GPS Solutions (2022) 26:133
https://doi.org/10.1007/s10291-022-01315-2

ORIGINAL ARTICLE

Applying principal components to analyze the distribution of model biases in GNSS tropospheric tomography for a case study in Northwestern Iran

Masoud Mashhadi Hossainali[1] · **Hanieh Tabatabaei**[1]

Received: 16 February 2021 / Accepted: 27 July 2022

Abstract
Although GNSS tropospheric tomography is a powerful tool in meteorology, available validation data limit its accuracy and precision analysis. Moreover, it is customary to accept the validation results as a measure of the model performance. This study shows that this is only possible when the sensitivity of the model elements to the input perturbations is the same. We propose the principal component analysis for studying the sensitivity of a tomography model for this purpose. Our model includes 17 GNSS stations in Northwestern Iran. To analyze the contribution of the applied constraints in the sensitivity results, we use the 3D Gaussian, horizontal, numerical weather prediction model and virtual reference stations (VRS) in our analysis. The results show that some parts of our model are more sensitive to perturbations of input parameters, and therefore, they are more prone to regularization bias. This depends not only on time but also on the applied constraints for computing a unique solution. Results show that the response of our model to input perturbations is considerably different when we use the VRS concept for constraining the model. Using the proposed method and the traditional ways of validating a tomography model, one can develop a lower bound limit for the bias in the sensitive parts of the model and an upper bound limit for the bias in the other parts.

Keywords GNSS troposphere tomography · Principal component analysis · Sensitivity analysis · Regularization bias

Introduction

Water vapor is one of the most significant and varying components of the troposphere. It plays a major role in many areas, such as the accurate estimation of tropospheric delays of GNSS signals (Shafei and Mashhadi-Hossainali 2020), weather forecasting (Zhao et al. 2020; Chen et al. 2017), and hazard mitigation (Bender and Raabe 2007). Water vapor has a short but complex lifespan (approximately ten days) caused by atmospheric dynamics, making its distribution heterogeneous in space and time (Guerova 2003). Because of this and the sparse set of available measurements, the determination of atmospheric water vapor is not an easy subject.

GNSS signals can be used to estimate water vapor through tomography. GNSS signals are delayed in the troposphere due to the refractivity parameter, which is usually divided into hydrostatic and wet components. If surface meteorological data is available, it is possible to estimate the hydrostatic component at a millimeter or more accuracy (Bevis et al. 1992). In contrast, the wet component depends on the amount of troposphere water vapor. GNSS tomography is a recent method for estimating this component (Braun et al. 1999), where the problem is expressed in terms of a Fredholm integral equation of the first kind. Therefore, this method results in an improperly posed problem in the sense that its solution is often extremely unstable and non-unique (Aster et al. 2019). The simultaneous system of integral equations is solved by dividing the troposphere into a finite set of elements known as voxels (Flores 1999). Tropospheric delays of GNSS signals are then used to reconstruct the wet refractivity in these voxels (Brenot et al. 2020). Wet refractivity is assumed to be constant within the voxels of such a

✉ Masoud Mashhadi Hossainali
hossainali@kntu.ac.ir

Hanieh Tabatabaei
h_tabatabai@email.kntu.ac.ir

[1] Faculty of Geodesy and Geomatics Engineering, K. N. Toosi University of Technology, No. 1346, ValiAsr Street, Mirdamad Cross, Tehran, Iran

Published online: 22 August 2022

model (basic assumption in GNSS tomography). However, wet refractivity can be distributed vertically using a power function if its values exhibit a smooth vertical profile that decrease exponentially with height (Douša and Eliaš 2014). Computational tomography demand appears to be reasonable on the occasions that an inversion of vertical profile occurs at some heights.

The imaging system of computerized tomography (CT) consists of ground-based receivers and GNSS satellites in space. Neither the GNSS receivers nor the constellation of GNSS satellites is designed for an optimum CT system. For example, some voxels remain empty due to the limited number of receivers and the elevation angle of satellites. Moreover, atmospheric parameters like wind speed challenge the basic assumption in GNSS tomography. Therefore, the dimension of the voxels should be changed such that the reconstructed refractivity parameters are truly representative of wet refractivity in the relevant parts of the atmosphere in the time span covered by the input data or the time response of the model. Reconstructed images are usually verified using radiosonde data. It is common to assume radiosondes provide a vertical profile of atmospheric parameters at the zenith of the launch point. Moreover, their spatial and temporal resolutions are poor compared to the spatial and temporal resolution of GNSS data. To conclude, the inadequacy of the imaging system, the instability and non-uniqueness of the sought solution, the need for the consistency of the voxel size with the dynamics of the atmosphere, the selection of an optimal time response for the model, the validity of the basic assumption in GNSS tomography and the insufficiency of the method normally used for validating tomography images are challenges in GNSS tomography. This study focuses on the validation part of GNSS-tomography.

Constraints on unknown parameters are normally used to compute a unique solution for the GNSS tropospheric tomography problem. Different methods have been proposed for this purpose: Hirahara (2000) added the constraints derived from radiosonde observations into the tomographic equations system. Liou et al. (2003) introduced surface meteorological measurements and assumed zero refractivity at a reasonable height to constrain the problem. Adavi and Mashhadi-Hossainali (2014) used the observations of virtual reference stations (VRS) to fix the rank deficiency of the problem. Shafei and Mashhadi-Hossainali (2020) applied reflected signals in an airborne reflectometry mission over Italy as constraints. Bender et al. (2011) used several vertical and horizontal constraints in a GPS water vapor tomography, including synoptic observations and a 3D Gaussian Filter. Rohm and Bosy (2011) used two types of constraint equations: First, they added horizontal inter voxel constraints, where each refractivity is a weighted mean of the refractivity of the surrounding voxels of the same layer. They also introduced additional parameters from the airflow analysis of the radiosonde data and the COAMPS weather prediction model.

The accuracy and precision of a tomographic model are normally analyzed using statistics such as the bias and standard deviation of the developed model. This analysis is usually done by comparing the refractivity profiles derived from the radiosonde data and the model. Radiosondes are typically launched a few times a day, and their vertical profiles are available for a limited time and only in the position of the radiosonde stations. Therefore, it is unclear whether the results of the accuracy and precision analysis in a tomographic model can be extended to other voxels in the desired region. In addition, to get a unique solution, the application of certain constraints is necessary. The new question is how the applied constraints contribute to the accuracy and precision of a tomographic model. We try to discuss these problems. To this end, we apply principal component analysis (PCA) as a sensitivity analysis method to a tomographic model in Iran.

Methodology

As a useful technique, GNSS tomography reconstructs the spatio-temporal variations of the wet refractivity in the troposphere. The Slant Wet Delays (SWDs) are the most important inputs in GNSS tomography. The fundamental equation in GNSS tomography is given by:

$$\mathrm{SWD} = 10^{-6} \int_S N_w \mathrm{d}s \tag{1}$$

where N_w is the wet component of refractivity, and S is the signal path between a satellite and a receiver.

The fundamental continuous integral in (1) is changed into a discrete form. The troposphere is substituted by voxels and refractivity is assumed to be constant in every voxel of this model. This changes (1) to:

$$\mathrm{SWD}_i = 10^{-6} \sum N_{wj} \Delta S_j \tag{2}$$

where index j represents the voxels located in the path of the ith signal and ΔS_j is the associated signal length in the jth voxel.

In GNSS data processing, the values of SWDs are not determined directly. Usually, they are calculated as a function of Zenith Wet Delay (ZWD), mapping functions, and tropospheric horizontal gradients as (Hanna et al. 2019):

$$\mathrm{SWD} = \mathrm{ZWD}.\mathrm{mf}_w(\varepsilon) + \mathrm{mf}_g(\varepsilon).[N.\cos(\alpha) + E.\sin(\alpha)] \tag{3}$$

where ε and α are the elevation and azimuth angles, N and E are the tropospheric horizontal gradients in the north–south and east–west, respectively. Also, mf_w is the wet mapping function for ZWDs that is calculated by global mapping

function (GMF) in this research (Böhm et al. 2006), and mf_g is the Chen–Herring mapping function for horizontal gradients (Chen and Herring 1997). ZWD values and tropospheric gradients are calculated by processing the GNSS observations.

In matrix notation, Eq. (2) results in the simultaneous system of observation equations as given in:

$$\mathbf{y} = \mathbf{B}\mathbf{x} \tag{4}$$

where $\boldsymbol{y}$ is the observations vector (SWDs), $\mathbf{x}$ consists of the unknown parameters (N_w), and $\boldsymbol{B}$ is the design matrix with $m \times n$ dimensions. m depends on the number of GNSS stations, GNSS satellites, and the time resolution of the model. Also, n is the number of unknown parameters or voxels. The elements of matrix $\boldsymbol{B}$ represent the length of signals within the corresponding voxels.

GNSS signals do not reach the lower layers of the model due to the low elevation of the satellites and the limited number of receivers. Therefore, some voxels are signal-free. This makes the model design matrix rank deficient. One of the common ways to solve this problem is to add constraints to (4) as follows:

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{y}_{\mathbf{constraints}} \end{bmatrix} = \begin{bmatrix} \mathbf{B} \\ \mathbf{B}_{\mathbf{constraints}} \end{bmatrix} \mathbf{x} \;\rightarrow\; \mathbf{d} = \mathbf{A}\mathbf{x} \tag{5}$$

Equation (5) is the final form of the observation equations for reconstructing a tomographic image at an arbitrary epoch.

We solve the problem using the gradient descent with constant step size, also known as the Landweber method (Nikazad 2007). The Landweber solution is:

$$\mathbf{m}^{(K+1)} = \mathbf{m}^{k} - \omega_k \mathbf{A}^T \left(\mathbf{A}\mathbf{m}^{(k)} - \mathbf{d}\right) \tag{6}$$

where ω_k (relaxation parameter) ensures the convergence of the solution. It is selected such that $0 < \omega_k < 2/\sigma^2_{\max}$. Here, $\sigma_{\max}$ is the largest singular value of the design matrix $\mathbf{A}$ (Aster et al. 2005). Various methods are available for computing the value of ω_k. Some of the most common ones are Ψ_1 and Ψ_2-based relaxation strategies, as well as the modified Ψ_1 and Ψ_2 methods. In these methods, the relaxation parameter is determined in each step to control the propagated noise component of the error (Elfving et al. 2010). We prefer the modified Ψ_2 method for its high convergence speed.

Radiosonde data is usually used to select the optimum number of iterations. For this purpose, the N_{w} profile derived from the radiosonde data is compared with the corresponding vertical profile reconstructed using the tomographic model. The optimum value of k is selected when the root-mean-square error (RMSE) is minimal. RMSE, bias, and standard deviation (SD) of the results are calculated as follows:

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(N^i_{\mathrm{model}} - N^i_{\mathrm{calculated}}\right)^2} \tag{7}$$

$$\mathrm{bias} = \frac{1}{N}\sum_{i=1}^{N}\left(N^i_{\mathrm{model}} - N^i_{\mathrm{calculated}}\right) \tag{8}$$

$$\mathrm{SD} = \sqrt{\mathrm{RMSE}^2 - \mathrm{bias}^2} \tag{9}$$

where N^i_{model} is the wet refractivity obtained from the tomography model in the ith voxel and $N^i_{\mathrm{calculated}}$ is the wet refractivity of the corresponding voxel calculated from the radiosonde data.

Constraints in GNSS tomography

Many constraints have been used to resolve the rank deficiency of the design matrix in GNSS tomography. Therefore, we only review some of the commonly used constraints: 3D Gaussian Filter constraints, horizontal smoothing constraints, constraints based on the NWP models and the VRS concept. The first two cases, also known as inter-voxel constraints, smooth the refractivity difference calculated from reconstruction methods or interpolated from NWP models between neighboring voxels.

3D Gaussian filter

3D Gaussian filter is an inter-voxel constraint that may limit the refractivity gradients between adjacent voxels. The refractivity N' in the voxel (i_0, j_0, k_0) is calculated by the weighted sum of the refractivity N within the domain $i_0 \pm \Delta i,\ j_0 \pm \Delta j,\ \pm k_0 \pm \Delta k$ (Bender et al. 2011),

$$N'_{i_0 j_0 k_0} = \frac{\sum_{i=i_0-\Delta i}^{i_0+\Delta i}\sum_{j=j_0-\Delta j}^{j_0+\Delta j}\sum_{k=k_0-\Delta k}^{k_0+\Delta k}\alpha_{i,j,k}N_{i,j,k}}{\sum_{i=i_0-\Delta i}^{i_0+\Delta i}\sum_{j=j_0-\Delta j}^{j_0+\Delta j}\sum_{k=k_0-\Delta k}^{k_0+\Delta k}\alpha_{i,j,k}} \tag{10}$$

where α is the Gaussian weighting factor, which is defined by

$$\alpha_{ijk} = e^{-\frac{(x-x_0)^2}{2\sigma_1^2}} e^{-\frac{(y-y_0)^2}{2\sigma_2^2}} e^{-\frac{(z-z_0)^2}{2\sigma_3^2}} \tag{11}$$

The number of neighboring voxels Δi, Δj, and Δk as well as the value of σ are selected such that the best result is acquired. Usually, one or two adjacent voxels and $\sigma = 0.5, ..., 2$ lead to good results, and this is the reason that $\sigma = 2$ and $\Delta i = \Delta j = \Delta k = 1$ are considered in this research (Bender et al. 2011). The number of adjacent voxels used

for each voxel depends on its location in the model. Accordingly, 26 adjacent voxels are used for a voxel inside the middle layers in (10).

Horizontal constraints

Rohm and Bosy (2009) used horizontal constraint equations in such a way that the refractivity of each voxel, for example, *k*th voxel, is a weighted mean of the refractivities of neighboring voxels in the same layer:

$$0 = \frac{1}{w_{k1}} N_{w_1} + \frac{1}{w_{k2}} N_{w_2} + \cdots - 1 \times N_{w_k} + \cdots + \frac{1}{w_{km}} N_{w_m} \tag{12}$$

where w_{kj} is the weight of refractivity between the constrained voxel (*k*th voxel) and the voxel number *j*. These weights are the inverse of the distance between the constrained voxel and all other voxels in the same layer. The sum of all weights should be equal to 1. It should be mentioned that 14 neighboring voxels are used for each voxel in this research because each layer of the model contains 15 voxels.

NWP constraints

The refractivity computed from interpolated NWP models in the center of voxels can be used for constraining the problem:

$$x_i - x_i^{\mathrm{NWP}} = 0 \tag{13}$$

where x_i is the unknown value of refractivity, and x_i^{NWP} is the refractivity computed from the NWP model at the center of the *j*th voxel. Weather Research and Forecasting model (WRF) is one of the most common atmospheric simulation models used in this research. The global larger scale NWP models such as the global forecast system (GFS), the coupled forecast system (CFS), and regional NWP models such as the North American Mesoscale model (NAM) are used to generate the initial and boundary conditions of the WRF model (Skamarock et al. 2008). This study takes the initial and boundary conditions from GFS 3 h forecast data. Also, the horizontal resolution of the GFS model outputs is $0.5° \times 0.5°$ (Yáñez-Morroni et al. 2018). NWP models predict the 3D structure of the troposphere for the next 6, 12, 18, … to 120 h. This study used 24 h predictions of the WRF model with spatial and temporal resolutions of 10 km and one hour, respectively. The required parameters (temperature, special humidity, and height) are calculated in 25 pressure levels using the WRF model. The coordinates of the eight vertices of each voxel were firstly calculated. Values of the pressure of water vapor and temperature acquired from the WRF model were then interpolated to the calculated positions. Finally, N_w values are computed through the following equation (Kleijer 2004):

$$N_w = K_2 \frac{e}{T} + K_3 \frac{e}{T^2} \tag{14}$$

Here, $K_2 = 71.2952$ K/hpa, $K_3 = 375463$ K^2/hpa, and the parameters e and T are the pressure of water vapor and the air temperature, respectively (Kleijer 2004). The average value of N_w at voxel vertices is considered the wet refractivity in the corresponding voxel.

VRS constraints

Adavi and Mashhadi-Hossainali (2014) used the concept of VRS as extra information to compute a unique solution for the problem of troposphere tomography. In their study, the interpolation of tropospheric error at the position of VRS stations is below 10^{-3} m when the distances of VRS and GNSS stations are less than 40–50 km. They found that the minimum number of virtual stations, which solves the rank deficiency of the design matrix, is attained when VRS stations are located at the corners of the voxels. We used the same method to determine the number and location of VRS stations with a maximum distance of 50 km from GNSS reference stations. Virtual signals at virtual reference stations add additional equations to the system of observation equations. The interpolation algorithm proposed by Zus et al. (2019) is used in this research to generate ZWDs in the position of VRS stations. For any VRS station, GNSS stations within the radius of 1° are considered, and the ZWD value in the VRS position is calculated using the following equation:

$$\begin{aligned} \mathrm{ZWD}_{\mathrm{VRS}} =& \Sigma w^i . \mathrm{ZWD}^i . \exp\left(-\frac{\Delta h^i}{H}\right) \\ &+ \Sigma w^i . \left(\frac{E_w^i}{C} R \cos\left(\varphi^i\right) \Delta\lambda^i + \frac{N_w^i}{C} R \Delta\varphi^i\right) . \exp\left(-\frac{\Delta h^i}{H}\right) \end{aligned} \tag{15}$$

Here, w is the weight, and the inverse distance weighting scheme is used. $\Delta\lambda$, $\Delta\varphi$, and Δh are the differences in longitude, latitude, and height between the VRS and the GNSS stations, respectively. The index *i* indicates the *i*-th GNSS station. The factor *C* is set to 4 km. The scale height of the ZWD, denoted H, is chosen to be 3 km. Also, E_w^i and N_w^i indicate the east and north gradients of the wet components, respectively. Similar to the ZTD, the tropospheric gradient can be written as the sum of the hydrostatic and wet components:

$$\begin{aligned} E &= E_h + E_w \\ N &= N_h + N_w \end{aligned} \tag{16}$$

where $N_h(N_w)$ is the north gradient of the hydrostatic (wet) component and $E_h(E_w)$ is the east gradient of this parameter.

The wet gradients are obtained from the tropospheric gradients by subtracting the hydrostatic gradients that are derived from the ray-trace algorithm in this research. First, 120 STDs are computed at elevation angles of 3°, 5°, 7°, 10°, 15°, 20°, 30°, 50°, 70°, 90°, and azimuths between 0° and 360° with a spacing of 30° (Zus et al. 2014). Next, we compute azimuth-independent STDs under the assumption of a spherically layered troposphere. Then, the differences between azimuth-dependent and azimuth-independent STDs are computed. Finally, the gradient components are determined by a least-square fitting (Zus et al. 2015). The Chen and Herring gradient mapping function is used. To calculate hydrostatic gradients, we compute the hydrostatic delays separately.

These values are then converted to SWDs using the GMF in the direction of VRS stations to GNSS satellites. The method is useful when the number of GNSS stations is not sufficient and they are not evenly distributed in the desired area.

Sensitivity analysis in GNSS-tomography

Consider the linear system of equations:

$$\mathbf{Ax}=\mathbf{b},\quad \text{where}:\quad \mathbf{A}\in R^{m\times n},\quad \mathbf{x}\in R^{n\times 1},\quad \mathbf{b}\in R^{m\times 1}\quad \text{and}\quad m\geq n \tag{17}$$

In practice, two types of errors can contaminate the elements of the coefficient matrix **A** and the vector **b**: observational and computational (rounding and truncation) errors. The main aim of the perturbation theory is to analyze the effect of such perturbations on the estimated parameters compared to the exact solution (Dief 1986). Assuming $n=m$ and **A** is not singular, the upper bound limit for the error in the exact solution **x** in the perturbed linear system $(A+\Delta A)\tilde{x}=b+\Delta b$ is,

$$\frac{\|\tilde{\mathbf{x}}-\mathbf{x}\|}{\|\mathbf{x}\|}\leq\frac{\mathbf{k}_2(\mathbf{A})}{1-\|\mathbf{A}^{-1}\Delta\mathbf{A}\|}\left[\frac{\|\Delta\mathbf{b}\|}{\|\mathbf{b}\|}+\frac{\|\Delta\mathbf{A}\|}{\|\mathbf{A}\|}\right] \tag{18}$$

where $\tilde{\mathbf{x}}$ is the vector of perturbed unknown parameters and $\mathrm{k}_2(\mathbf{A})=\|\mathbf{A}\|_2\,\|\mathbf{A}^{-1}\|_2$ is the condition number (Jain and Iyengar 2003). Ignoring the perturbations ΔA in the perturbed linear system above gives the following upper limit for the error in the exact solution:

$$\frac{\|\tilde{\mathbf{x}}-\mathbf{x}\|}{\|\mathbf{x}\|}\leq k_2(\mathbf{A})\frac{\|\Delta\mathbf{b}\|}{\|\mathbf{b}\|} \tag{19}$$

Equation (19) shows in a system of linear equations that the upper bound limit for perturbations in the exact solution depends on the input noise and is proportional to the condition number of the system of observation equations. Assuming $m>n$ and **A** is not singular, the upper bound limit for the exact least-squares solution due to perturbations $\Delta\mathbf{A}$ and $\Delta\mathbf{b}$ is given by Higham (2002):

Theorem 1 *Let* $\mathbf{A}\in\mathbf{R}^{m\times n}$, $(m\geq n)$ *and* $\mathbf{A}+\Delta\mathbf{A}$ *be of full rank and let:* $\|\mathbf{b}-\mathbf{Ax}\|_2=\min$, $\mathbf{r}=\mathbf{b}-\mathbf{Ax}$; $\|(\mathbf{b}+\Delta\mathbf{b})-(\mathbf{A}+\Delta\mathbf{A})\tilde{x}\|=\min$, $\mathbf{a}=\mathbf{b}+\Delta\mathbf{b}-(\mathbf{A}+\Delta\mathbf{A})\tilde{x}$; $\|\Delta\mathbf{A}\|\leq\varepsilon\|\mathbf{A}\|,\|\Delta\mathbf{b}\|\leq\varepsilon\|\mathbf{b}\|$, *then provided that* $k_2(\mathbf{A})\varepsilon<1$,

$$\frac{\|\tilde{\mathbf{x}}-\mathbf{x}\|}{\|\mathbf{x}\|}\leq\frac{\mathrm{k}_2(\mathbf{A})\varepsilon}{1-\mathrm{k}_2(\mathbf{A})\varepsilon}\left(2+\left(\mathrm{k}_2(\mathbf{A})+1\right)\right)\frac{\|\mathbf{r}\|}{\|\mathbf{A}\|\,\|\mathbf{x}\|} \tag{20}$$

It follows from (20) that the sensitivity of the least-squares solution is measured by $\mathrm{k}_2(\mathbf{A})$, when the residuals are small or zero, and by $\mathrm{k}_2(\mathbf{A})^2$ otherwise.

In GNSS tomography, the noise of the input data differs from one element to the next because the measurement noise depends on the azimuth and elevation angle of the GNSS satellites. As a result, the residual or misfit of the model to the input data is different in different parts of a tomographic model. Considering this point and the above theorem, the model's sensitivity is different between various layers of a model and in the voxels of the same layer. As a result, we cannot expect the same accuracy and precision as for voxels where reference information such as radiosonde data is used to validate the model. Here, we propose a method for sensitivity analysis of a tomographic model based on principal component analysis. Although using this method, it is not possible to quantify or predict the accuracy and precision of a model at locations without validation data, it provides more details on the expected quality of the inverse solution in sensitive locations.

Principal component analysis can be implemented for any positive-definite matrix. In this research, we apply this method to the normal matrix of our tomographic model. For the positive-definite matrix $\underline{\sum}$ and the random vector variable $\mathbf{X}=\left[x_1\ x_2\ \ldots\ x_p\right]^T$ with eigenvalue–eigenvector pairs $(\lambda_1,\mathbf{e_1}),(\lambda_2,\mathbf{e_2}),\ldots,(\lambda_p,\mathbf{e_p})$, where $\lambda_1\geq\lambda_2\geq\ldots\geq\lambda_p\geq 0$, the hth principal component is given by Johnson and Wichern (2002):

$$y_h=\mathbf{e_h^T X}=e_{1_h}x_1+e_{2_h}x_2+\ldots+e_{p_h}x_p\,,\quad h=1,2,\ldots,p \tag{21}$$

The correlation coefficient of the jth variable and the hth principal component is calculated through (Johnson and Wichern 2002):

$$cor\left(x_j,y_h\right)=\frac{e_{jh}\sqrt{\lambda_h}}{\sigma_j},\quad j,h=1,2,\ldots,p \tag{22}$$

where e_{jh} is the jth element of the hth eigenvector and $\sigma_j=\sqrt{\mathrm{var}(x_j)}$ is the square root of the jth diagonal element in the normal matrix.

PCA can be used as a tool for indexing the unknowns in a system of simultaneous equations with respect to the spectral values of the corresponding coefficient matrix. This indexing process helps arrange the parameters of a system according to their interrelation with various singular values. Mashhadi-Hossainali (2006) proposed PCA application for the sensitivity analysis of strain parameters to perturbations of geodetic observations in deformation analysis. For this purpose, the normal matrix of the tomographic model is first expressed in its spectral form. Then, the corresponding principal components are established using (21). Next, the correlation coefficients between the principal components and all unknown parameters are estimated using (22). Finally, in every principal component, the computed correlation coefficient is used to sort the unknowns, i.e., the refractivity parameters, in ascending order. If a singular value was small or close to zero, i.e., in a badly conditioned system of simultaneous equations, random variables with larger correlations would be more sensitive to perturbation of the input data, or the inverse solution for such variables (voxels) would be more biased when the regularized solution is computed.

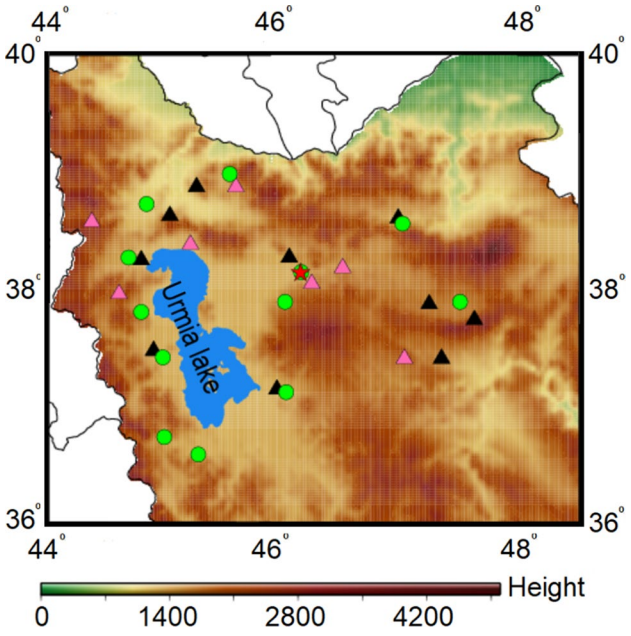

Fig. 1 Location of the study domain together with GPS stations (triangles), radiosonde station (red star), and synoptic stations (green circles). Black and pink triangles represent GPS stations with and without meteorological sensors, respectively

Data

Iran has diverse climate conditions. Altitude, latitude, sea effects, seasonal winds, and proximity to mountain slopes or deserts all play an important role in daily and seasonal atmospheric fluctuations in this country. The Iranian Permanent GNSS Network for Geodynamics (IPGN), with 125 permanent GNSS stations throughout the country, has a good density. In this research, only GPS observations are used because in 2011, active GLONASS satellites were limited in number and could not be tracked by IPGN receivers. We use 17 permanent GPS and a radiosonde station in the northwest of Iran. Figure 1 illustrates the distribution of the GNSS, radiosonde, and synoptic stations used in this study, with the topography of this area. The GNSS stations with and without meteorological sensors are marked separately in this figure.

The Bernese GNSS software version 5.2 is used to process the GNSS measurements based on a double-difference strategy. The desired parameter is the Zenith Total Delay (ZTD). We used the surface meteorological parameters and the Saastamoinen model to compute the Zenith Hydrostatic Delays (ZHDs) (Kleijer 2004). The hydrostatic component is then subtracted from the total zenith delay to calculate the wet component. Finally, ZWDs are converted to Slant Wet Delays (SWDs) using (3).

In the area selected for this research, 10 GPS stations are equipped with meteorological sensors. Together with 12 synoptic stations, these stations were used and interpolated spatially to nearby GNSS stations without meteorological sensors (see Fig. 1). At the farther GNSS stations, the ECMWF Reanalysis v5- Land (ERA5-Land) model was used to calculate the surface meteorological parameters. ERA5 is the latest climate reanalysis produced by ECMWF, providing hourly data on many atmospheric, land-surface and sea-state parameters. ERA5-Land is a replay of the land component of the ERA5 climate reanalysis with a spatial resolution of 9 km and provides hourly high-resolution data of surface variables (https://climate.copernicus.eu/climate-reanalysis).

Results and discussion

A tomography model was developed with a horizontal resolution of 40 km. For the bottom eight and top seven layers of this model, the vertical resolutions were 500 m and 1 km, respectively. The developed model consists of 225 voxels located in 15 horizontal layers and 5×3 voxels in each layer. The model extends to the height of 11 km from the surface of the earth. In this research, DOY 300 and 250 in 2011 were selected to investigate the effect of time and weather conditions on the results. Based on the radiosonde data, the selected area and processing time, the temperature and relative humidity values on the DOY 250 were 16.34 °C and 35.63%, whereas the values of the same quantities on the DOY 300 were 4.27 °C and 80.47%. Also, precipitation water was calculated by radiosonde data for our processing

time as 0.01 and 2.1 mm for DOY 250 and 300, respectively. In addition, 15 GPS stations were considered to obtain the results of DOY 300. Then, two new stations were added to the model of DOY 250 to examine the impact of the number of stations on the results (see Fig. 4). Moreover, a time resolution of 1 h and a minimum elevation cut-off angle of 5 degrees were applied. Finally, the first epoch, which includes GNSS observations from (00 h:00 m:00 s- to 01 h:00 m:00 s in UTC), is considered due to the limited radiosonde data in Iran. Figure 2 illustrates the 3D structure of the model with the signals received from GPS satellites in the first 30 s of the DOY 300 in 2011 (00 h:00 m:00 s-00 h:00 m:30 s UTC). The scales of this figure are not uniform for better visualization.

The design matrix of the model is calculated based on signal length in every element of the model. The rank deficiency of the model is 16 at the first epoch of DOY 300. The relevant empty voxels are shown in Fig. 3.

As expected, many empty voxels are located in the lower layers of the model. To analyze the model's sensitivity, we first fix the rank defect by adding additional constraints to the problem. To this end, and to see if the contribution of applied constraints on the sensitivity of a tomographic model is remarkable, we use the 3D Gaussian, horizontal, NWP, and VRS constraints. Before adding the constraints, the condition number of the design matrix is 1.0967×10^{18}. Added constraints not only remove the rank deficiency of the design matrix but also improve the conditioning of the problem: The condition number of the coefficient matrix improves to 771.3600, 252.5676, 235.3039, and 838.7082 when the 3D Gaussian, horizontal, NWP and VRS constraints are used at this specific epoch, respectively. We follow Adavi and Mashhadi-Hossainali (2014) for selecting the number and the spatial distribution of the virtual reference stations. Figure 4 shows the distribution of GPS and VRS stations in the lowest layer of the tomographic model.

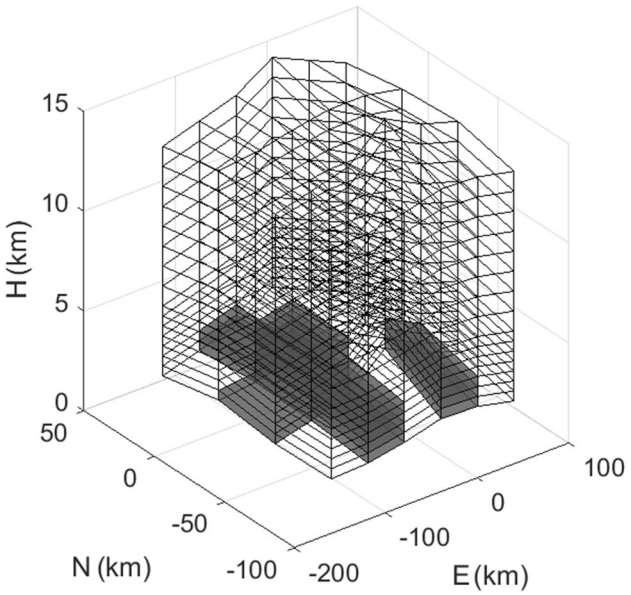

Fig. 3 Empty voxels shaded in gray in the first epoch of DOY 300, 2011

For eigenvalues close to zero, the N_w parameters whose correlation coefficient is higher with the related principal component are prone to larger bias when we derive the inverse solution. In perturbation terminology, such voxels are more sensitive to input perturbations than the other elements of a tomography model. The smallest eigenvalue depends on the constraints we apply to find a unique solution. This is vividly seen from the condition numbers of the coefficient matrix, as already reported for the first epoch. In order to see if the applied constraints contribute to the model's sensitivity, we use the $\sum_{i=1}^{j} \lambda_i / \sum_{i=1}^{p} \lambda_i$ ratio as a relative measure to select the smallest eigenvalue in the spectral representation of the coefficient matrix. We accept 90% as the threshold value in our analysis. This threshold value is reached at the 21st eigenvalue when 3D-Gaussian, horizontal, and NWP

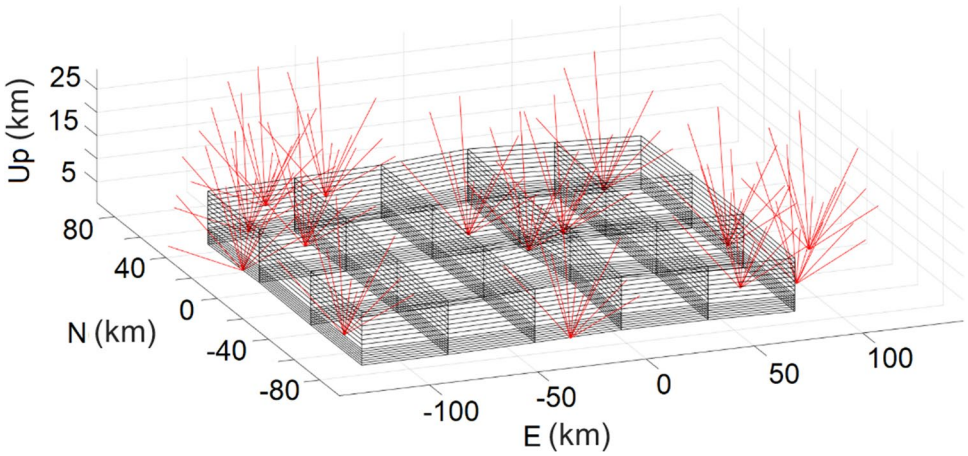

Fig. 2 GPS Signals passing through the voxels of our tomography model in the first 30 s of DOY 300, 2011

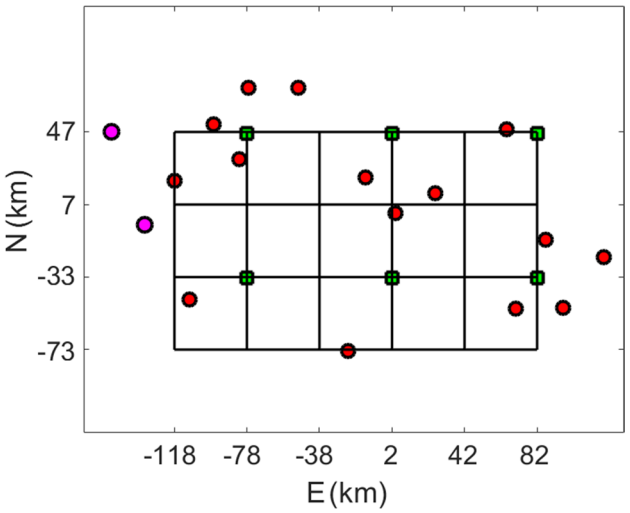

Fig. 4 Spatial distribution of GPS (circles) and VRS stations (squares) on the model. The pink circles show the GPS stations added to the model in DOY 250

constraints are used, and at the 6th eigenvalue when the VRS constraint is used, assuming eigenvalues are listed in ascending order. Figure 5 illustrates the sensitivity analysis results for the first epoch of measurements.

The shaded voxels in Fig. 5 show the sensitive parts of the model. Here, only the voxels whose correlation coefficient with the desired PC is larger than a maximum threshold have been taken into account. This threshold is the maximum correlation of the desired PC with the N_w parameters relevant to voxels through which the radiosonde profile passes. Therefore, the values 0.0332, 0.0394, 0.0382, and 0.0877 in the top left, top right, bottom left, and bottom right panels, respectively, are selected. According to the obtained results, the sensitive parts of the model are not necessarily composed of signal-free voxels and may also occur in other parts of the model. Also, the sensitive parts of the model are often similar when the first three methods are used to constrain the model. This is clearly seen in Fig. 5. Table 1 reports the IDs of sensitive voxels when four types of constraints are used. Subscripts in parentheses refer to the number of layers to help understand the position of sensitive voxels.

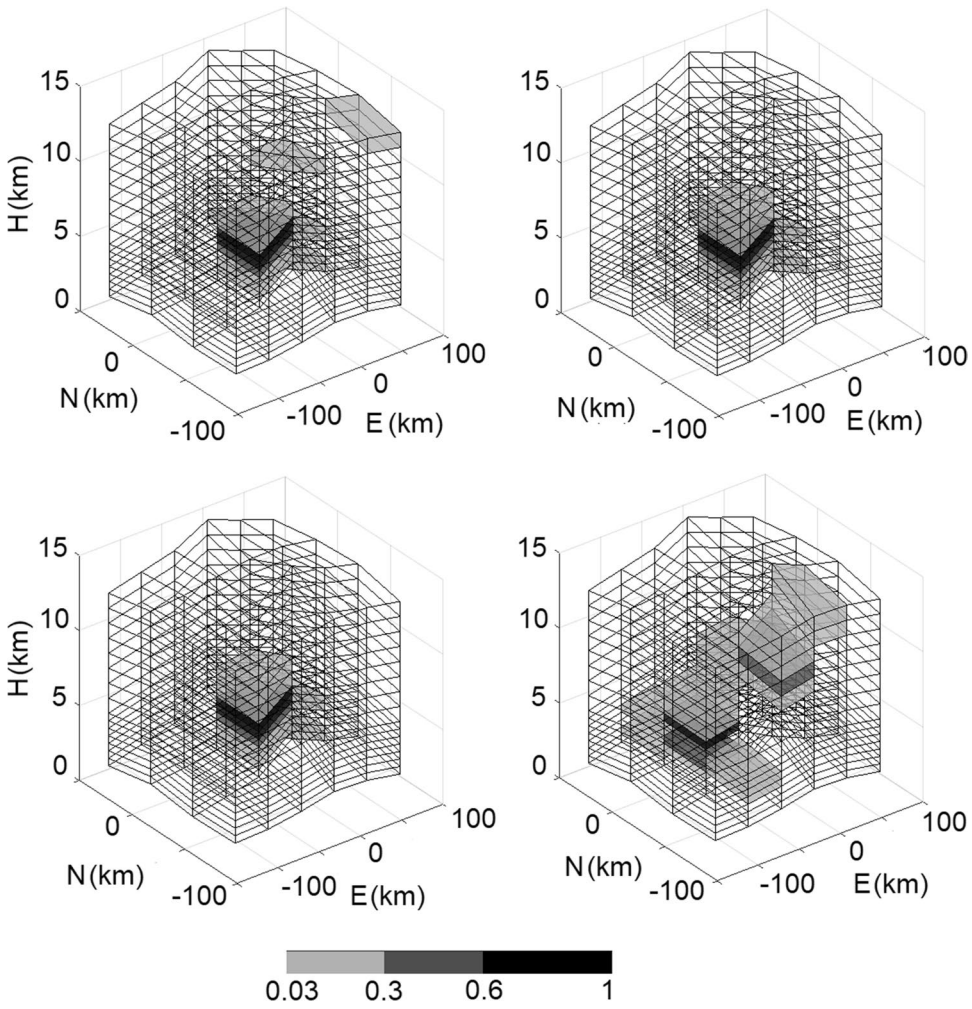

Fig. 5 Sensitivity analysis of tomography model constrained using 3D Gaussian (top left), horizontal (top right), NWP (bottom left), and VRS (bottom right) constraints at the first epoch of DOY 300, 2011

Table 1 IDs of the sensitive voxels derived from PCA technique for the model constrained using four types of constraints for the first epoch of DOY 300, 2011. Subscripts in parentheses specify layer numbers for sensitive voxels

Constraints	DOY 300
3D Gaussian	$159_{(11)}$, $29_{(2)}$, $215_{(15)}$, $9_{(1)}$, $23_{(2)}$, $24_{(2)}$, $38_{(3)}$, $69_{(5)}$, $113_{(8)}$, $84_{(6)}$, $68_{(5)}$, $98_{(7)}$, $83_{(6)}$
Horizontal	$29_{(2)}$, $128_{(9)}$, $9_{(1)}$, $23_{(2)}$, $24_{(2)}$, $38_{(3)}$, $53_{(4)}$, $69_{(5)}$, $84_{(6)}$, $113_{(8)}$, $68_{(5)}$, $98_{(7)}$, $83_{(6)}$
NWP	$128_{(9)}$, $9_{(1)}$, $23_{(2)}$, $24_{(2)}$, $38_{(3)}$, $53_{(4)}$, $69_{(5)}$, $84_{(6)}$, $113_{(8)}$, $68_{(5)}$, $98_{(7)}$, $83_{(6)}$
VRS	$37_{(3)}$, $72_{(5)}$, $143_{(10)}$, $62_{(5)}$, $183_{(13)}$, $214_{(15)}$, $27_{(2)}$, $47_{(4)}$, $127_{(9)}$, $158_{(11)}$, $153_{(11)}$, $199_{(14)}$, $112_{(8)}$, $168_{(12)}$, $52_{(4)}$, $97_{(7)}$

The obtained pattern for the model's sensitivity significantly changes when we use the VRS constraint. Application of 3D Gaussian, horizontal, and NWP constraints to the simultaneous system of observation equations results in an augmented coefficient matrix by appending a rectangular identity or close to identity sub-matrix to the original one. Therefore, matrix A and the normal matrices of the systems of observation equations are almost the same when these types of constraints are used. It is usually necessary to apply a minimum number of constraints to an inverse problem in order to ensure that the solutions will satisfy the constraints. By over-constraining a problem, the solution reflects the information inherent in the applied constraints rather than the information hidden in the input data. On the other hand, VRS constraints are applied by adding virtual observations to the simultaneous system of observation equations. The spatial distribution and the number of virtual reference stations are dependent on the available GNSS stations in the test area. Moreover, the relevant virtual observations are used for the entire period adopted as the time response of the model (here, 1 h). Similar to GNSS observations, the length of the virtual signals passing through the model elements (voxels) is used to augment the coefficient matrix of the original problem. Consequently, the coefficient matrix and the normal matrix are significantly different from when the other constraints are used. As a result, a significant difference between the VRS and other results is expected.

To get an idea of the expected range of bias in reconstructed tomographic images, we use the Landweber method for computing our inverse solutions. This is repeated for the four types of constraints we have focused. The initial values for the refractivities required in the applied recursive approach are supplied by NWP models. In this research, the WRF model has been used. The reconstructed tomographic profiles for the wet refractivities at the radiosonde location are shown in Fig. 6 for DOY 300 and 250 in 2011 at the first epoch.

Panels 6a to 6d represent the reconstructed tomographic results for the model constrained with 3D-Gaussian, horizontal, NWP, and VRS constraints, respectively, for DOY 300, and panels 6e to 6 h show the same results for DOY 250. The values of RMSE, SD, and bias of the tomography model are calculated using the radiosonde and tomography results. These values are given in Table 2. Table 3 compares the mean correlation coefficients (see (22)) calculated for the voxels that the radiosonde profile passes through and the voxels already identified as sensitive.

According to Table 3, since the mean correlation in the voxels of the radiosonde profile is much lower than the mean correlation in sensitive voxels for both days, the magnitude of the bias calculated using the radiosonde profile cannot be generalized to the entire model.

Since the design matrix changes over time, the model's sensitivity is also dependent on time. In the following, we verify this by repeating our analysis for the first epoch of DOY 250 in 2011 to reconstruct the tomographic image.

The comparison of Figs. 5 and 7 confirms the above assertion. The threshold values adopted for the correlation coefficient at the first epoch of DOY 250 are 0.022, 0.0519, 0.0514 and 0.0213. The shaded voxels in Fig. 7 indicate where the model is expected to have larger biases than the maximum amount of bias resulting from the assessment of the model using the radiosonde profile. The IDs of the shaded voxels are given in Table 4. Subscripts in parentheses refer to the number of layers to help understand the position of sensitive voxels.

According to Table 4 and similar to Table 1 for DOY 300, the model's sensitivity is quite different for VRS constraints. Moreover, voxels that are more prone to regularization error are not only scattered through the model, but the expected amount of bias is also different for them. In other words, for the test area in this study, the type of added constraints impacts the bias of the model. Such results also imply that the traditional validation method, i.e., using a radiosonde profile to evaluate the results and generalize them to the entire model, is not sufficient. Due to the existing limitations in this study, for example, a limited number of GPS stations and lack of GPS data on other days of the year, further investigation encompassing tomography models with GPS + GLONASS data, different number of GNSS stations, and more testing days are encouraged. However, it is not possible to expect similar results in another test area, because none of the test conditions are repeatable. For example, the sky view of GNSS

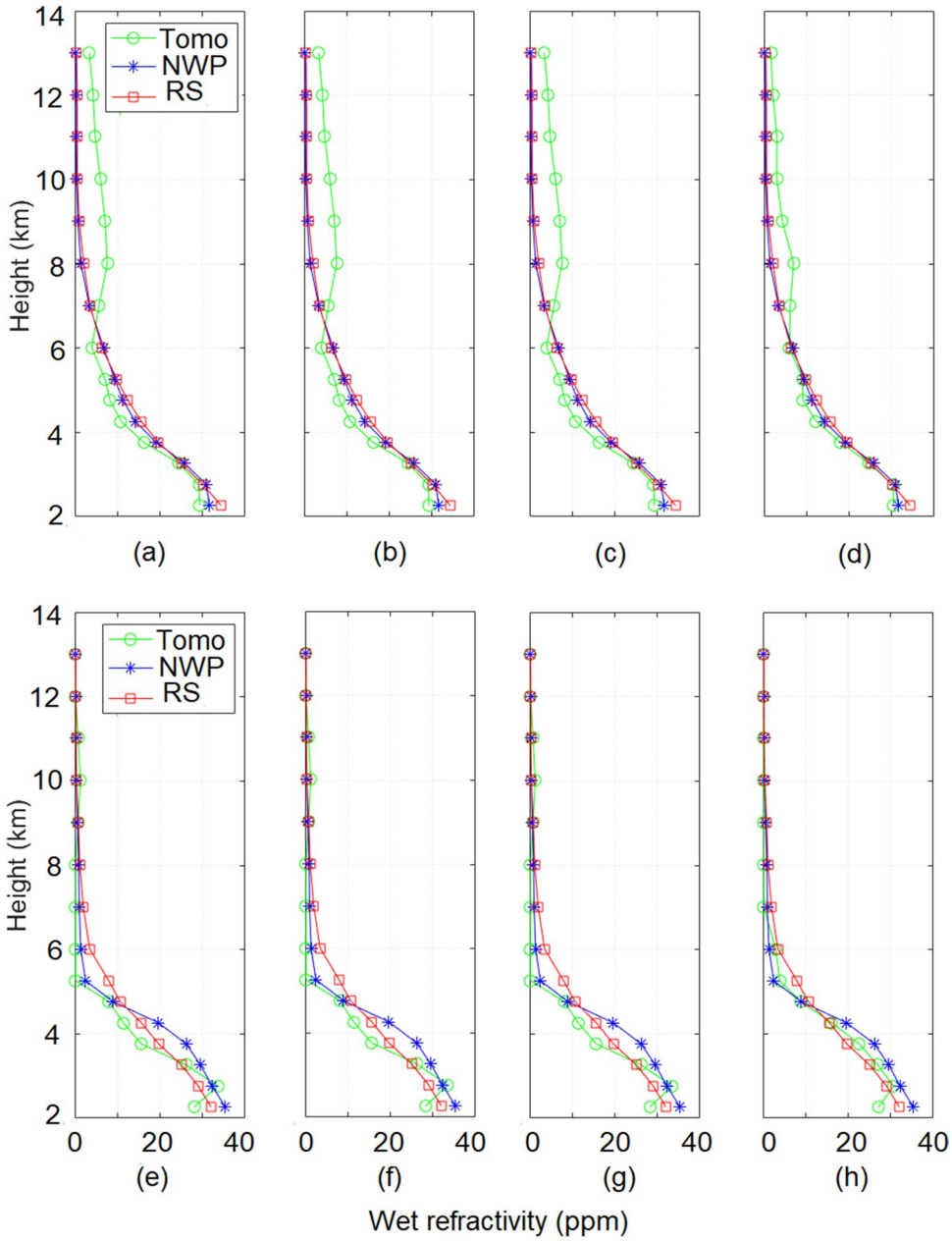

Fig. 6 Tomography results when using four types of constraints together with the radiosonde and NWP profiles for the first epoch of DOY 300 (top) and DOY 250 (bottom), 2011

Table 2 Bias, RMSE, and SD of tomography model constrained using four types of constraints for the first epoch of DOY 300 and 250, 2011

Constraints	DOY 300			DOY 250		
	Bias (ppm)	SD (ppm)	RMSE (ppm)	Bias (ppm)	SD (ppm)	RMSE (ppm)
3D Gaussian	0.415	3.969	3.990	1.467	2.905	3.254
Horizontal	0.415	3.969	3.990	1.461	2.886	3.235
NWP	0.415	3.969	3.991	1.463	2.883	3.234
VRS	0.334	2.560	2.581	0.535	2.001	2.071

satellites changes from one experiment to another, especially when test areas are far apart. This changes the imaging system from one test area to another. Therefore, this study is not intended to suggest a general interpretation of the performance of the proposed PCA-based technique, but it emphasizes the necessity of the sensitivity analysis of a tomography model. The proposed method helps understand if similar validation results are expected in

Table 3 Mean correlation in radiosonde voxels and sensitive voxels of the model for the first epoch of DOY 300 and 250, 2011

Constraints	DOY 300		DOY 250	
	Mean correlation of sensitive voxels	Mean correlation of radiosonde voxels	Mean correlation of sensitive voxels	Mean correlation of radiosonde voxels
3D Gaussian	0.227	0.010	0.228	0.007
Horizontal	0.248	0.011	0.293	0.013
NWP	0.266	0.011	0.350	0.014
VRS	0.207	0.017	0.332	0.006

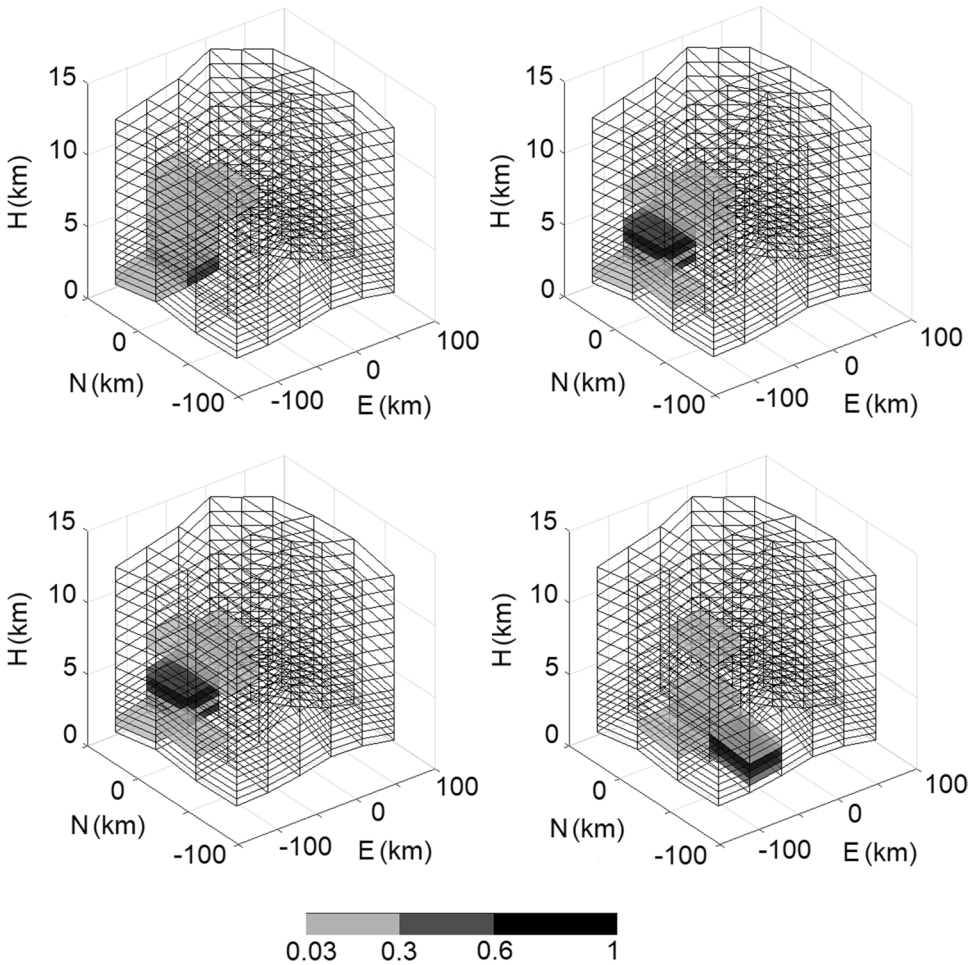

Fig. 7 Sensitivity analysis of tomography model at the first epoch of DOY 250, using 3D Gaussian (top left), Horizontal (top right), NWP (bottom left), and VRS (bottom right) constraints

Table 4 IDs of the sensitive voxels derived from PCA technique for the model constrained using four types of constraints for the first epoch of DOY 250, 2011. Subscripts in parentheses specify layer numbers for sensitive voxels

Constraints	DOY 250
3D Gaussian	$12_{(1)}$, $27_{(2)}$, $72_{(5)}$, $57_{(4)}$, $87_{(6)}$, $42_{(3)}$, $117_{(8)}$, $157_{(11)}$, $26_{(2)}$, $11_{(1)}$, $102_{(7)}$, $147_{(10)}$, $132_{(9)}$
Horizontal	$27_{(2)}$, $57_{(4)}$, $72_{(5)}$, $87_{(6)}$, $157_{(11)}$, $51_{(4)}$, $132_{(9)}$, $42_{(3)}$, $127_{(9)}$, $26_{(2)}$, $77_{(6)}$, $117_{(8)}$, $11_{(1)}$
NWP	$27_{(2)}$, $57_{(4)}$, $72_{(5)}$, $87_{(6)}$, $157_{(11)}$, $51_{(4)}$, $132_{(9)}$, $42_{(3)}$, $127_{(9)}$, $26_{(2)}$, $77_{(6)}$, $117_{(8)}$, $11_{(1)}$
VRS	$32_{(3)}$, $47_{(4)}$, $17_{(2)}$, $52_{(4)}$, $67_{(5)}$, $97_{(7)}$, $62_{(5)}$, $157_{(11)}$, $51_{(4)}$, $7_{(1)}$

other parts of a model or not. The reconstructed parameters in the voxels whose sensitivity to the perturbations of the input parameters (here the SWDs) is higher than the voxels used in validation are prone to a larger bias than the validation process foresees for the model. Finally, it has to be mentioned that the spatial distribution and the required number of virtual reference stations depend on the configuration of the GNSS stations available in the test area.

Summary and conclusion

The precision and the bias of a tomographic model are normally analyzed using radiosonde profile(s). Although the results obtained are vividly valid for a linear vertical profile, they are usually considered quality measures for the whole model. We propose the principal component analysis as a method for sensitivity analysis of a tomographic model for the first time. The application of this method to the test study of this research suggests that some of the voxels are more prone to regularization errors than the other voxels in our model. For the test field and the data we used, the spatial distribution and the expected amount of bias depend on the time and constraints that we normally use to fix the rank deficiency of the problem. Therefore, the results of one vertical profile could not be generalized to the whole model as a quality criterion. The proposed method only suggests a lower bound limit for the accuracy of the model where no validation data is available. Validation results also provide an upper bound limit for the error in the other parts of the model. We applied four types of constraints in our research. According to our results, the sensitivity of the tomography model changes significantly when the VRS constraints are used. VRS constraints add virtual observations to the simultaneous system of observation equations. Although a minimum number of VRS stations are required to resolve the rank deficiency of the normal matrix, the relevant observations are applied for the entire time period, which is adopted as the time response of the model. Therefore, the VRS stations overconstrain the problem compared to other types of constraints, such as 3D-Gaussian, horizontal, and NWP constraints as reported in this research.

Acknowledgements For this study, we would like to thank the National Cartographic Center (NCC) of Iran for providing the GNSS data that were used in this work. We particularly appreciate the meteorological organization of Iran for giving us access to radiosonde profiles at the stations of this research.

Data availability All datasets used in this study can be obtained from the corresponding author.

References

Adavi Z, Mashhadi-Hossainali M (2014) 4D-tomographic reconstruction of the tropospheric wet refractivity using the concept of virtual reference station, case study: North West of Iran. Meteorol Atmos Phys 125:193–205

Aster R, Borchers B, Thurber C (2005) Parameter estimation and inverse problems, 1st edn. Elsevier Academic Press, New York

Aster R, Borchers B, Thurber C (2019) Parameter estimation and inverse problems, 3rd edn. Elsevier Academic Press, New York

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

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. Adv Space Res 47:1704–1720

Bevis M, Businger S, Herring TA, Rocken C, Anthes RA, Ware RH (1992) GPS meteorology: remote sensing of atmospheric water vapor using the global positioning system. J Geophys Res Atmos 97:15787–15801

Böhm J, Niell A, Tregoning P, Schuh H (2006) Global mapping function (GMF): a new empirical mapping function based on numerical weather model data. Geophys Res Lett. https://doi.org/10.1029/2005GL02554

Braun J, Rocken C, Meertens C, Ware R (1999) Development of a water vapor tomography system using low cost L1 GPS receivers. In: Ninth ARM science team meeting proceedings, San Antonio

Brenot H, Rohm W, Kačmařík M, Möller G, Sá A, Tondaś D, Rapant L, Biondi R, Manning T, Champollion C (2020) Cross-comparison and methodological improvement in GPS tomography. Remote Sens 12:30

Chen G, Herring TA (1997) Effects of atmospheric azimuthal asymmetry on the analysis of space geodetic data. Geophys Res 102:20489–20502. https://doi.org/10.1029/97JB01739

Chen BY, Liu ZZ, Wong WK, Woo WC (2017) Detecting water vapor variability during heavy precipitation events in Hong Kong using the GPS tomographic technique. J Atmos Ocean Technol 34:1001–1019

Dief A (1986) Sensitivity analysis in linear systems. Springer

Douša J, Elias M (2014) An improved model for calculating tropospheric wet delay. Geophys Res Lett 41:4389–4397. https://doi.org/10.1002/2014GL060271

Elfving T, Nikazad T, Hansen PC (2010) Semi-convergence and relaxation parameters for a class of SIRT algorithms. Electron Trans Numer Anal 37:321–336

Flores A (1999) Atmospheric tomography using satellite radio signals. Ph.D. Dissertation, Technical University of Catalonia

Guerova G (2003) Application of GPS derived water vapour for numerical weather prediction in Switzerland. Ph.D. Dissertation, University of Bern, Bern

Hanna N, Trzcina E, Möller G, Rohm W, Weber R (2019) Assimilation of GNSS tomography products into the weather research and forecasting model using radio occultation data assimilation operator. Atmos Meas Tech 12(9):4829–4848

Higham NJ (2002) Accuracy and stability of numerical algorithms. SIAM, Philadelphia

Hirahara K (2000) Local GPS tropospheric tomography. Earth Planets Space 52(11):935–939

Jain MK, Iyengar SRK, Jain RK (2003) Numerical methods for scientific and engineering computation. New Age International (P) Limited Publishers, Delhi

Johnson RA, Wichern DW (2002) Applied multivariate statistical analysis. Prentice Hall, Upper Saddle River

Kleijer F (2004) Troposphere modeling and filtering for precise GPS leveling, Delft, The Netherlands

Liou YA, Lin YJ, Chiang CC, Huang CY (2003) Reconstructing 3D wet refractivity structures of the lower troposphere from GPS measurements, In: Proceedings of the GPS-MET conference, Tsukuba, Japan

Mashhadi-Hossainali M (2006) A comprehensive approach to the analysis of the 3D kinematics of deformation. Ph.D. dissertation, Institute of Physical Geodesy, Darmstadt University of Technology, p 152

Nikazad T (2007) The use of Landweber algorithm in image reconstruction. Ph.D. dissertation, Linkopings University, Linkopings

Rohm W, Bosy J (2009) Local tomography troposphere model over mountains area. Atmos Res 93(4):777–783

Rohm W, Bosy J (2011) The verification of GNSS tropospheric tomography model in a mountainous area. Adv Space Res 47:1721–1730

Shafei MJ, Mashhadi-Hossainali M (2020) Application of the GPS reflected signals in tomographic reconstruction of the wet refractivity in Italy. J Atmos Solar Terr Phys. https://doi.org/10.1016/j.jastp.2020.105348

Skamarock WC, Klemp JB, Dudhia J, Gill DO, Barker D, Duda MG, Powers JG (2008) A description of the advanced research WRF version 3. NCAR technical note NCAR/TN-475+STR. National Center for Atmospheric Research, Boulder, Colorado, USA. https://doi.org/10.5065/D68S4MVH

Yáñez-Morroni G, Gironás L, Caneo M, Delgado R, Garreaud R (2018) Using the weather research and forecasting (WRF) model for precipitation forecasting in an Andean region with complex topography. Atmosphere 9:304. https://doi.org/10.3390/atmos9080304

Zhao Q, Liu Y, Ma X, Yao W, Li X (2020) An improved rainfall forecasting model based on GNSS observations. IEEE Trans Geosci Remote Sens. https://doi.org/10.1109/TGRS.2020.2968124

Zus F, Dick G, Douša J, Heise S, Wickert J (2014) The rapid and precise computation of GPS slant total delays and mapping factors utilizing a numerical weather model. Radio Sci 49:207–216. https://doi.org/10.1002/2013RS005280

Zus F, Dick G, Heise S, Wickert J (2015) A forward operator and its adjoint for GPS slant total delays. Radio Sci 50:393–405. https://doi.org/10.1002/2014RS005584

Zus F, Douša J, Kačmařík M, Václavovic P, Balidakis K, Dick G, Wickert J (2019) Improving GNSS zenith wet delay interpolation by utilizing tropospheric gradients: experiments with a dense station network in central Europe in the warm season. Remote Sens 11(6):674

Masoud Mashhadi Hossainali is an associate professor in Geodesy at K. N. Toosi University of Technology (KNTU). He received his Ph.D. degree at the Darmstadt University of Technology. His current research interests are GNSS meteorology, positioning and navigation, deformation monitoring, and constellation design for various space missions.

Hanieh Tabatabaei received her M.Sc. degree in Geodesy at K. N. Toosi University of Technology (KNTU) in 2013. She is currently a Ph.D. candidate in Geodesy at KNTU and working on ionospheric tomography using statistical approaches. Her research interests are GNSS meteorology, positioning and navigation, and optimization methods.