approaches or a (truncated) singular value decomposition is a computationally demanding task even if no matrix inversion is required. The MART1 algorithm provides results within less than 10 minutes on an ordinary PC (Core2Duo Intel processor, 2.3 GHz, 2 GB RAM) using only a single core. $\lambda = 0.2$ as taken from tab. 1 and a fixed number of 150 iterations

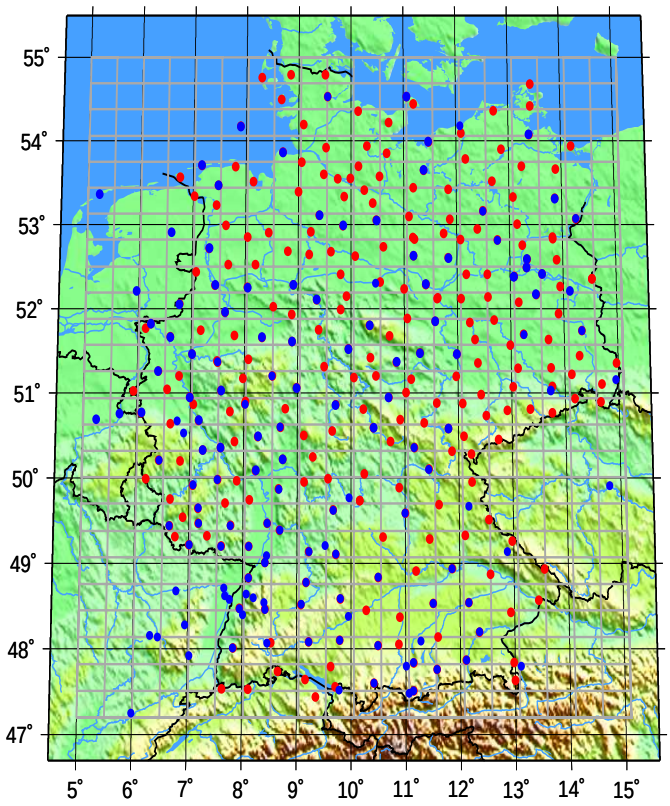

Figure 4: GPS stations in Germany and nearby countries providing STD data for the GNSS water vapour tomography. The blue stations were already available during COPS in summer 2007, the red stations were added in 2009. The grid used for the tomographic reconstruction (fig. 6) is shown in grey.

was chosen to obtain the results presented below. To get a more homogeneous distribution the Gaussian filter was applied three times between different reconstruction steps but not after the last step, i. e. the reconstruction was repeated four times.

The backprojections obtained with the initial field and the finally reconstructed field give a first impression of the reconstruction quality. The ART algorithms start with an initial field which is iteratively improved by computing the back projection $\boldsymbol{A}^i\boldsymbol{x}^k$ and applying small corrections weighted by this quantity. The quality of the initial filed ($k = 0$) and the result of the reconstruction ($k = N_{\mathrm{iter}} = 150$) as well can therefore be estimated by comparing the backprojection with the observations, i. e. by computing the distribution of $\Delta_i = m_0^i - \boldsymbol{A}^i\boldsymbol{x}^k$ for all observations i. Both Δ-distributions are shown in fig. 5. A broad distribution with two peaks at $\Delta = -1.4$ cm and $\Delta = 1.7$ cm was found for the initial field as the correlation with the observations is rather poor. After running MART1 a narrow distribution centred at $\Delta \approx 0$ was obtained. To receive such a distribution the refractivity along the majority of slant paths must have been improved considerably. Unfortunately, this is not a sufficient condition for a good reconstruction as there is in general an infinitely large set of solutions which can lead to the same distribution Δ. However, if the initial field was close enough to the real one a narrow distribution around $\Delta = 0$ indicates a reasonable result. Poor results show very often asymmetric side lobes and are not centered near $\Delta = 0$.

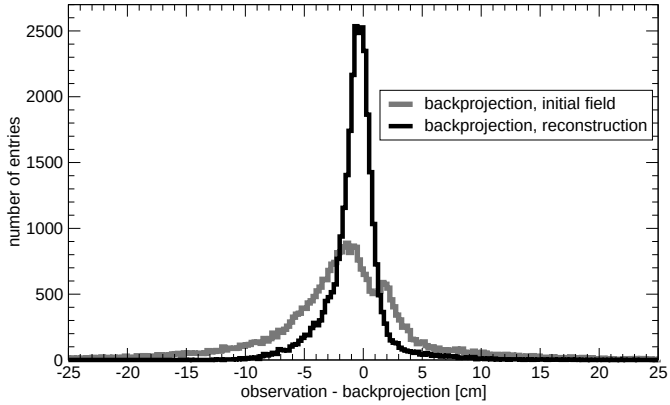

Figure 5: Distribution of $\Delta_i = m_0^i - \boldsymbol{A}^i\boldsymbol{x}^k$ for all 31222 slant wet delays used by the reconstruction. $\boldsymbol{A}^i\boldsymbol{x}^k$ is the backprojection of the ith slant path onto the observed wet delay m_0^i. Δ_i is the difference between the observed and the reconstructed value and gives an impression of the reconstruction quality. The number of entries in each 0.25 cm histogram bin is given.

The reconstruction of the wet refractivity at 6 August, 2007, 22:00 UTC is shown in fig. 6. A convergence line followed by a cold front was moving eastwards over Germany. The IWV field (fig. 6, right) shows distinct meridional isolines. The tomographic reconstruction used 31222 slant delays from one hour (22:00 - 23:00 UTC) of GPS observations but no other data. A COSMO-DE analysis from 18:00 UTC was used to initialise the field which should lead to comparable results as a 4 hours forecast. A horizontal layer from the reconstructed field of the wet refractivity N_{w} is shown in fig. 6 (centre). The layer at an altitude of 676 m ASL can be compared with the 22:00 UTC COSMO-DE analysis (fig. 6, left) interpolated to the same grid level and the IWV distribution (fig. 6, right) as obtained from the GPS processing. The horizontal water vapour distribution in the lower part of the troposphere should be close to the IWV distribution but it must be considered that the wet refractivity has contributions from the temperature field (equ. 3) and the N_{w} distribution can therefore be different from the IWV distribution. However, as the temperature gradients are usually correlated with the humidity gradients the shape of both distributions are in most cases very much alike.

In the current case, the 22:00 UTC COSMO-DE field is quite different from the observed IWV distribution. Most of the humidity is concentrated in the southwestern part of Germany with strong gradients to a very dry region in the southeastern part. The northern part of Germany is considerably too dry. Regarding the COSMO-DE analysis the convergence line was coming from southwest rather than from west. The tomographic reconstruction shows a

humidity field which is in close resemblance to the IWV distribution but shows still some pattern of the initial COSMO-DE distribution and some artefacts of the tomography. The isolated structures in the reconstructed field are mainly due to the limited number of stations. Station distances between $\sim$ 30 km and $\sim$ 60 km lead to a rather small number of slant paths per voxel and therefore to local pattern around some stations. The shape of the convergence line is rather close to the shape of the IWV distribution only the northern part is too dry. This is due to the sparse GPS network in northern Germany which doesn't provide enough slant delays to override the initially wrong field. The unevenly distributed GPS stations lead also to a clustered structure of the reconstructed humidity field which becomes more pronounced if the number of iterations is increased. However, the tomography results in a humidity field which is altogether much closer to the reality than the analysis of the numerical weather model.

A vertical cut through the humidity fields at $\lambda = 7.52°$ E is shown in the lower part of fig. 6. The upper profile was taken from the COSMO-DE field, the lower one from the tomographic reconstruction. Again it can be seen that the dry region in the northern part became much more humid after the tomographic reconstruction. It seems that humid air masses are in some regions located at higher altitudes than in the model field but layers above $\sim$ 3500 m are at the same time much dryer. The vertical field shows a clustered structure just as the horizontal view. Further validation studies, e. g. using radiosonde profiles, are necessary to verify if these structures are real atmospheric pattern or artefacts caused by the inhomogeneous distribution of GPS stations.

Three vertical profiles were compared with the COSMO-DE profiles (fig. 7). The humidity field around Potsdam (fig. 7, left) in the northeastern part of Germany was not considerably changed by the tomography and the profiles are rather similar, except the somewhat smoother profile of the tomography. Two radiosonde (RS) profiles from Lindenberg (about 100 km east of Potsdam) support these results, whereby the 12:00 UTC RS profile is close to the tomographic reconstruction while the 0:00 UTC profile is more similar to the COSMO-DE profile. The COSMO-DE field around Osnabrück (fig. 7, centre) in the southwestern part of Germany was essentially too wet and the tomography leads to a much dryer boundary layer but to an increasing humidity between $\sim$ 400 m and $\sim$ 2000 m. Above $\sim$ 2000 m the tomographic reconstruction is again much dryer. However, both profiles are too wet as compared to the RS profile from Essen (150 km NE from Osnabrück). The third region around Landshut (fig. 7, right) in the southeastern part was considerably too dry. As in the previous case the boundary layer of the reconstructed field is much dryer than the COSMO-DE field but between $\sim$ 200 m and $\sim$ 1000 m there is much more humidity in the reconstructed field which is to some degree compensated by the dryer region between $\sim$ 1000 m and $\sim$ 2000 m. The RS profile from München (75 km SW of Landshut) shows considerably more humidity than the COSMO-DE field but at higher altitudes than the tomographic reconstruction.

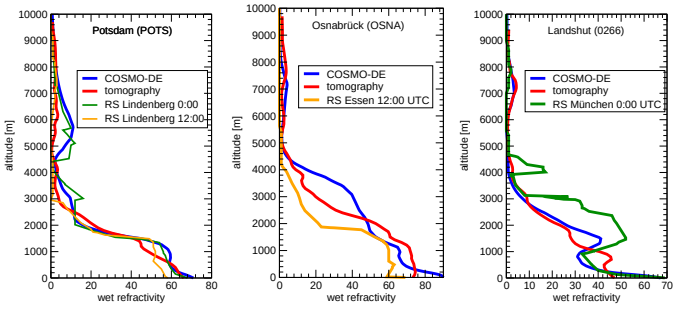

Figure 7: Vertical N_w profiles at three representative stations: Potsdam (left), 13.07° E, 52.38° N, Osnabrück (centre), 8.05° E, 52.25° N and Landshut (right), 12.16° E, 48.54° N. The tomographically reconstructed field (red) and the COSMO-DE fields (blue) were interpolated to the station coordinates. Some profiles of nearby radio soundings are also shown. In all cases the altitude above ground is given.

Altogether, the tomography leads to reasonable vertical profiles as compared with a numerical weather model and some radiosonde profiles. However, the validation of tomographically reconstructed humidity fields must be carried out very carefully. The spatial coverage of the atmosphere by slant paths can change very fast as the GNSS satellite constellation varies. The GPS contribution to the tomography is therefore highly variable in space and time and an uniform quality of the reconstructed fields can therefore not be expected. The GNSS tomography can, like any other observation system, not provide any information if no observations from the given region are available. Validating certain regions covered by the tomography grid should therefore always be accompanied by an estimation of the information provided by the GNSS observations within that region.

7. Conclusion

A GNSS water vapour tomography system was developed by the GFZ and the LIM to reconstruct spatially resolved humidity fields in the troposphere. The tomography is part of the GNSS atmosphere sounding system of the GFZ and utilises the spatial information provided by the STD data to extend the established horizontal IWV distributions along the vertical axis. Designed as a part of the near real-time GNSS processing system the tomography must be completed in less than 15 minutes in order to supply all atmospheric products within 30 minutes. These requirements are demanding if the slant delay data from Germany, i. e. more than 50000 STDs per hour, must be processed with a temporal resolution between 15 and 60 minutes, a horizontal resolution of 30–40 km and a vertical resolution better than 500 m. As more GNSS satellites become available in near future and the networks

are extended to cover larger regions with an improved station density the tomography must have the potential to meet even considerably higher conditions. Several members from the family of ART algorithms were therefore investigated with respect to their potential for the tropospheric GNSS tomography. These iterative techniques were successfully used in the GNSS ionosphere tomography and can easily be parallelised to run on multiple processor cores or on computer clusters.

It was found that all tested algorithms of the ART family are suitable for the GNSS tomography but show a different convergence behaviour and lead to somewhat varying results. The important relaxation parameter λ was systematically modified to estimate the range of feasible λ's and to identify an optimal λ parameter. Very different λ ranges were found for different algorithms. A special version of the MART algorithm (MART1) seems to give the best combination of reconstruction speed and quality.

However, finding the best algorithm is closely related to the evaluation of the reconstruction quality. This is a rather difficult task as the spatial coverage of the atmosphere by slant paths depends on the satellite constellation and changes very fast. It can therefore not be expected that all parts of the region are reconstructed equally well at all times. The capability of the reconstruction algorithms must therefore be separated from the quality of the input data. Another aspect of this problem is the convergence behaviour of the iterative reconstruction algorithms and the definition of stop criteria. To stop the iteration in an optimal state the reconstruction quality must be estimated continuously. The criteria used in this work vary in a wide range and it is difficult to define thresholds which can be used to stop the iterations. Further investigations on this topic are required.

It was also found that white noise added to the simulated SWD data does not disturb the ART algorithms too much as long as the noise does not exceed 50 % of the SWV. The quality of the reconstruction decreases with an increasing noise level but the reconstruction process itself remains stable, i. e. the degradation is a continuous process. This is a very nice feature of an inverse procedure which very often become instable and lead at a certain point to meaningless results.

Several options to use additional observations and to introduce constraints were described. Synoptic observations can be extrapolated onto the spatial grid to initialise the tomography and can be used as point observations which are processed together with the slant delays. Fitted with an adequate weight the surface layer can be constrained to these observations. Averaged ZWDs from the IWV analysis can be used to constrain the total amount of water vapour inside the reconstructed region and the upper grid level can be set to some small value to preselect solutions with approximately exponential vertical profiles. Inter-voxel constraints implemented as a Gaussian filter reduce the gradients within the reconstructed humidity field and lead to a spreading of the information through the grid. This is important to adjust regions which are not covered by any observations to the overall distribution. The impact of these additional observations and the constraints will be investigated in future work. As the primary intention of this work is the description of the GNSS tomography system and its reconstruction algorithms only an example of a reconstructed field is given to demonstrate the applicability of the system.

Acknowledgments

This work was supported by the Deutsche Forschungsgemeinschaft (DFG) under the research grants RO 2330/5-1, TE 51/26-1. The authors would like to thank W. Wergen, Ch. Schraff and Ch. Koziar from the German Meteorological Service (DWD) for providing COSMO analyses and weather observations.

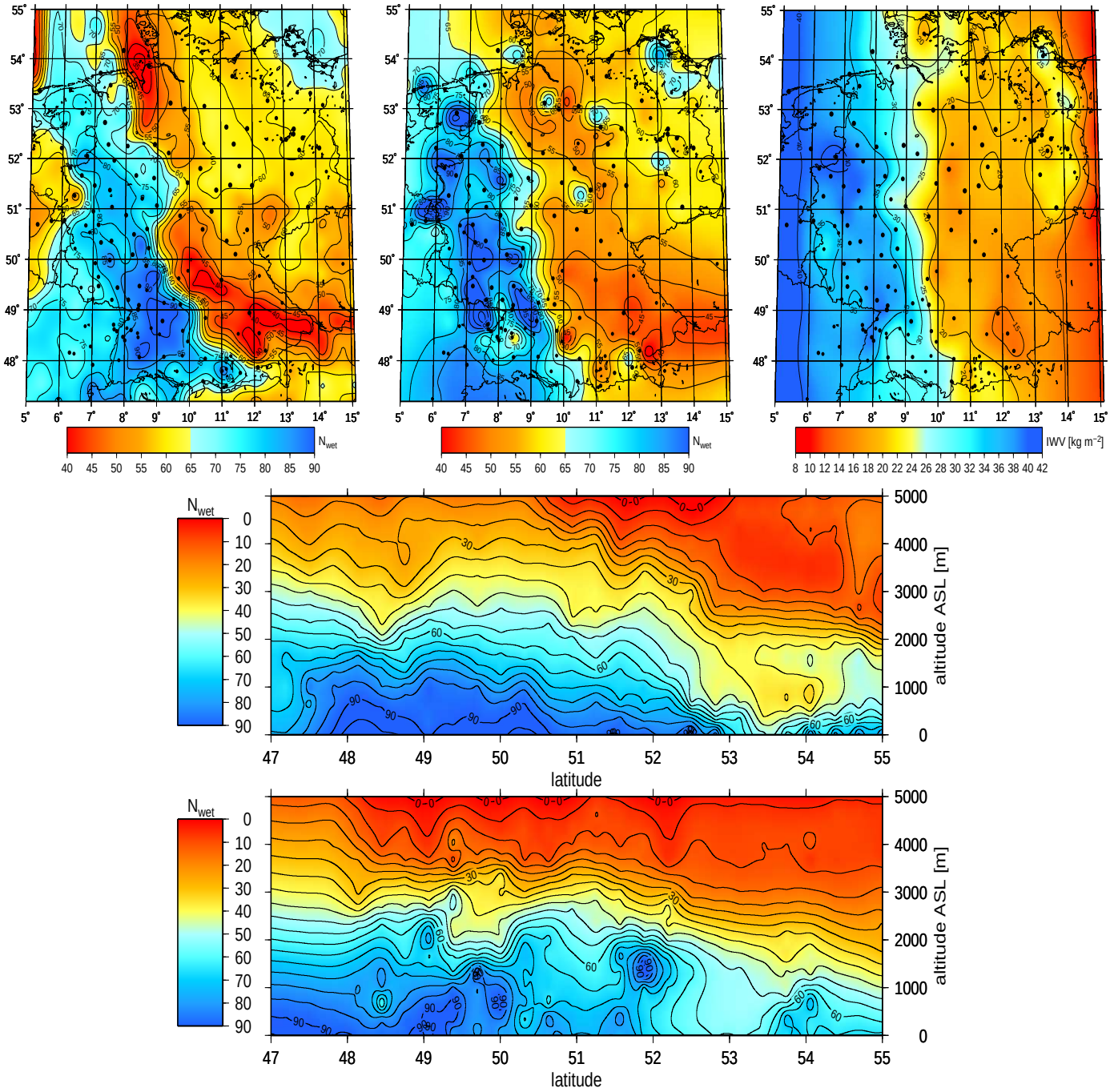

Figure 6: Tomographic reconstruction of the wet refractivity at 22:00 UTC, 6.8.2007 (top, centre), the corresponding COSMO-DE field (left) and the observed IWV distribution (right). In case of the reconstruction and the COSMO field a layer at 676 m ASL is shown. The IWV always represents the total amount of humidity in the atmosphere. The stations providing slant delays for the tomography are given by black dots. In the lower part two vertical cuts through the N_w fields at $\lambda = 7.52°$ E are shown. The upper slice represents the COSMO-DE field the lower one the tomographic reconstruction.

References

Alber, C., Ware, R., Rocken, C., et al., Obtaining single path delays from GPS double differences, Geophys. Res. Lett. 27 (7), 2661–2664, 2000.

Bender, M., Dick, G., Wickert, J., et al., Estimates of the information provided by GPS slant data observed in Germany regarding tomographic applications, J. Geophys. Res. 114, D06303, 2009, doi:10.1029/2008JD011008.

Bender, M., Raabe, A., Preconditions to ground-based GPS water vapour tomography, Ann. Geophysicae 25 (8), 1727–1734, 2007.

Bengtsson, L., Hodges, K. I., On the impact of humidity observations in numerical weather prediction, Tellus 57 (5), 701–708, 2005, doi:10.1111/j.1600-0870.2005.00142.x.

Bevis, M., Businger, S., Chiswell, S., et al., GPS meteorology: Mapping zenith wet delays onto precipitable water, J. Appl. Meteor. 33 (3), 379–386, 1994, doi:10.1175/1520-0450.

Bevis, M., Businger, S., Herring, T. A., et al., GPS meteorology: Remote sensing of atmospheric water vapor using the global positioning system, J. Geophys. Res. 97 (D14), 15787–15801, 1992.

Boehm, J., Kouba, J., Schuh, H., Forecast vienna mapping functions 1 for real-time analysis of space geodetic observations, J. Geodesy 83 (5), 397–401, 2008, doi:10.1007/s00190-008-0216-y.

Boehm, J., Niell, A., Tregoning, P., et al., Global Mapping Function (GMF): A new empirical mapping function based on numerical weather model data, Geophys. Res. Lett. 33, L07304, 2006, doi:10.1029/2005GL025546.

Braun, J., Rocken, C., Ware, R., Validation of line-of-sight water vapor measurements with GPS, Radio Sci. 36 (3), 459–472, 2001.

Censor, Y., Finite series-expansion reconstruction methods, Proceedings of the IEEE 71, 409–419, 1983.

Champollion, C., Masson, F., Bouin, M.-N., et al., GPS water vapour tomography: preliminary results from the ESCOMPTE field experiment, Atmos. Res. 74 (1-4), 253–274, 2005.

Chen, G., Herring, T. A., Effects of atmospheric azimuthal asymmetry on the analysis of space geodetic data, J. Geophys. Res. 102 (B9), 20489–20502, 1997.

Davis, J., Herring, T., Shapiro, I., et al., Geodesy by radio interferometry: Effects of atmospheric modeling errors on estimates of baseline length, Radio Sci. 20 (6), 1593–1607, 1985.

de Brito Mendes, V., Modeling the neutral–atmosphere propagation delay in radiometric space techniques, Ph.D. thesis, University of New Brunswick, Fredericton, New Brunswick, Canada, Department of Geodesy and Geomatics Engineering, 1999, technical Report No. 199.

de Haan, S., van der Marel, H., Observing three dimensional water vapour using a surface network of GPS receivers 8 (5), 17193 – 17235, 2008.

Deng, Z., Bender, M., Dick, G., et al., Retrieving tropospheric delays from GPS networks densified with single frequency receivers, Geophys. Res. Lett. 36, L19802, 2009, doi:10.1029/2009GL040018.

Dick, G., Gendt, G., Reigber, C., First experience with near real-time water vapor estimation in a German GPS network, J. Atmos. Sol. Terr. Phys. 63 (12), 1295–1304, 2001.

Flores, A., de Arellano, J.-G., Gradinarsky, L. P., et al., Tomography of the lower troposphere using a small dense network of GPS receivers, IEEE Trans. on Geosci. and Rem. Sens. 39 (2), 439 – 447, 2001.

Foelsche, U., Kirchengast, G., Tropospheric water vapor imaging by combination of ground-based and spaceborn GNSS sounding data, J. Geophys. Res. 106 (D21), 27221–27231, 2001.

Ge, M. R., Calais, E., Haase, J., Sensitivity of zenith total delay accuracy to GPS orbit errors and implications for near-real-time GPS meteorology, J. Geophys. Res. 107 (D16), 2002, doi:10.1029/2001JD001095.

Gendt, G., Dick, G., Reigber, C., et al., Near real time GPS water vapor monitoring for numerical weather prediction in Germany, J. Meteor. Soc. Japan 82 (1B), 361–370, 2004.

Gendt, G., Dick, G., Rius, A., et al., Comparison of software and techniques for water vapor estimation using German near real-time GPS data, Phys. Chem. Earth (A) 26 (6-8), 417–420, 2001a.

Gendt, G., Reigber, C., Dick, G., Near real-time water vapor estimation in a German GPS network - first results from the ground program of the HGF GASP project, Phys. Chem. Earth (A) 26 (6-8), 413–416, 2001b.

Gilbert, P. F. C., Iterative methods for three-dimensional reconstruction of an object from its projections, J. Theo. Biol. 36 (1), 105–117, 1972, doi:10.1016/0022-5193(72)90180-4.

Gordon, R., Bender, R., Herman, G. T., Algebraic reconstruction technique (ART) for three-dimensional electron microscopy and X-ray photography, J. Theo. Biol. 29 (3), 471–481, 1970, doi:10.1016/0022-5193(70)90109-8.

Gradinarsky, L. P., Jarlemark, P., Ground-based GPS tomography of water vapor: Analysis of simulated and real data, J. Meteor. Soc. Japan 82 (1B), 551–560, 2004.

Hirter, H., Mehrdimensionale Interpolation von meteorologischen Feldern zur Berechnung der Brechungsbedingungen in der Geodäsie, Technical Report Mitteilungen Nr. 64, Institut für Geodäsie und Photogrammetrie, 1996.

Hobiger, T., Ichikawa, R., Takasu, T., et al., Ray-traced troposphere slant delays for precise point positioning, Earth Planets Space 60 (5), e1–e4, 2008a.

Hobiger, T., Kondo, T., Koyama, Y., Constrained simultaneous algebraic reconstruction technique (C-SART) - a new and simple algorithm applied to ionospheric tomography, Earth Planets Space 60 (7), 727–735, 2008b.

Hofmann-Wellenhof, B., Lichtenegger, H., Collins, J., Global Positioning System – Theory and Practise, Springer, Wien, New York, 2nd edition, 1993.

Jin, S., Luo, O. F., Variability and climatology of PWV from global 13-year GPS observations, IEEE Trans. on Geosci. and Rem. Sens. 47 (7), 1918–1924, 2009, doi:10.1109/TGRS.2008.2010401.

Jin, S., Luo, O. F., Gleason, S., Characterization of diurnal cycles in ZTD from a decade of global GPS observations, J. Geodesy 83 (6), 537–545, 2009, doi:10.1007/s00190-008-0264-3.

Jin, S., Luo, O. F., Park, P., GPS observations of the ionospheric F2-layer behavior during the 20th November 2003 geomagnetic storm over South Korea, J. Geodesy 82 (12), 883–892, 2008, doi:10.1007/s00190-008-0217-x.

Jin, S., Park, J.-U., GPS ionospheric tomography: A comparison with the IRI-2001 model over South Korea, Earth Planets Space 59 (4), 287–292, 2007.

Jin, S., Park, J.-U., Cho, J.-H., et al., Seasonal variability of GPS-derived zenith tropospheric delay (1994 - 2006) and climate implications, J. Geophys. Res. 112, D09110, 2007, doi:10.1029/2006JD007772.

Kak, A. C., Slaney, M., Principles of Computerized Tomographic Imaging, The lnstitute of Electrical and Electronics Engineers, New York, 1999.

Kunitsyn, V. E., Tereshchenko, E. D., Ionospheric Tomography, Springer, Berlin, Heidelberg, 2003.

Lu, W., Yin, F. F., Adaptive algebraic reconstruction technique, Med. Phy. 31 (12), 3222–3230., 2004.

Moritz, H., Least–squares collocation, volume 75 of Deutsche Geodätische Kommission bei der Bayerischen Akademie der Wissenschaften : Reihe A, Theoretische Geodäsie, München, 1973.

Niell, A. E., Global mapping functions for the atmosphere delay at radio wavelengths, J. Geophys. Res. 101 (B2), 3227–3246, 1996, doi:10.1029/95JB03048.

Poli, P., Moll, P., Rabier, F., et al., Forecast impact studies of zenith total delay data from European near real-time GPS stations in Météo France 4DVAR, J. Geophys. Res. 112, D06114, 2007, doi:10.1029/2006JD007430.

Raymund, T. D., Comparisons of several ionospheric tomography algorithms, Ann. Geophysicae 13 (12), 1254–1262, 1995.

Rocken, C., Johnson, J. M., Braun, J. J., et al., Improving GPS surveying with modeled ionospheric corrections, Geophys. Res. Lett. 27 (23), 3821 – 3824, 2000.

Rocken, C., Sokolovskiy, S., Johnson, J. M., et al., Improved mapping of tropospheric delays, J. Atmos. Oceanic Technol. 18 (7), 1205–1213, 2001.

Rocken, C., Ware, R. H., Hove, T. V., et al., Sensing atmospheric

water vapor with the global positioning system, Geophys. Res. Lett. 20 (23), 2631–2634, 1993.

Saastamoinen, J., Atmospheric correction for the troposphere and stratosphere in radio ranging of satellites, in The Use of Artificial Satellites for Geodesy, (editors) Henriksen, S. W., Mancini, A., Chovitz, B. H., volume 15 of Geophysical Monograph Series, pp. 247–251, American Geophysical Union (AGU), 1972.

Saastamoinen, J., Contributions to the theory of atmospheric refraction Part II. Refraction corrections in satellite geodesy, Bull. Geod. 107, 13–34, 1973.

Seko, H., Shimada, S., Nakamura, H., et al., Three-dimensional distribution of water vapor estimated from tropospheric delay of GPS data in a mesoscale precipitation system of the baiu front, Earth Planets Space 52 (11), 927–933, 2000.

Smith, E. K., Weintraub, S., The constants in the equation for atmospheric refractive index at radio frequencies, Proceedings of the I.R.E. 41, 1035–1037, 1953.

Stolle, C., Three-dimensional imaging of ionospheric electron density fields using GPS observations at the ground and onboard the CHAMP satellite, Ph.D. thesis, Universität Leipzig, Institut für Meteorologie, 2004.

Stolle, C., Schlüter, S., Heise, M., et al., A GPS based three-dimensional ionospheric imaging tool: Process and assessment, Adv. Space Res. 38 (11), 2313–2317, 2006, doi: 10.1016/j.asr.2006.05.016.

Subbarao, P. M. V., Munshi, P., Muralidhar, K., Performance of iterative tomographic algorithms applied to non-destructive evaluation with limited data, NDT & E International 30 (6), 359–370, 1997, doi:10.1016/S0963-8695(97)00005-4.

Thayer, G. T., An improved equation for the radio refractive index of air, Radio Sci. 9 (10), 803–807, 1974.

Tomassini, M., Gendt, G., Dick, G., et al., Monitoring of integrated water vapour from ground-based GPS observations and their assimilation in a limited-area NWP model, Phys. Chem. Earth 27 (4-5), 341–346, 2002.

Troller, M., Geiger, A., Brockmann, E., et al., Determination of the spatial and temporal variation of tropospheric water vapour using GPS networks, Geophys. J. Int. 167 (2), 509–520, 2006a.

Troller, M., Geiger, A., Brockmann, E., et al., Tomographic determination of the spatial distribution of water vapor using GPS observations, Adv. Space Res. 37 (12), 2211–2217, 2006b, doi: 10.1016/j.asr.2005.07.002.

Ware, R., Alber, C., Rocken, C., et al., Sensing integrated water vapor along GPS ray paths, Geophys. Res. Lett. 24 (4), 417–420, 1997, doi:10.1029/97GL00080.

Wulfmeyer, V., Behrendt, A., Bauer, H.-S., et al., The convective and orographically-induced precipitation study: A research and development project of the World Weather Research program for improving quantitative precipitation forecasting in low-mountain regions, Bull. Amer. Meteor. Soc. 89 (10), 1477–1486, 2008, doi: 10.1175/2008BAMS2367.1.

Zhai, Y., Cummer, S. A., A flexible and robust direct reconstruction method for magnetospheric radio tomography, Radio Sci. 40 (3), RS3004, 2005.

Zumberge, J. F., Heflin, M. B., Jefferson, D. C., et al., Precise point positioning for the efficient and robust analysis of GPS data from large networks, J. Geophys. Res. 102 (B3), 5005–5018, 1997, doi: 10.1029/96JB03860.

Zus, F., Grzeschik, M., Bauer, H.-S., et al., Development and optimization of the IPM MM5 GPS slant path 4DVAR system, Meteorol. Z. 17 (6), 867–885, 2008, doi:10.1127/0941-2948/2008/0339.