GPS Solutions (2022) 26:77
https://doi.org/10.1007/s10291-022-01264-w

ORIGINAL ARTICLE

Development of a hybrid tomography model based on principal component analysis of the atmospheric dynamics and GPS tracking data

Elaheh Sadeghi[1] · Masoud Mashhadi Hossainali[2] · Abdolreza Safari[1]

Received: 24 July 2021 / Accepted: 19 April 2022

Abstract
In voxel-based tropospheric tomography, water vapor is computed for a set of voxels, each covering a specific part of the troposphere. The basic assumption is that water vapor is homogenous in any voxel. In this study, the Principal Component Analysis (PCA) has been applied to select the horizontal size of voxels. The application of PCA helps determine the horizontal size of voxels while keeping the atmospheric changes and the basic assumption intact. Application of the proposed method to a numerical atmospheric model in our case studies shows that the largest horizontal changes in wet refractivity (N_w) occurred in layers close to the earth. We also applied this method to tomography models with 30, 40, and 50 km horizontal resolutions. Results were compared using an initial model as a reference. The results indicated that in the lower layers of the atmosphere, the 40 km model has the strongest similarity with our reference in terms of scattering after the 30 km one. Due to limited changes in N_w at the upper layers, the size of voxels was increased to 50 km. N_w was reconstructed for tomographic models with horizontal resolutions of 40 and 50 km and a hybrid model with a combined horizontal resolution in which the model resolution is 40 km in the first seven layers and 50 km in the rest. Validity checks against the radiosonde profile show that the hybrid model reduces the bias and increases the redundancy of our model.

Keywords GPS tomography · Optimize voxel size · PCA · Atmospheric dynamics

Introduction

An accurate estimate of the three-dimensional distribution of water vapor is the main challenge in weather and climatic change prediction. Water vapor measurement techniques include lidar (Bock et al. 2002; Troller et al. 2002), water vapor radiometer (Braun and Rocken 2003), terrestrial sensors (Troller 2004; Lutz 2008), and radiosonde. However, the relatively low spatial and temporal resolution of these sensors and their high cost are among the shortcomings of these technologies. Due to high clarity and accuracy, the radiosonde is accepted as the most common technology and used as a reference technique for validation purposes (Caldas-Alvarez et al. 2021). Today, GNSS is used owing to its adequate spatial and diverse temporal resolution (Haji-Aghajany et al. 2020b; Wang and Dessler 2020).

Three-dimensional tomographic models reflecting the spatial changes in water vapor are generated based on the permanent GNSS stations in an area (Brenot et al. 2014; Haji-Aghajany et al. 2020b). Bevis et al. (1992) first reported high-resolution water vapor GPS-tomography. Since then, others have tried to improve the GPS-tomographic models; for example, Guo et al. (2016) introduced an optimal tomography technique based on the weighted equations. Zhao et al. (2018) introduced the function-based water vapor tropospheric tomography. Haji-Aghajany et al. (2020a) used NWP model outputs and topography of the area to reduce the number of unknown parameters and, therefore, the number of required constraints. Tomography is an inverse method that reconstructs cross-sectional images through penetrating rays or waves (Kak et al. 2002). The troposphere is divided into several elements known as voxels (Flores et al. 2000).

✉ Elaheh Sadeghi
Sadeghi.Elaheh@ut.ac.ir

[1] School of Surveying and Geospatial Engineering, College of Engineering, University of Tehran, Central Building of the College of Engineering, North Kargar Street, Tehran 1439957131, Iran

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

Published online: 11 May 2022

One of the fundamental flaws in GPS tomography is the absence or insufficiency of signals in some voxels. This is due to limitations in the number of receivers and/or elevation angle of satellites, resulting in the non-uniqueness of the sought solution (Bender et al. 2011; Xia et al. 2013). Various methods have been proposed for finding a unique solution to this problem (Flores et al. 2000; Bender et al. 2009). GPS tomography assumes that the water vapor distribution is homogenous across the entire space occupied by a voxel (Zhang et al. 2020; Rohm 2013). So, the size of voxels should include enough observations and should satisfy this requirement. To eliminate voxels with insufficient number of observations, the size of voxels is determined only based on the number of observations in each voxel or the distance between GPS stations on the ground (Adavi and Mashhadi-Hossainali 2014, Yang et al. 2018).

Since weather condition changes with the height of layers, initial knowledge on the atmosphere can help set new constraints on the size of voxels taking the dynamics of the atmosphere into account. With regard to the voxel size, for example, Adavi and Mashhadi-Hossainali (2014) determined the optimal horizontal voxel size by investigating the resolution matrix. Yao and Zhao (2016, 2017) suggested increasing the number of voxels using the concept of non-uniform symmetrical division of horizontal voxels and eliminating the voxels with no intersection or insufficient observations.

This study proposes a method for determining the size of voxels based on the adequate number of observations and incorporating the dynamics of water vapor in the test area. The proposed method is based on PCA as a method in Exploratory Data Analysis (EDA) (Komorowski et al. 2016). Numerical Weather Prediction (NWP) model outputs have as a priori information about the atmosphere. By running the PCA on NWP model outputs in every layer of the tomographic model, it was decided to expand or reduce the voxel sizes of the tomographic model. A brief summary of the tomographic reconstruction of refractivity (N_w) using GPS observations and the theoretical background of the applied method are explained next. Our method has been applied to four tomographic models with different horizontal resolutions. The section on numerical results reports the outcomes and discusses this method. Concluding remarks are given in the last section.

Methodology

Exploratory Data Analysis (EDA) is a way to analyze data that use various techniques and summarizes the main specifications of datasets to visualize the relation between the samples and variables. One of the most successful EDA methods is PCA. PCA is a modeling technique that extracts information from data sets by accessing the relation between variables (Demšar et al. 2013).

A type of PCA suitable for spatial data has been applied in this study. In contrary to non-spatial data that incorporate the measurement of variables or properties, spatial data refer to data determined by space or location attributed to its respective measurement. In contrast, typical data, i.e., non-spatial data, incorporate only the measurement of variables or their properties. Two methods of applying PCA to spatial data can be investigated: the standard non-spatial PCA and adapted PCA. In the standard non-spatial PCA, spatial challenges such as spatial heterogeneity and spatial autocorrelation are avoided, while the adapted PCA considers spatial effects with respect to spatial heterogeneity or autocorrelation. Because it ignores geographic effects, and PCA is only performed on the spatial attribute, we use standard non-spatial PCA, which we explain in detail in the section on principal component analysis of spatial data.

It is possible to analyze the similarities and differences of a group of variables through a comparison of the principal components (PCs) if they are measured on several different samples (Krzanowski 1979). In this study, PCA is used to inspect tomographic models and suggest the one that keeps the atmospheric conditions of the test area intact, taking into account the inter-distance of GPS stations and other available or required data. The main objectives are hence: (a) to inspect horizontal atmosphere changes in the various height layers of the tomographic model, (b) to investigate scattering changes in N_w by increasing the horizontal voxel sizes of a tomographic model, (c) to develop a hybrid tomography model based on the proposed approach.

Principles of GPS tomography

GPS tomography is based on modeling atmospheric delays of GPS signals either in terms of the tropospheric refractivity or electron density in the ionosphere. The relevant mathematical model is the Fredholm Integral Equation (FIE) of the first kind. When the Slant Wet Delays (SWD) are used as the input data, GPS tomography reconstructs the N_w images in three or four dimensions (Flores et al. 2000)

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

Here, S is the signal path between a satellite and a receiver.

The continuous integral (1) is changed to a discrete form. The study region is developed into a series of cubic elements or voxels. Water vapor density or N_w are assumed to change only from one voxel to another (Bender et al. 2013). If n is the number of such elements, then

$$SWD_i = 10^{-6} \sum_{j=1}^{n} N_{wj}\, \Delta s_j \tag{2}$$

In this equation, j represents the voxel confluence by the ith signal with the length of Δs_j. In matrix notation (Flores et al. 2000):

$$\mathbf{SWD} = \mathbf{AN_w} \tag{3}$$

in which $\mathbf{A}$ is the coefficient matrix and includes the signal lengths Δs_j (Rohm and Bosy 2009). $\mathbf{SWD}$ is the vector of observations derived from the analysis of GPS measurements (Adavi and Mashhadi-Hossainali 2014).

GPS tomography is a mixed-determined inverse problem with no unique solution but is sensitive to perturbations of the input (Rohm and Bosy 2009; Bender et al. 2011). For a unique solution, the simultaneous system of observation equations is changed to a constrained system of equations. In this study, the method of virtual reference stations has been used (Adavi and Mashhadi-Hossainali 2014). Moreover, a regularization method is used to find the solution. Recursive methods are used to acquire a regular solution in large-scale ill-posed problems. Landweber algorithm is one of the classical and efficient recursive methods in regularization techniques (Landweber 1951). With this method, the solution of the simultaneous equations $\mathbf{SWD} = \mathbf{AN_w}$ is computed by (Elfving et al. 2010):

$$\mathbf{N}_w^{s+1} = \mathbf{N}_w^{s} + \lambda_s \mathbf{A}^T \left(\mathbf{SWD} - \mathbf{AN}_w^{s}\right) \tag{4}$$

In this equation, λ_i is the relaxing parameter. Using $0 < \lambda_s < 2/\sigma^2_{\max}$, in which $\sigma^2_{\max}$ is the largest eigenvalue of $\mathbf{A}$, ensures solution convergence (Aster et al. 2018). Here, the modified Ψ_2 strategy (Elfving et al. 2010) has been used for computing this parameter. The accuracy and precision of the results should be calculated to investigate the effectiveness of the applied tomographic model. Statistical quantities including Bias, Root-Mean-Square Error (RMSE) and Standard Deviation (STD) are the measures normally used for this purpose (Guerova 2003):

$$\text{Bias} = \frac{1}{n} \sum_{i=1}^{n} \left(N^i_{\text{wm}} - N^i_{\text{wo}}\right) \tag{5}$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (N^i_{\text{wm}} - N^i_{\text{wo}})^2} \tag{6}$$

$$\text{STD} = \sqrt{\text{RMSE}^2 - \text{Bias}^2} \tag{7}$$

In these equations, N_{wm} is the reconstructed N_w in voxel m. N_{wo} is the calculated N_w from the radiosonde data, and n is the number of voxels.

Principal component analysis of spatial data

The PCA maps the original p-dimensional observation matrix ($\mathbf{G}$) of p variables and n measurements onto a new r-dimensional orthogonal space, in which the new axes are toward the largest variance in the data, also known as PCs (Fodor 2002):

$$\mathbf{S}_{n\times r} = \mathbf{G}_{n\times r}\mathbf{U}_{r\times r} \tag{8}$$

$\mathbf{S}$ and $\mathbf{U}$ are the score and loading matrix and $r = \text{rank}(\mathbf{G})$. Columns of $\mathbf{U}$ are in the directions of maximum variance of the data. Therefore, $\mathbf{S}$ transforms the original matrix to the new space. These directions correspond to the eigenvectors of the covariance matrix $\mathbf{\Sigma_X}$ and they are calculated according to (Fodor 2002):

$$\mathbf{\Sigma_G} = \frac{\mathbf{G}\,\mathbf{G}^2}{n}, \qquad \mathbf{\Sigma_G} = \mathbf{U}\Lambda\mathbf{U}^T \tag{9}$$

where

$$\begin{aligned} &\Lambda = diag(\lambda_1,\ \lambda_2,\ \ldots \lambda_r,\ 0,\ \ldots,\ 0) \\ &\quad \lambda_1 \geq \lambda_2 \ldots \geq \lambda_r \geq 0 \end{aligned} \tag{10}$$

and λ_i is the ith eigenvalue of $\mathbf{\Sigma_G}$.

Dimensionality reduction is made by keeping the first k PCs, such that $k \ll n$. The k dimensional space contains most of the information compatible with certain parameters. Choosing the value of k is a key decision and is made by studying the data. PCA is done either in the standard or clustering mode. The standard or R-mode is used in classifying samples by discovering the relation of variables; its goal is to identify the important variables among the remaining. In the second method, known as the Q-mode, loading and score matrices are used to identify the important samples and clusters of the variables (Lee et al. 2017). Due to the nature of the data and the purpose of this study, R-mode PCA is used.

Standard non-spatial PCA on spatial data

The standard non-spatial PCA has been used to analyze data. Four methods are available for this purpose: spatial objects PCA, raster data PCA, atmospheric science PCA, and PCA on flows (Demšar et al. 2013). The raster data PCA has been preferred because its results are used as pre-processing for clustering, classification, change the identification, and tracking of specifications and properties. Here, the PCs are calculated for datasets in which data elements are cells of raster surfaces or spaces at the center of the raster cells with measurements of several variables in each space. Geographic effects are ignored and PCA is only performed on the spatial attribute. Ordered spatial position measurement is essential in this method and the analysis of the first few PCs would be sufficient (Demšar

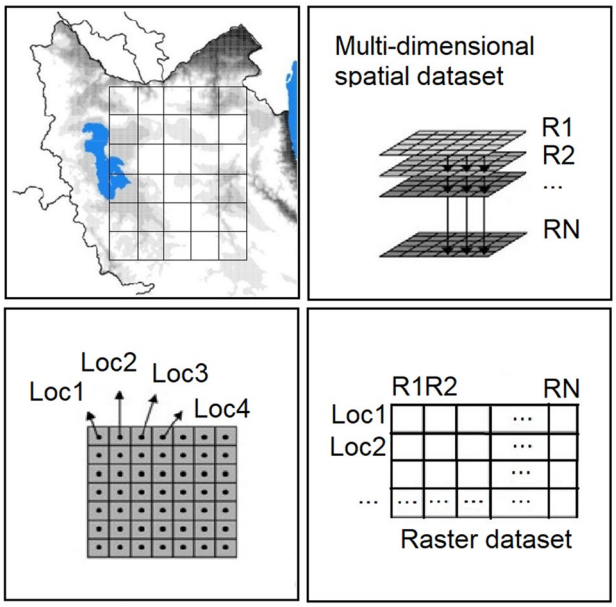

Fig. 1 Visualizing raster data PCA in space domain. Top left: raster visualization of the research area. Top right: raster dataset. R1 … RN are the layers of a tomography model in this research. Bottom left: location of the attributes assigned to each pixel. Voxels in a tomography model are the pixels and Loc1, Loc2, etc., are the voxel centers in one of the layers in a tomography model. Bottom right: arrangement of raster data in raster PCA

et al. 2013). In raster data PCA, the data matrix is constructed with columns equal to the number of measured variables and rows equal to the cell centers in the network under study (Fig. 1).

As illustrated in Fig. 1, measured quantities in corresponding cells of all layers make up one column of the data matrix. In raster datasets, it is important that the cell sizes across the entire network are equal in size.

Comparison of PCs between multivariate samples

Discovering the similarities and differences of samples is the focus when the same variables are measured in multiple samples. Comparison of the PCs among groups with the same variables is an analytical approach for this purpose. Available methods for this comparison are given in the following sections.

Comparison of PCs between two samples

Assuming multivariate samples $\mathbf{G_1}$ and $\mathbf{G_2}$ include p values of the same variable, say $x_1, \ldots, x_p$, each measured n_1 and n_2 times having transposed loading matrices $\mathbf{L}$ and $\mathbf{M}$, respectively, then:

$$y_i = \sum_{j=1}^{p} l_{ij}x_j \quad , \quad z_i = \sum_{j=1}^{p} m_{ij}x_j \qquad (i = 1, \ldots, k) \tag{11}$$

The symbols y_i and z_i are the PCs of datasets $\mathbf{G_1}$ and $\mathbf{G_2}$, respectively. Dimension reduction is made by keeping the first k PCs, i.e., first k rows of matrices $\mathbf{L}$ and $\mathbf{M}$. The transferred data are in the new k dimensional space with orthogonal axes $y_1, \ldots, y_k$ and $z_1, \ldots, z_k$.

The coincidence angle between two subspaces can be used to find the value of k: the smallest angle between a given vector of the first subspace and a vector from the second one almost parallel to it (resulting from its own map on the second subspace). The angle in question is computed as $\theta = \cos^{-1}\left\{(\lambda_1)^{\frac{1}{2}}\right\}$ in which λ_1 is the largest eigenvalue of the matrix $\mathbf{F} = \mathbf{L}\mathbf{M}^T\mathbf{M}\mathbf{L}^T$ (Krzanowski 1979). Therefore, the sum of the eigenvalues of $\mathbf{F}$ equals the sum of squares of the cosines between each of the kth principal components of two matrices $\mathbf{G_1}$ and $\mathbf{G_2}$,

$$\sum_{i=1}^{k} \lambda_i = \text{trace } \mathbf{F} = \sum_{i=1}^{k}\sum_{j=1}^{k} \cos^2 \theta_{ij} \tag{12}$$

The sum of the eigenvalues of $\mathbf{F}$ is between k and zero. If the sum is equal to k, then the two spaces are coincident and if zero, they are orthogonal. Therefore, this value can serve as a general similarity criterion for two datasets.

Comparison of PCs between more than two samples

Assuming g datasets with p variables, the PCA is performed on each dataset individually. The first k PCs of each group are then considered for further analysis. The transpose of the loading matrix is specified $L^{(t)}$ with (i,j) elements, i.e., $L_{ij}^{(t)}$ corresponds to the jth variables on the ith principal component of tth group so that $(i = 1, \ldots, k;\ j = 1, \ldots, p;\ t = 1, \ldots, g)$.

Assuming $\mathbf{b}$ is a vector of the original p dimensional space and δ is the angle between this vector and the vector most nearly parallel to it in the space resulting from the k PCs of the group t $(t = 1, \ldots, g)$, the vector closest to all groups, i.e., the vector for which $V = \sum_{t=1}^{g} \cos^2 \delta_t$ reaches its minimum value, is the result of the eigenvector corresponding to the largest eigenvalue of $\mathbf{H} = \sum_{t=1}^{g} \mathbf{L}_t^T\mathbf{L}_t$ (Krzanowski 1979). If this vector is known as $\boldsymbol{b}$, then vector $\boldsymbol{b}$ is the closest vector to all groups. The angle δ can be used as a criterion for describing the degree of separation of this vector from the tth group. This angle is calculated using the equation (Krzanowski 1979),

$$\delta_t = \cos^{-1}\left\{\left(b_1^T \mathbf{L}_t^T \mathbf{L}_t b_1\right)^{\frac{1}{2}}\right\} \quad (13)$$

Therefore, similarities and differences between groups can be determined by analyzing the eigenvector (**b**) and the separation angle (δ) of **H**. We applied this analytical approach to discover the similarities and differences among our tomography models.

Numerical results and discussion

Choosing the optimal size for the voxels is a challenging task in GNSS tomography. Voxel sizes must be selected in a way to include enough observations as well as be adequately homogeneous regarding the desired parameter N_w. Here, one of the Numerical Weather Prediction (NWP) models has been used to extract initial information on water vapor required for selecting voxels with optimal size in our problem.

As mentioned earlier, PCA is used in multivariate analysis. An important advantage of this method is that it specifies which variable has the greatest variety in the measured sample in multivariate descriptions. This feature convinced us to use PCA to exploit atmospheric changes in the height layers, matching layers of tomographic models, and investigate to what extent the horizontal voxel sizes of our tomographic model can be increased while maintaining the horizontal atmospheric changes intact.

Study area

The northwest region of Iran was selected as the study area for several reasons, including its mountainous terrain in the foothills, a widespread and relatively dense network of permanent GPS stations, and reasonably accurate weather forecasting data (Fig. 2). Along with permanent GPS stations, six virtual stations have been used to constrain the tomography of this study.

Tomographic models

NWP models predict the 3D structure of the troposphere for 6, 12, 18,…to 120 h. Predictions are more accurate in shorter time lags. The Weather Research and Forecasting model (WRF) is one of Iran's most common atmospheric simulation models used for weather prediction. In this study, 24 h predictions have been used on DOY 250 (summer) and 300 (autumn) in 2011 with the resolutions of 10 km and one hour in space and time. The required parameters (temperature, specific humidity, and height) are calculated at 25 pressure levels with pressures of 50, 100, 150,...750, 775,…,975, and 1000 millibar using the WRF model. The weather condition

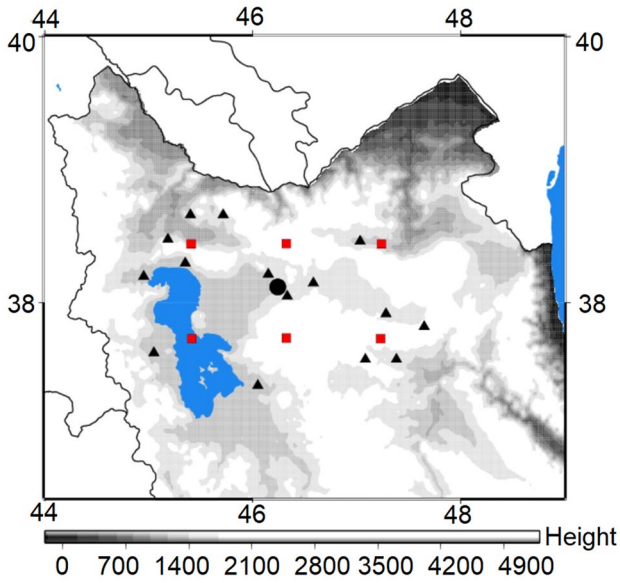

Fig. 2 Distribution of GPS stations (triangles), virtual stations (rectangles), and radiosonde station (circle)

is different for the selected days: According to synoptic data in this area, the average relative humidity in DOY 250 and 300 are 55 and 88%, and the average temperatures are 4.5 °C and 19 °C, respectively.

Since the horizontal resolution of the applied WRF model is 10 km, it has been considered as the initial or reference for the voxel size, assuming that the homogeneity condition is met. Using the horizontal resolution of 10 km, the vertical resolution of 500 m to the height of 4 km and 1 km to an approximate height of 10 km from the surface of the ground together with a temporal resolution of one hour and the existing permanent GPS stations (Fig. 2) the rank deficiency of coefficients matrix **A** (3) becomes large, i.e., 2020 for DOY 250 and 2790 for DOY 300. Computing a unique solution for this problem will cause the final result to be strictly controlled by the constraints rather than the GPS observations. This will also happen for a horizontal resolution of 20 km in which the rank deficiency of coefficients matrix reduces to 251 for DOY 250 and 312 for DOY 300.

Resolution matrices were also calculated for the 30, 40, and 50 km models. For the 30 km model, 13% and 18% of the model parameters are resolved poorly by the inverse solution in the first epoch of DOY 250 and 300, respectively. In other words, the resolution of such parameters is either zero or close to zero (Aster et al. 2018). This is 3% and 2%, respectively, in 40 and 50 km models in the first epoch of DOY 250 and is 6% and 2% in the first epoch of DOY 300. The issue is to what extent the voxel size can be increased while reducing modeling error and maintaining the atmosphere dynamics in each layer simultaneously. PCA has been applied here for this purpose.

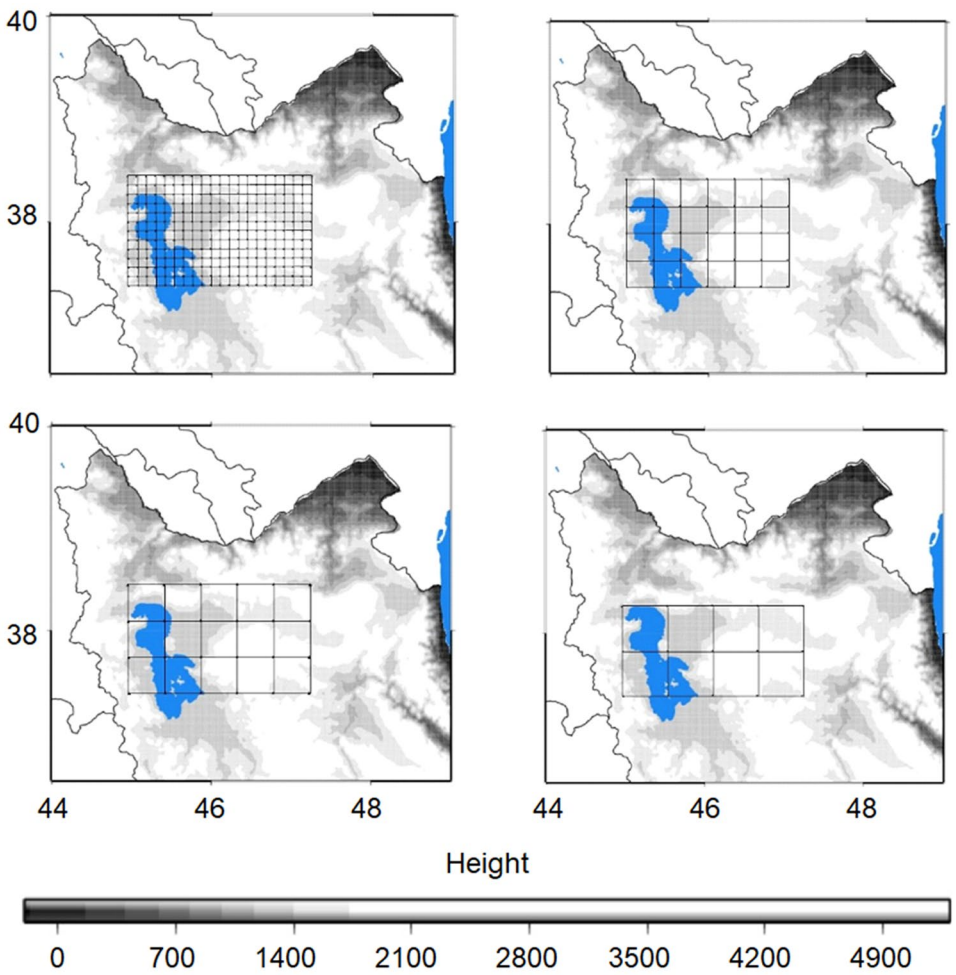

Fig. 3 Tomographic models with horizontal resolutions of 10 km (top left), 30 km (top right), 40 km (bottom left) and 50 km (bottom right)

Three tomographic models have been developed with resolutions of 30, 40, and 50 km (target models). Each target model is then compared to the initial model in terms of homogeneity of the atmospheric conditions and the resolution of the sought solutions. Figure 3 illustrates the top view of the reference and target models.

PCA in tomography models

Coordinates of the grid points were calculated in the coordinate system compatible with the WRF model. Values of water vapor pressure and temperature acquired from the WRF model were then interpolated at the positions derived in the previous step to compute the N_w values through the following equation (Kleijer 2004):

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

Here, e is the water vapor pressure, T is the air temperature in Kelvin, and $K_3 = 375463$ K^2/hpa and $K_2 = 71.2952$ K/hpa are constants (Kleijer 2004). Each voxel includes eight grid points. The average of N_w values at the voxel grid points is calculated and saved as the representative value of this parameter in the corresponding voxel.

Raster data PCA

Since this study aims to compare the horizontal resolutions of tomographic models while taking the horizontal atmospheric changes into account, voxels in the same height layer are put together in one class as the measurements of one variable. Therefore, our raster data set is a $(n \times p)$ matrix in which $p = 14$ is the number of vertical layers (variables) in tomographic models, and n is the number of units measured, i.e., the number of voxels in each layer of the models. For the horizontal resolutions of 10, 30, 40 and 50 km, $n = 240,\ 24,\ 15,\ 8$, respectively.

To analyze the variation of N_w across the proposed models, the PCA has been performed on the reference model, i.e., the models with a horizontal resolution of 10 km. When climatic changes in terms of N_w scattering with respect to the initial model is noticeable, the horizontal voxel sizes are expanded from 10 to 30, 40, and 50 km. Figures 4 and 5

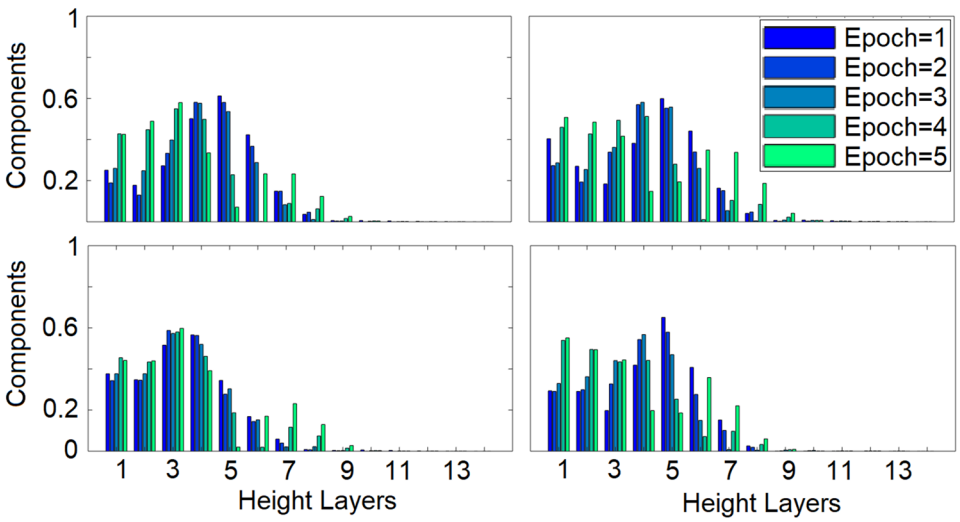

Fig. 4 First PC on DOY250 in five consecutive epochs for tomographic models: 10 km (top left), 30 km (top right), 40 km (bottom left), and 50 km (bottom right)

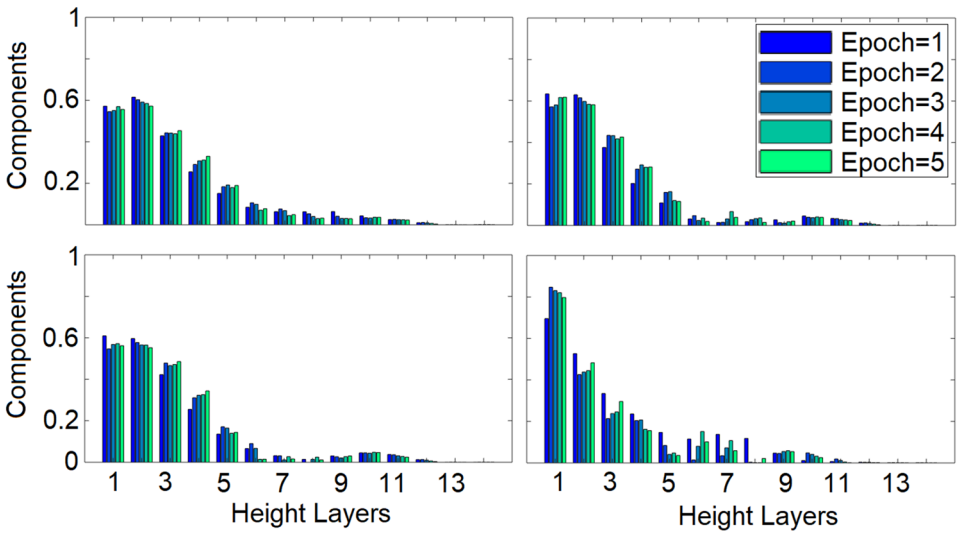

Fig. 5 First PCs on DOY 300 in five consecutive epochs for tomographic models: 10 km (top left), 30 km (top right), 40 km (bottom left), and 50 km (bottom right)

illustrate the first PC for the initial and target models in five consecutive epochs of DOY 250 and 300.

For the initial model: In DOY 250 (Fig. 4) the largest coefficients are seen in height layers 5 and 4 during the first three epochs and in height layers 3, 2 and 1during epochs 4 and 5, respectively. In DOY 300, the largest coefficients of the first PC are seen in height layers 2, 1 and 3 during the complete time of this experiment (Fig. 5). Accordingly, for the experiments, the largest horizontal changes in the atmosphere happen in the lower height layers. Therefore, increasing the voxel sizes can result in reconstructed images lacking information about the ongoing climate variations. The horizontal voxel sizes can be increased if the N_w scattering exhibit behaviors similar to the initial model. In other words, voxels can be expanded as long as climatic changes in a certain height layer can be shown properly.

According to Figs. 4 and 5, the 30 km model closely resembles the initial one, and therefore, the largest scattering in the N_w value is similar to the reference model. In this regard, and DOY 250, the 40 km model is different from the initial model as far as the order of the height layers is concerned, but in DOY 300, it resembles the reference model. There is a remarkable difference between the 50 km and the initial model on both days compared to the other models. All models illustrate that the highest scattering of the N_w parameter occurs in the lower layers. Therefore, the voxel sizes cannot be selected arbitrarily to reduce the loss of information in reconstructed tomographic images. Finally, the application of PCA to our tomographic models and during the time interval of this study shows that the largest horizontal changes in the atmosphere happen in the lower height layers. Therefore, increasing the voxel sizes in these layers

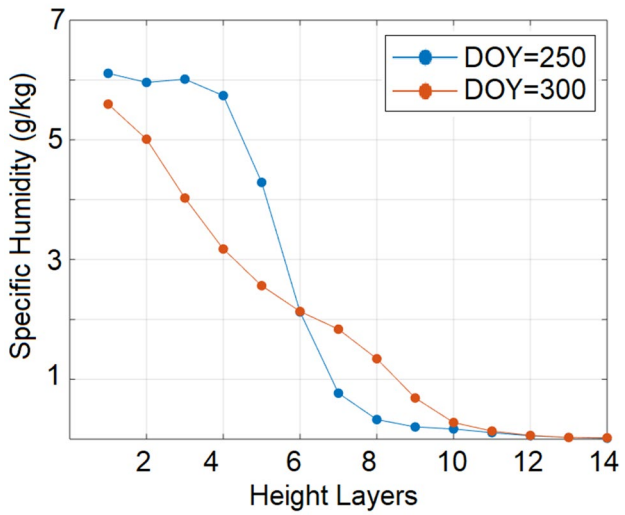

Fig. 6 Average specific humidity to the height layers of the tomographic models obtained from the WRF model in DOY 250 and 300, 2011

other in pairs or simultaneously, i.e., comparing all models at the same time. The first approach helps study the scattering of N_w in each of the target models against the initial one. In other words, the similarity of the target models to the reference model is analyzed in every layer. The extent of this general resemblance to the initial model can be estimated too. But the second approach explains which of the models under study is different from the others in terms of the scattering of N_w values and in which of the height layers this occurs. The outcome of the second approach can be used to form a hybrid tomographic model.

Two-by-two analysis of models

The two-by-two analysis approach is based on comparing each tomographic model separately with the reference model. This helps determine which height layer in a target model the scattering of N_w values shows the most resemblance with the initial model.

The coincidence angle (θ) and the total sum of the eigen-

Table 1 Coincidence angles between subspaces of dimensions 1–6

DOY	Coincidence angles in Degrees	K					
		1	2	3	4	5	6
250	Between 10 and 30 km subspaces	13.45	5.84	0.57	0.26	0.07	0.01
	Between 10 and 40 km subspaces	29.15	2.31	1.48	0.27	0.10	0.03
	Between 10 and 50 km subspaces	9.77	5.07	2.59	0.43	0.10	0.04
300	Between 10 and 30 km subspaces	9.69	2.09	1.08	0.37	0.14	0.02
	Between 10 and 40 km subspaces	4.84	4.05	0.94	0.83	0.42	0.02
	Between 10 and 50 km subspaces	11.98	6.63	3.87	0.91	0.81	0.04

is impossible, which can cause a loss of climatic changes in the reconstructed tomographic images.

However, the reason for the decline of the components of the first PC from height layer 7 upwards needs to be studied. The average specific humidity for the height layers of the tomographic models is then calculated using the WRF model at hour 0 UTC of both days (Fig. 6).

According to Fig. 6, the average specific humidity from height layer 7 upwards declined below of 2 (g/kg). Therefore, scattering in the value of N_w declined from height layer 7 upwards.

Although the computed PCs provide an overview of how each of the four models represents the climatic changes, the comparative analysis of the initial and target models inevitably reveals a horizontal resolution that keeps the atmospheric changes intact.

Inter-models comparison of PCs

The comparison of tomographic models in terms of the PCs may be made either by comparing the models against each

values are the measures that are used in this comparison. Finding appropriate dimensions (k) for the relevant subspaces is the first challenge. That is normally done by trial and error (Deměsar et al. 2013). The numerical value of k is not large in Raster PCA. Usually, small subspaces adequately represent the original space (Deměsar et al. 2013).

The smallest coincidence angle has been used to find the adequate dimensions for the subspaces under the comparison. We calculated the coincidence angle between subspaces of one to six dimensions in each of the three comparison groups of the first epoch for both days to find the appropriate dimensions. The resulting values are shown in Table 1.

According to the given results, coincidence angles are significant in all groups when the dimension of subspaces is one. Moreover, they decrease when the subspace dimension changes to 2 and 3 for both days. When the subspace dimensions are set to higher than 3, the coincidence angles reduce to less than 1 degree. This represents an increasing coincidence between the axes of the corresponding spaces. Therefore, spaces with four dimensions and more are fit for

Table 2 Coincidence angles between 4D subspaces of the initial model and target tomographic models

DOY	Coincidence angles in Degrees	Epoch				
		1	2	3	4	5
250	Between 10 and 30 km subspaces	0.26	0.14	0.35	0.33	0.17
	Between 10 and 40 km subspaces	0.27	0.23	0.44	0.69	0.17
	Between 10 and 50 km subspaces	0.43	0.49	0.44	0.87	0.98
300	Between 10 and 30 km subspaces	0.37	0.31	0.13	0.17	0.16
	Between 10 and 40 km subspaces	0.83	0.88	0.34	0.82	0.74
	Between 10 and 50 km subspaces	0.91	0.61	0.69	0.90	1.07

Table 3 Total sum of the eigenvalues of F in 4D subspaces and the average percentage of similarity for subspaces in comparison

DOY	$\sum_{i=1}^{4}\lambda_i = \sum_{i=1}^{4}\sum_{j=1}^{4}\cos^2\theta_{ij}$	Epoch					Average (%)
		1	2	3	4	5	
250	10 and 30 km spaces	3.97	3.98	3.56	3.33	3.66	92.5
	10 and 40 km spaces	3.91	3.74	3.36	3.86	3.82	93.5
	10 and 50 km spaces	3.88	3.93	3.77	2.98	3.84	92.0
300	10 and 30 km spaces	3.98	3.97	3.98	3.97	3.93	99.2
	10 and 40 km spaces	3.97	3.96	3.97	3.95	3.89	98.7
	10 and 50 km spaces	3.56	3.37	2.95	3.22	3.28	81.9

comparison. Since k must be the smallest number among the selected dimensions (Krzanowski 1979), it is concluded that the four-dimensional subspaces can be chosen as the appropriate ones for our comparison.

Using $k=4$, the coincidence angles have been calculated for each of the three groups in comparison. Corresponding results for the five epochs of DOY 250 and 300 are given in Table 2. According to these results, subspaces relevant to the 10 and 30 km models have the smallest separation angle on both days. The next smaller value belongs to the subspaces of 10 and 40 km models. The largest separation angle belongs to the 10 and 50 km models. This analysis demonstrates that surface climatic changes will not be properly maintained when the horizontal resolution of the tomographic model decreases from 10 to 50 km. This conforms to the observed largest surface climatic changes in the lower atmosphere (Figs. 4 and 5).

The total sum of the eigenvalues of **F** can be considered a criterion to explore the general similarity of the two datasets. In this study, four-dimensional subspaces have been compared with one another using this measure as well. If the sum is 4, then the two subspaces coincide completely, and if the sum is 0, the two subspaces are orthogonal. This sum has been calculated in all three comparison groups for both days. In addition, the average percentage of similarity for subspaces in comparison has been calculated. The results are shown in Table 4.

The coincidence of two subspaces implies that the two spaces are identical. According to Table 3 for DOY 250, the 40 km model, on average, maintains surface N_w changes to 93. 5%. This is 1 and 1.5% more than the 30 and 50 km models, respectively. Assuming similar results for the other epochs, small differences in the coincidence angles together with the inter-distance of GPS stations imply that the horizontal resolution of the tomographic model can be 40 or even 50 km in DOY 250. In DOY 300, the average percentage of similarity for the 30 km model is almost the same as the 40 km one, while it is 17% higher than the 50 km model. For the resolution matrix of the 30 km model, 18% of the model parameters are resolved poorly by the inverse solution. This number is 2.7% and 1.8% in DOY 250 and 6% and 2% in DOY 300 for the 40 km and 50 km models. Based on this discussion, the 40 km model is suggested as a good substitute for the reference model for both days due to the high coincidence of this model with the reference and its remarkable resolution compared to the 30 km. This model can be a good substitute for the initial model.

Simultaneous analysis of models

Here, the initial and target models are compared simultaneously. This provides a measure for analyzing the overall similarity of the model performances in reconstructing the N_w values in our test area. For this purpose, the closest vector to all subspaces is computed. Since the first principal component is in the direction of maximum changes, the results obtained from the comparison of models when $k = 1$ can expose the majority of similarities and differences. Therefore, with the consideration of $k = 1$, the closest vector

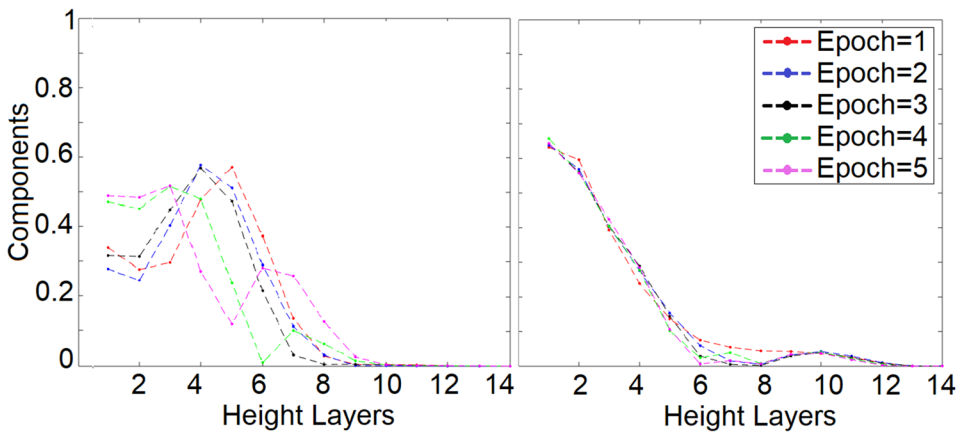

Fig. 7 Direction closest to each subspace in the first five epochs in DOY 250 (left) and DOY 300 (right), 2011. The value of the components is between zero and one

Table 4 Minimum angular separation of each subspace with the direction of the closest vector in the first five temporal epochs in DOY 250 and 300, 2011

DOY	Tomography models	Epoch					Average
		1	2	3	4	5	
250	Reference Model	2.7	1.2	1.8	1.7	2.3	1.9
	Model 30 km	1.4	2.7	1.7	3.4	1.3	2.1
	Model 40 km	3.1	2.1	4.9	2.9	3.7	3.3
	Model 50 km	3.2	4.4	4.3	3.3	6.1	4.3
300	Reference model	4.7	8.3	9.2	10.6	9.9	8.5
	Model 30 km	6.9	5.8	5.2	3.9	3	5.0
	Model 40 km	3.4	7.5	6.5	7.6	7.3	6.5
	Model 50 km	9.5	19.8	19.3	20.2	17.7	17.3

to all models in the five temporal epochs was computed for studied days. The results are shown in Fig. 7.

The largest vectors components are obtained in the layers where the scattering of N_w is maximum, i.e., the lower layers in all models. This shows that the four tomographic models share the highest scattering similarity in these layers and for both studied days. The angular separation of each vector resulting from the spaces being studied must now be calculated. The model with the lowest similarity to the other ones is determined using angular separation. For this purpose, the angle δ was calculated for each of the five epochs in the studied days. Table 4 shows the results.

According to the results in Table 4, the 50 km model has the largest average of angular separation on both days. This shows that the 50 km model has the lowest similarity to the other ones in the lower atmospheric layers. Analysis of the first PCs concludes that the highest scattering in the values N_w is in the lower layers of the atmosphere, which happened to have higher humidity. The scattering of N_w values decreases in the higher layers. Therefore, voxels with a larger size can be a homogeneous representative of the N_w in higher layers. These findings lead to the idea of forming a hybrid tomography model with smaller voxel sizes in lower layers and larger ones in the upper parts of the model.

Tomographic reconstruction of N_w

In order to check the performance of the outlined theory, N_w values have been reconstructed using three tomographic models with the horizontal resolutions of 40, 50 km, and hybrid tomographic models with the horizontal resolution of 40 km for the seven layers closer to earth and 50 km elsewhere.

N_w images are reconstructed by using each of the three models in the first epoch of both days of this experiment. To validate reconstructed images, we compared the N_w profile resulting from each of the three models with the radiosonde profile. The statistical measures of Bias, RMSE, and STD are used. Corresponding results are reported in Table 5 and also the N_w profiles are given in Figs. 8 and 9.

According to the findings outlined in Table 5, the hybrid model has the lowest value of RMSE and STD as compared to the other models. The 40 and 50 km models are in the second and third ranks, respectively, based on these measures. Moreover, the principal component comparison of the models proves that the 40 km model is most similar to the reference model. In addition, this approach to model analysis shows from the 8th layer upward that the similarity of the 50 and 40 km models to the reference

Table 5 Bias, RMSE and STD for reconstructed N_w images

DOY	Tomography Models	Bias (ppm)	RMSE (ppm)	STD (ppm)
250	40 km model	0.76	2.66	2.54
	50 km model	1.72	3.21	2.71
	Hybrid model	0.97	2.49	2.29
300	40 km model	0.61	4.20	4.15
	50 km model	2.58	5.79	5.19
	Hybrid model	1.73	2.44	1.46

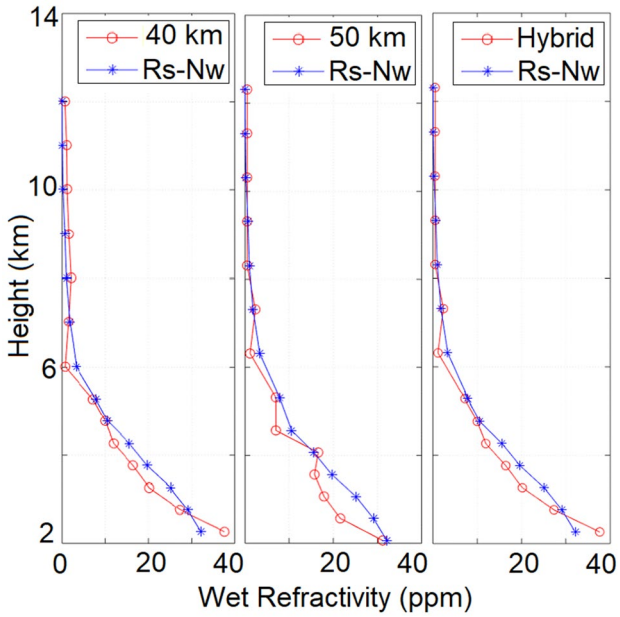

Fig. 8 Wet refractivity profiles reconstructed by tomographic models and derived from radiosonde data (non-solid versus solid circles) at the first epoch of DOY 250, 2011

model is the same. Increasing voxel sizes to 50 km in this part of the model reduces the number of unknowns by 64.

However, the hybrid model has the lowest value of RMSE and STD. Moreover, the number of unknowns is also less than the 40 km model. Therefore, the 40 km model is no more superior to the hybrid model when the RMSE, STD, and similarity measures are all considered. Considering that the N_w parameters exhibit sharp reduction from the medium to top layers; here, from 1.8 to 0.02 and 3.4 to 0.2 for the 9th to the last layer in DOY 250 and 300, respectively, and that increasing the voxel size can considerably decrease the computation time, the higher accuracy and redundancy of the hybrid model is remarkable. Moreover, for the test area of this research, there is not a remarkable difference between the precision of the two models.

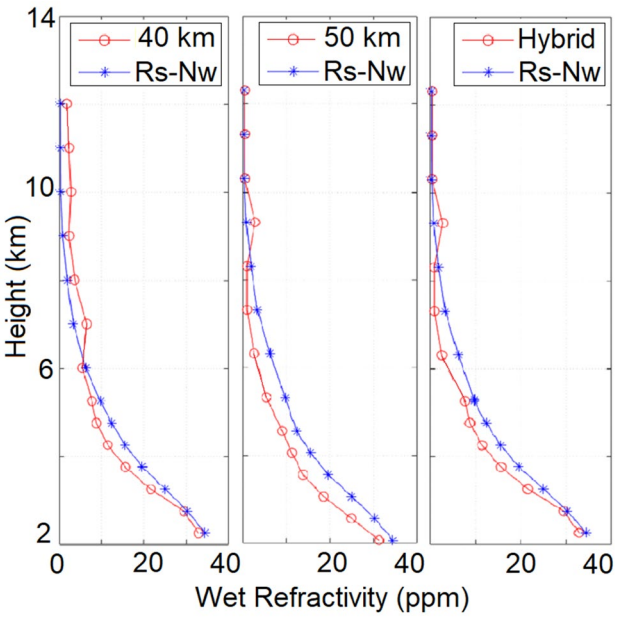

Fig. 9 Wet refractivity profiles reconstructed by tomographic models and derived from radiosonde data (non-solid versus solid circles) at the first epoch of DOY 300, 2011

Conclusion

Based on PCA, we propose a new method for determining the horizontal resolution of a GPS tomography model. The method is used for reconstructing wet refractivity in Iran. Northwest of Iran was selected to test the performance of this method because it is mountainous, and the density of GPS stations is the highest in Iran. WRF forecasts with spatiotemporal resolutions of 10 km and 1 h were used as the initial information on the atmosphere in our test area. Our experiment includes five successive epochs (totaling 5 h). Taking the model with a horizontal resolution of 10 km as the initial or reference model, the similarity analysis of this research shows that in terms of scattering in N_w parameter, the performance of the hybrid model with a horizontal resolution of 40 km at lower and 50 km at upper layers is similar to the applied reference in the test areas of this research. Considering that the number and distribution of GPS stations, i.e., the geometry of the imaging system, has a key role in the accuracy of reconstructed images, the model resolution, the angular distance of subspaces constructed by the PCs, and the bisector as the key measures have been used for deciding on the size of voxels in our experiments.

Application of the proposed method in our test areas suggests a hybrid model with the horizontal resolution of 40 km for layers 1…7 and 50 km for the remaining layers of the model, as the most accurate with reasonable precision for reconstructing N_w in our study area. The hybrid model also reduces the computation time. Therefore, its application

helps (near) real-time GNSS tomography, especially when the study area is large.

Acknowledgements We are grateful to the National Cartographic Center (NCC) of Iran for providing the observation files of the northwestern part of the Iranian Permanent GPS Network (IPGN). We particularly appreciate the meteorological organization for giving us access to radiosonde profiles with dense pressure levels at the stations of this research. We would also like to thank K. N. Toosi University for granting us permission to use the KNTU tomography model developed in the department of Geodesy, Faculty of Geodesy and Geomatics Engineering of this university.

Data availability All the 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: northwest of Iran. Meteorol Atmos Phys 126(3):193–205

Aster RC, Borchers B, Thurber CH (2018) Parameter estimation and inverse problems. Elsevier

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(10):1704–1720

Bender M, Dick G, Wickert J, Ramatschi M, Ge M, Gendt G, Rothacher M, Raabe A, Tetzlaff G (2009) Estimates of the information provided by GPS slant data observed in Germany regarding tomographic applications. J Geophys Res Atmosph, 114(D6).

Bender M, Dick G, Heise S, Zus F, Deng Z, Shangguan M, Ramatschi M, Wickert J (2013) GNSS water vapor tomography. In: Deutsches Geo Forschungs Zentrum. GFZ.

Bevis M, Businger S, Herring TA, Rocken C, Anthes RA, Ware RH (1992) GPS meteorology: Remote sensing of atmospheric water vapour using the global positioning system. J Geophys Res Atmosph 97(D14):15787–15801

Bock O, Tarniewicz J, Thom C, Pelon J (2002) The effect of inhomogeneities in the lower atmosphere on coordinates determined from GPS measurements. Phys Chem Earth, Parts a/b/c 27(4–5):323–328

Braun J, Rocken C (2003) Water vapor tomography within the planetary boundary layer using GPS. In: International workshop on GPS meteorology, Tsukuba, Japan, pp 14–17. Citeseer.

Brenot H, Walpersdorf A, Reverdy M, Van Baelen J, Ducrocq V, Champollion C, Masson F, Doerflinger E, Collard P, Giroux P (2014) A GPS network for tropospheric tomography in the framework of the Mediterranean hydro meteorological observatory Cévennes-Vivarais (southeastern France). Atmosph Measure Tech 7(2):553–578

Caldas-Alvarez A, Khodayar S, Knippertz P (2021) The impact of GPS and high-resolution radiosonde nudging on the simulation of heavy precipitation during HyMeX IOP6. Weather Climate Dyn Discuss 2(3):561–580

Champollion C, Masson F, Bouin M-N, Walpersdorf A, Doerflinger E, Bock O, Van Baelen J (2005) GPS water vapour tomography: pre-preliminary results from the escompte field experiment. Atmos Res 74(1–4):253–274

Chen K, Horton RM, Bader DA, Lesk C, Jiang L, Jones B, Zhou L, Chen X, Bi J, Kinney PL (2017) Impact of climate change on heat-related mortality in Jiangsu province, China. Environ Pollut 224:317–325

Demšar U, Harris P, Brunsdon C, Fotheringham AS, McLoone S (2013) Principal component analysis on spatial data: an overview. Ann Assoc Am Geograph 103(1):106–128

Elfving T, Nikazad T, Hansen PC (2010) Semi-convergence and relaxation parameters for a class of SIRT algorithms. Electron Trans Numer Anal 37(274):321–336

Flores A, Ruffini G, Rius A (2000) 4D tropospheric tomography using GPS slant wet delays. Ann Geophys 18:223–234

Fodor IK (2002) A survey of dimension reduction techniques. (No. UCRL-ID-148494). Lawrence Livermore National Lab., CA (US).

Guerova G (2003) Application of GPS derived water vapour for numerical weather prediction in Switzerland. Doctoral dissertation . University of Bern

Guo J, Yang F, Shi J, Xu C (2016) An optimal weighting method of global positioning system (GPS) troposphere tomography. IEEE J Sel Top Appl Earth Obs Remote Sens 9(12):5880–5887

Haji-Aghajany S, Amerian Y, Verhagen S, Rohm W, Ma H (2020a) An optimal troposphere tomography technique using the WRF model outputs and topography of the area. Remote Sens 12(9):1442. https://doi.org/10.3390/rs12091442

Haji-Aghajany S, Amerian Y, Verhagen S (2020b) B-spline function-based approach for GPS tropospheric tomography. GPS Solutions 24(3):1–12

Kak AC, Slaney M, Wang G (2002). Principles of computerized tomographic imaging.

Kleijer F (2004) Troposphere modeling and filtering for precise GPS leveling. Delft Univ. of Technol, Delft, Netherlands

Komorowski M, Marshall DC, Salciccioli JD, Crutain Y (2016) Exploratory data analysis. In: MIT Critical Data (ed.), Secondary analysis of electronic health records. pp 185–203, Springer

Krzanowski WJ (1979) Between-groups comparison of principal components. J Am Stat Assoc 74(367):703–707

Landweber L (1951) An iteration formula for Fredholm integral equations of the first kind. Am J Math 73(3):615–624

Lee LC, Liong C-Y, Jemain AA (2017) Q-mode versus r-mode principal component analysis for linear discriminant analysis (LDA). In: AIP conference proceedings, vol 1842, pp 030024. AIP Publishing LLC.

Lutz SL (2008) High-resolution GPS tomography in view of hydrological hazard assessment. Ph.D. thesis, ETH Zurich.

Rohm W (2013) The ground GNSS tomography–unconstrained approach. Adv Space Res 51(3):501–513

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

Troller M, Burki B, Cocard M, Geiger A, Kahle H-G (2002) 3-d refractivity field from GPS double difference tomography. Geophys Res Lett 29(24):2–1

Troller MR (2004) GPS based determination of the integrated and spatially distributed water vapor in the troposphere. Ph.D. thesis, ETH Zurich.

Wang X, Dessler AE (2020) The response of stratospheric water vapor to climate change driven by different forcing agents. Atmos Chem Phys 20(21):13267–13282

Xia P, Cai C, Liu Z (2013) GNSS troposphere tomography based on two- step reconstructions using GPS observations and cosmic profiles. Ann Geophys 31:1805–1815 (**Copernicus GmbH**)

Yang F, Guo J, Shi J, Zhou L, Xu Y, Chen M (2018) A method to improve the distribution of observations in GNSS water vapor tomography. Sensors 18(8):2526

Yao Y, Zhao Q (2016) Maximally using GPS observation for water vapor tomography. IEEE Trans Geosci Remote Sens 54(12):7185–7196

Yao Y, Zhao Q (2017) A novel, optimized approach of voxel division for water vapor tomography. Meteorol Atmos Phys 129(1):57–70

Zhang W, Zhang S, Ding N, Zhao Q (2020) A tropospheric tomography method with a novel height factor model including two parts: Isotropic and anisotropic height factors. Remote Sensing 12(11):1848
Zhao Q, Yao Y, Yao W (2018) troposphere water vapor tomography: a horizontal parameterized approach. Remote Sens 10(8):1241

Elaheh Sadeghi received her M.Sc. degree in Geodesy at K. N. Toosi University of Technology in 2014, and she is a Ph.D. candidate of Geodesy in the School of Surveying and Geospatial Engineering, College of Engineering, University of Tehran. Her research interests are GNSS meteorology and satellite Geodesy and signal processing.

Masoud Mashhadi Hossainali received his Ph.D. degree from the Darmstadt University of Technology and is an associate professor in Geodesy and Dean of Faculty of Geodesy and Geomatics Engineering at K. N. Toosi University of Technology. His professional interests are deformation modeling and analysis, satellite Geodesy and signal processing.

Abdolreza Saferi is a Professor at the University of Tehran. He received his Ph.D. in geodesy from the School of Surveying and Geospatial Engineering, College of Engineering, University of Tehran. His professional interests are the mathematical modeling of remote sensing and geodetic data.

Terms and Conditions

Springer Nature journal content, brought to you courtesy of Springer Nature Customer Service Center GmbH (“Springer Nature”).
Springer Nature supports a reasonable amount of sharing of research papers by authors, subscribers and authorised users (“Users”), for small-scale personal, non-commercial use provided that all copyright, trade and service marks and other proprietary notices are maintained. By accessing, sharing, receiving or otherwise using the Springer Nature journal content you agree to these terms of use (“Terms”). For these purposes, Springer Nature considers academic use (by researchers and students) to be non-commercial.
These Terms are supplementary and will apply in addition to any applicable website terms and conditions, a relevant site licence or a personal subscription. These Terms will prevail over any conflict or ambiguity with regards to the relevant terms, a site licence or a personal subscription (to the extent of the conflict or ambiguity only). For Creative Commons-licensed articles, the terms of the Creative Commons license used will apply.
We collect and use personal data to provide access to the Springer Nature journal content. We may also use these personal data internally within ResearchGate and Springer Nature and as agreed share it, in an anonymised way, for purposes of tracking, analysis and reporting. We will not otherwise disclose your personal data outside the ResearchGate or the Springer Nature group of companies unless we have your permission as detailed in the Privacy Policy.
While Users may use the Springer Nature journal content for small scale, personal non-commercial use, it is important to note that Users may not:

1. use such content for the purpose of providing other users with access on a regular or large scale basis or as a means to circumvent access control;
2. use such content where to do so would be considered a criminal or statutory offence in any jurisdiction, or gives rise to civil liability, or is otherwise unlawful;
3. falsely or misleadingly imply or suggest endorsement, approval , sponsorship, or association unless explicitly agreed to by Springer Nature in writing;
4. use bots or other automated methods to access the content or redirect messages
5. override any security feature or exclusionary protocol; or
6. share the content in order to create substitute for Springer Nature products or services or a systematic database of Springer Nature journal content.

In line with the restriction against commercial use, Springer Nature does not permit the creation of a product or service that creates revenue, royalties, rent or income from our content or its inclusion as part of a paid for service or for other commercial gain. Springer Nature journal content cannot be used for inter-library loans and librarians may not upload Springer Nature journal content on a large scale into their, or any other, institutional repository.
These terms of use are reviewed regularly and may be amended at any time. Springer Nature is not obligated to publish any information or content on this website and may remove it or features or functionality at our sole discretion, at any time with or without notice. Springer Nature may revoke this licence to you at any time and remove access to any copies of the Springer Nature journal content which have been saved.
To the fullest extent permitted by law, Springer Nature makes no warranties, representations or guarantees to Users, either express or implied with respect to the Springer nature journal content and all parties disclaim and waive any implied warranties or warranties imposed by law, including merchantability or fitness for any particular purpose.
Please note that these rights do not automatically extend to content, data or other material published by Springer Nature that may be licensed from third parties.
If you would like to use or distribute our Springer Nature journal content to a wider audience or on a regular basis or in any other manner not expressly permitted by these Terms, please contact Springer Nature at

onlineservice@springernature.com