Investigation of Height-Dependent Systematic Component of ZTDs Using Spherical Harmonic Functions (SHF) and Empirical Orthogonal Functions (EOF)

İlke DENİZ, Turkey

Key words: GNSS/GPS, positioning, tropospheric zenith delay, spherical harmonic, empirical orthogonal functions

SUMMARY

Spherical harmonic (SH) is one of the tools used in modelling scalar fields on the spherical surface. Employing SH doesn't cause a problem in the matrix calculations of global and regional areas. However, the (weighted) normal equations' matrix of the least squares solution becomes ill-conditioned in local areas of several degrees. Apart from the density and distribution of data, noise and blunders affect the results of SH modeling. For unique and stable solutions, it is recommended to use the smoothed or filtered data.

Empirical Orthogonal Function (EOF) analysis is a statistical tool utilized in the determination of spatial and temporal variation in a physical field, the separation of the signal from the noise, the prediction and filtering of data.

Tropospheric Zenith Delay (ZTD) time series derived from continuous GNSS stations (CORS) depends on the topography of the field (position) and the meteorological parameters. Location-dependent systematic effects in ZTDs, and especially the effect of station heights, can be determined by EOF analysis with high precision.

In this study, hourly ZTD time series are derived at 16 TUSAGA-Active (Turkish CORS) stations (at distances of 80-100 km) for 20 days in a test area limited to 30°–34° northern latitudes and 39°–42° eastern longitudes. The precision of the least squares SHF modelling and the SHF modelling of the results obtained from EOF analysis (the height-dependent systematic components (PCs) and the reconstructed data (FRs) from these PCs) are investigated. It is shown that ZTDs depend on height. SHF modelling was carried out using 16 stations for every hour. SHF modelling was employed to these stations for the arithmetic mean of standard deviation of total of 456 models M_{avg} different radius vectors, PC_1 (first Principal Component) and FR_{12} (the constructed data from 12 PC). Whether the ZTDs contain a height-dependent systematic component and their effects on the precision of 3D SHF modelling are verified by the values of M_{avg} For unique and stable solutions, the radius of the Earth is taken as r=6375.0 km+hkm and the arithmetic average of standard deviation is found to be M_{avg}=±8.5 mm.

Investigation of Height-Dependent Systematic Component of ZTDs Using Spherical Harmonic Functions (SHF) and Empirical Orthogonal Functions (EOF)

İlke DENİZ, Turkey

1. INTRODUCTION

Continuous GNSS networks (CORS) already cover large areas and is rapidly growing. Long-term Total Zenith Delay (ZTD) time series collected from these networks enables the analysis of temporal and spatial variations of ZTDs.
In recent years, the investigation of height-dependent systematic component of ZTDs (Schüler, 2001, Steigenberger, 2009, Qian, 2016), the precision of ZTD (Shrestha, 2003, Haase et al., 2003), and the modelling and interpolation of ZTDs on global, continental or local scales (Zus et al., 2019) have become popular.
CORS stations have been equipped with meteorological sensors and converted to GNSS-MET stations to estimate near real-time water vapor. Similarly, these network for water vapor prediction can be intensified by interpolating ZTD values to meteorological stations around CORS stations.
Spherical harmonic functions (SHF) are the most natural tools in the modeling of scalar fields outside the spherical surface. Therefore, SHF are used for the modelling of physical processes such as gravity, electricity and heat transfer (Brett, 1988, Barthelmes and Köhler, 2016) and scalar quantities such as temperature, pressure, ZTD and mapping function (Böhm et al., 2007, Fritsche et al., 2012, Yao et al., 2015).
SHF do not cause a problem in the least squares computations for global and continental areas, whereas they cause ill-conditioned linear problems in small areas of several degrees. Different algorithms have been developed for solving these types of problems. In this context, one of the following methods can be selected (Brett, 1988, Hansen, 1994, Yagle, 2005, Meurant, 2010, Voss and Lampe, 2011, Xiang and Zou, 2013): singular value decomposition, sensitivity and conditioning solution of linear system of equations, regularizations, iterative solution of linear systems of equations.
Empirical Orthogonal Function (EOF) analysis is a vital mathematical and statistical tool for the study of spatial and temporal variation in a physical field. This method seeks patterns with the largest variance in a spatial-temporal dataset. It is used for analyzing the relationship between variables in the dataset and the characteristics of this relation. Due to these features of EOF analysis, it has been used in meteorology since the late 1940s. EOF analysis obtains special orthogonal patterns or principal components (PC) of an uncorrelated time series in any space-time meteorological field. In EOF analysis, it is possible to generate a new version of dataset by using the largest PCs in the dataset. Due to this feature, it is used for many applications such as data compression, data prediction, filtering, and factor analysis. Moreover, it is an important tool for separating the signal from the noise in the dataset.
EOF analysis is a widely used method because of its analytic derivation being simple. However, the orthogonality of EOFs in space and time is a debatable issue. There is no orthogonality in

physical phenomenon. To overcome this issue, linear transformations of EOFs based on rotation is used (rotated empirical orthogonal functions (REOFs)). REOFs enhance the physical interpretation of the field (North et al., 1982, Preisendorfer, 1988, Hannachi A., 2004, Hannachi et al., 2007, Wilks, 2011).

In this study, a test area is limited 39.0°–42.0° northern latitudes and 30.5°–35.5° eastern longitudes comprising of 16 Turkish Continuously Operated Reference Stations (TUSAGA-Active) spacing of 80 - 100 km. 20 days of hourly ZTD time series derived from these stations are evaluated by EOF analysis. It is aimed to investigate whether the ZTDs contain a height-dependent systematic component, the precision and filtering of ZTDs, to model the filtered and the original ZTD time series using SHF and the precision of the modeling.

2. COMPUTATION OF EOFs

A full mathematical derivation of EOFs is given in references (Preisendorfer, 1988, Hannachi et al., 2007, Von Storch, and Navarra, 2013). Here, the computation of EOFs using MATLAB will be given.

Data matrix M is formed with the dimension of txp where p denotes the number of stations, and t is the number of time series. Trends are removed from the dataset:

$$F_{txp} = \text{detrend(M,0)} \tag{1}$$

The covariance matrix of F is formed:

$$R_{pxp} = (F' * F) / N \tag{2}$$

where N is the number of observations.

The eigenvalues and eigenvectors of the covariance matrix R is computed:

$$[C, L] = eig(R) \tag{3}$$

where C_{pxp} is the eigenvector matrix and L_{pxp} the diagonal eigenvalues matrix.

If EOFs are required to be dimensional, the eigenvector matrix of C is normalized:

$$C = normc(C) \tag{4}$$

Hence, EOFs or MODs are computed for each station:

$$\begin{aligned} EOF_{1pxr} &= C(:,p) \\ EOF_{2pxr} &= C(:,p-1) \\ &\vdots \end{aligned} \tag{5}$$

The PCs for each station are obtained:

$$PC_{i_{tx1}} = F * EOF_i \qquad \text{i=1,...,p-1} \tag{6}$$

The amount of variance defined by each eigenvalue is computed:

$$\lambda = diag(L) / trace(L) \tag{7}$$

λ values are sorted in decreasing order. $\lambda(i)$ is paired with the variance of EOF_i.

One of the most important characteristic of EOF is to reconstruct the data matrix using PCs. The data matrix is reconstructed (FR) by;

$$\begin{aligned} FR_1 &= PC_1 * EOF_1' \\ FR_2 &= [PC_1 \; PC_2] * [EOF_1 \; EOF_2]' \\ &\vdots \end{aligned} \tag{8}$$

and in general by the sum of the modes

$$FR_{i_{txp}} = A * E' \tag{9}$$

where A= [PC_1 PC_2 ..] and E=[EOF_1 EOF_2 ...].

The statistical significance of the computed EOF can be determined by North's rule of thumb (North et al., 1982). This rule is chosen for its being simple, useful as well as powerful (Björnsson and Venegas, 1997). According to this rule, the largest EOF_1 is assumed to be significant.

$$\Delta\lambda_k \approx \sqrt{\frac{2}{N}}\lambda_k \text{ (test value)} \tag{10}$$

$$\Delta\lambda = \lambda_j - \lambda_k \tag{11}$$

EOFs corresponding to the variance percentages (λ) at the lower levels are used in Eqs. 10-11. If $\Delta\lambda > \Delta\lambda_k$ (hypothesis), λ_j; EOF corresponding to the percentage of variance that comes after λ_k and the corresponding PC is found to be significant. Here; N is the degree of freedom, λ_k is the first variance. According to the rule, the significance of λ at the lower levels is determined gradually depending on significant λ.

In the case of reconstruction of the data matrix by Eq. 9, the residual v for every reconstruction of the data matrix can be calculated:

$$\mathrm{v}_i = F - FR_i \quad \text{i=1,...,p-1} \tag{12}$$

where F is the observation matrix and FR is the reconstruction matrix.

Root mean square error (RMSE) is computed from the residuals:

$$[\mathrm{v}_i \mathrm{v}_i] = v^T v \tag{13}$$

$$m_{o_i} = \sqrt{[\mathrm{vv}]/(N-i)} \tag{14}$$

where N is the number of observations.

3. SPHERICAL HARMONICS FUNCTIONS MODELLING

In general, three-dimensional spherical harmonics function outside the spherical surface with a curvature radius R is given;

$$f(r,\varphi,\lambda) = \sum_{n=0}^{\infty} \left(\frac{R}{r}\right)^{n+1} \sum_{m=0}^{n} P_{nm} \sin(\varphi) \left[A_{nm}.\cos(m\lambda) + B_{nm}.\sin(m\lambda) \right] \tag{15}$$

where n; degree, m; order, r; radius vector, φ; spherical latitude, λ; spherical longitude and $P_{nm}\sin(\varphi)$; Legendre function. If R/r is taken as 1, the equation becomes two dimensional.

The least squares method for modelling the scalar quantities is given;

$$a_{nm} = \left(\frac{R}{r}\right)^{n+1} P_{nm}(\sin\varphi).\cos(m\lambda) \tag{16}$$

$$b_{nm} = \left(\frac{R}{r}\right)^{n+1} P_{nm}(\sin\varphi).\sin(m\lambda) \tag{17}$$

Eq.16 and Eq.17 are expanded to degree n and order m and the coefficients of the design matrix A is obtained.

$$A = [a_{00} \;\; a_{10} \;\; a_{11} \;\; b_{11} \;\; a_{20} \;\; a_{21} ...] \tag{18}$$

Hence the coefficients of A_{nm} and B_{nm} are computed by the least squares method:

$$x = \left(A^T A\right)^{-1} A^T l \tag{19}$$

where l is the matrix of observations.

Residuals of the observations is computed:

$$v = Ax - l \tag{20}$$

The mean square error of an observation is computed:

$$m_0 = \pm\sqrt{\frac{[vv]}{n-u}} \tag{21}$$

where n; number of observations, u; number of unknowns. The design matrix A is positive definite, provided that n>u. However, since the coefficients a and b attain values close to each other for smaller areas, the normal equations' matrix N (A^TA) can be ill-conditioned. So Generalized Tikhonov Regularization algorithm can be used for solving ill-conditioned linear problem as in Eq. 22 in the least squares estimation, (Moszynski, 1995, Gkioulekus, 1996, Matlab Central, 2015, Tikhonov regularization, 2015).

$$x = \left(A^T A + \lambda^2 B^T B\right)^{-1} A^T l \tag{22}$$

where λ is the regularization parameter and $B = \begin{bmatrix} 1 & 0 & \dots \\ -1 & 1 & \dots \\ \vdots & \vdots & \vdots \end{bmatrix}$ (Yagle, 2005).

For the optimal solution, m_0s are calculated by replacing λ parameter. The graphic of the L-curve is drawn with log λ and log m_0 axes. The λ corresponding to the corner of the curve (the point of maximum curvature) will give the optimal solution (Hansen, 1994, Voss and Lampe, 2011).

4. STUDY AREA AND DATA

ZTD time series derived from 16 TUSAGA-Active stations in a test area limited to 30°–34° northern latitudes and 39°–42° eastern longitudes are used in this study. 20 days of hourly ZTD values are obtained by processing GNSS observations of these stations between 29 June and 29 July 2017 using GAMIT/GLOBK software.

4.1 PCS and Precision of EOF Analysis

Systematic PCs of 20 days of hourly ZTD time series are computed by Eqs. 1-7. To test whether these PCs are significant, test data is derived by Eq. 10 and 11.
According to North's rule of thumb (North et al., 1982), the largest EOF_1 is assumed to be significant. The significance of other EOFs is examined by utilizing the differences between variance percentages (λ) of eigenvalues of consecutive PCs (λ_i-λ_{i+1}). The variance percentages

of PCs (λ_i), the variance percentage differences (λ_i-λ_{i+1}) and the test value $\Delta\lambda$ are given in Table 1.

Table 1 The results of EOF analysis (Deniz, 2019)

% λ_i i=1,...,p-1	λ_i-λ_{i+1}	$\Delta\lambda = \lambda_{i+1}\sqrt{\frac{2}{N}}$	m_{o_i} (mm) (1σ)
0.7723	0.7723	0.6752	17.2535
0.0971	0.6752	0.1023	13.0669
0.0327	0.0644	0.0129	11.3157
0.0249	0.0078	0.0043	9.7731
0.0149	0.0100	0.0033	8.7185
0.0111	0.0038	0.0020	7.8434
0.0100	0.0011	0.0015	6.9594
0.0076	0.0024	0.0013	6.2083
0.0070	0.0006	0.0010	5.4240
0.0060	0.0010	0.0009	4.6449
0.0040	0.0020	0.0008	4.0462
0.0032	0.0007	0.0005	3.4854
0.0029	0.0004	0.0004	2.9010
0.0026	0.0003	0.0004	2.2386
0.0023	0.0003	0.0003	1.4164
0.0015	0.0008	0.0003	1.4165

Time series can be reconstructed from the computed PCs using Eq. 8 and 9. Thus, $FR_1=PC_1$, $FR_2=PC_1+PC_2$, $FR_3=PC_1+PC_2+PC_3$,.... are reconstructed. The mean square error values (m_{o_i}) are computed using Eqs. 12-14 by comparing each reconstructed time series to the original time series F. These values are given in the m_{o_i} column of Table 1.

The calculated PCs may not have any physical properties. Therefore, they must be examined. Here, the largest component is examined.

1. day and 1. hour values of FR_1 computed from PC_1 is visualized in Fig. 1b. When the topography of the test network (Fig. 1a) is compared with Fig. 1b, it is observed that FR_1 values (negative values) also increase in parallel with the increase in height when moving away from the shore.

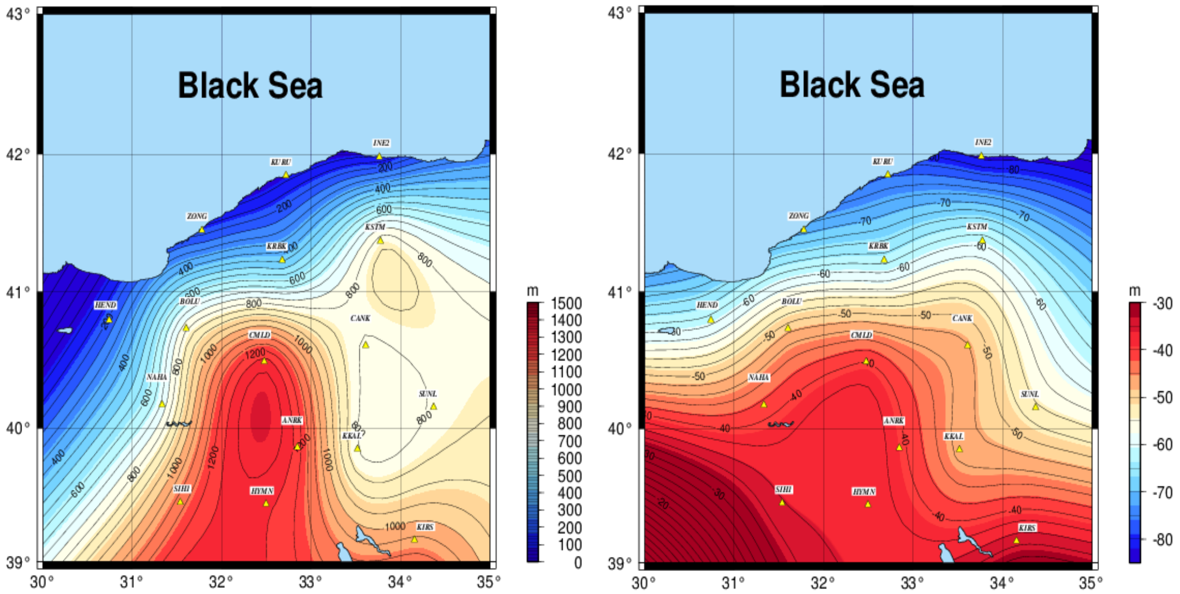

Figure 1: (a) Topography of the test network (b) 1. day and 1. hour values of FR_1

Fig. 2 shows that there is a high correlation between FR_1 values and the height of stations. Table 1 shows that PC_7 and PC_9 are not significant. However, PC_8, PC_{10} and the rest are significant. Here, the sum of these significant components FR_{12}, calculated by Eq. 9, will be the filtered original F data. FR_{12} is the sum of systematic components of F data. The differences F- FR_{12} will give the noise in the F data. Thus, the value of $m_0 = \pm$ 3.5 mm calculated for FR_{12} can be taken as the root mean square error of 1 ZTD observation.

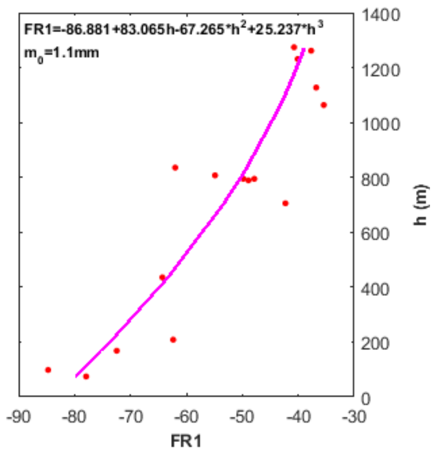

Figure 2: Height-FR_1 diagram

4.2 SHF Modeling

In the least squares estimation using Eqs. 15-21, ill-conditioned linear problem occurs because the test area is small. Therefore, Generalized Tikhonov Regularization algorithm (GTRA) given in Eq. 22 is employed to the data.
The most optimal regularization parameter λ is chosen with L-curve criteria by the expansion of 3rd degree of SHF (n=m=3, 10 unknowns). The graphic of L-curve is shown in Fig. 3.

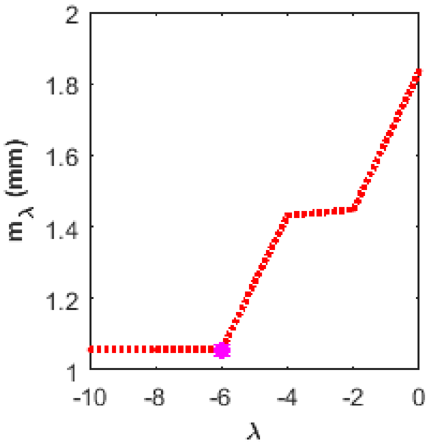

Figure 3: The graphic of L-curve

Consequently, the most optimal regularization parameter $\lambda = 1.10^{-6}$.
Whether these unknowns are significant has been tested empirically by adding or removing unknowns. In addition, n=m=3 expansion is found to be the optimal expansion for modeling in the test area.
The original matrix F and the reconstructed data FR_{12} are used in SHF modelling. The modelling becomes two dimensional if r is taken as 1, whereas it becomes three dimensional if $r = \frac{R}{R+h}$. Where R=6375.0 km (Gauss curvature radius for the test network) and h is the ellipsoidal height of the stations.
SHF modelling is employed separately for each hour (total of 480 solutions) and the arithmetic average of m_0 values is taken. The arithmetic average of precision (m_0) is given in Table 2.

Table 2 The arithmetic average of precision

Data	**2D** **r=1.0** **m_0 (mm)**	**3D** $\boldsymbol{r = \frac{R}{R+h}}$ **m_0 (mm)**
F (original)	60.60	12.11
FR_{12} (filtered)	38.61	8.54

It is apparent from Table 2 that 3D SHF modelling yields better results. This result can also be interpreted as the evidence of the station height effects in ZTDs.
The filtered FR_{12} matrix from 3D SHF modelling have produced the precision of ± 8.54 mm. Considering that the test network is 80-100 km apart, this result needs to be evaluated. In denser networks, a smaller precision can be expected.

5. CONCLUSIONS AND DISCUSSIONS

EOF analysis can be used for separating the signal and the noise in time series. In this study, ZTD time series are filtered by separating the signal and the noise, and reconstructing the series. The filtering has increased the precision of SHF modeling. The noise in ZTD time series is found to be ± 3.5 mm. It has been proven that ZTD time series in the test area contain height-dependent systematic component.
In SHF modeling of ZTD, 3D modelling yields better results than 2D modelling. This result also shows that ZTD contain height-dependent systematic component. ZTD time series in the test network spacing of 80 - 100 km can be modelled with ± 8.5 mm precision.
The height-dependent systematic components in ZTDs derived by EOF analysis and SHF should be investigated with long-term series in different networks. The scaling of these components obtained by EOF analysis should also be examined.

REFERENCES

Barthelmes, F. and Köhler, W., 2016, International Centre for Global Earth Models (ICGEM), in: Drewes, H., Kuglitsch, F., Adám, J. et al., The Geodesists Handbook 2016, Journal of Geodesy (2016), 90(10), pp 907-1205,doi: 10.1007/s00190-016-0948-z

Björnsson, H., Venegas S. A., 1997, A Manual for EOF and SVD Analyses of Climatic Data, CCGCR Report 97.1., McGill University, Montreal, Canada, pp112-134

Böhm, J., Heinkelmann, R., Schuh, H., 2007, Short note: a global model of pressure and temperature for geodetic applications, Journal of Geodesy, 81(10), 679-683

Brett, L. I. N. D. Y., 1988, Methods of spherical harmonic analysis. Quarterly Journal of the Royal Astronomical Society, 29, 129

Deniz, I., 2019, An empirical orthogonal function (EOF) analysis of Tropospheric Zenith Delay (ZTD), 7th International Colloquium on Scientific Fundamental Aspects of GNSS, 04-06.09.2019, Switzerland

Fritsche, M., Döll, P., Dietrich, R., 2012, Global-scale validation of model-based load deformation of the Earth's crust from continental watermass and atmospheric pressure variations using GPS, Journal of Geodynamics, 59, 133-142

Gkioulekus, E., 1996, Algorithms with Matlab, Universtity of Washington

Haase, J., Ge, M., Vedel, H., Calais, E., 2003, Accuracy and variability of GPS tropospheric delay measurements of water vapor in the western Mediterranean, Journal of Applied Meteorology, 42(11), 1547-1568

Hannachi A., 2004, A Primer for EOF Analysis of Climate Data. Department of Meteorology, University of Reading, pp1-33

Hannachi, A., Jolliffe, I. T., Stephenson, D. B., 2007, Empirical Orthogonal Functions and Related Techniques in Atmospheric Science: A Review, Int. J. Climatol., 27(9), 1119-1152

Hansen, P. C., 1994, Regularization tools: a Matlab package for analysis and solution of discrete ill-posed problems, Numerical algorithms, 6(1), 1-35

Matlab Central, 2015, Online at http://www.mathworks.com/matlabcentral (as of 2015)

Meurant, G., 2010, Solving discrete ill-posed problems with Tikhonov regularization and generalized cross validation, Plzen University, Czech Republic

Moszynski, K., 1995, On Solving Linear Algebraic Equations with an Ill-conditioned Matrix, Applicationes Mathematicae, 22(4), 499-513

North, G. R., Bell, T. L., Cahalan, R. F., Moeng, F. J., 1982, Sampling Errors in the Estimation of Empirical Orthogonal Functions, Mon. Weather Rev., 110(7), 699-706

Preisendorfer, R. W., 1988, Principal Component Analysis in Meteorology and Oceanography, 1-425, Amsterdam, Elsevier Science Publishers

Qian, C., 2016, Tropospheric correction modeling in SAPOS reference network under large height difference condition, MSc Thesis, Munich Technical University

Schüler, T., 2001, On ground-based GPS tropospheric delay estimation, PhD Thesis , Munich Technical University

Shrestha, S. M., 2003, Investigations Into the Estimation of Tropospheric Delay and Wet Refractivity Using GPS Measurements, MSc Thesis, University of Calgary, Department of Geomatics Engineering

Steigenberger, P., 2009, Reprocessing of a global GPS network, PhD Thesis, Munich Technical University

Tikhonov regularization, 2015, Online at http://en.wikipedia.org/wiki/Tikhonov-regularization (as of 2015)

Von Storch, H., Navarra, A. (Eds.), 2013, Analysis of Climate Variability: Applications of statistical techniques proceedings of an autumn school organized by the Commission of the European Community on Elba from October 30 to November 6, 1993, Springer Science & Business Media

Voss, H., Lampe, J., 2011, Regularization of Large Scale Total Least Squares Problems, Communications in Computer and Information Science. 245. 22-37. 10.1007/978-3-642-27245-5_6

Wilks, D. S., 2011, Statistical Methods in the Atmospheric Sciences, International Geophysics Series, Vol. 100. Elsevier Academic Press, Oxford, UK, pp1-676

Xiang, H., Zou, J., 2013, Regularization with randomized SVD for large-scale discrete inverse problems, Inverse Problems, 29(8), 085008

Yagle, A. E., 2005, Regularized matrix computations, Department of EECS, The University of Michigan

Yao, Y., Xu, C., Shi, J., Cao, N., Zhang, B., Yang, J., 2015, ITG: A new global GNSS tropospheric correction model, Scientific reports, 5(1), 1-9

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 Sensing, 11(6), 674

BIOGRAPHICAL NOTES

İlke Deniz is an Assistant Professor in the Department of Geomatics Engineering at Zonguldak Bülent Ecevit University, Turkey. She graduated from Yıldız Technical University in 2004 as Geodesy and Photogrammetry Engineer. She obtained M.Sc. degree from Geodesy Department of Boğaziçi University and PhD degree from Geomatics Engineering of Bülent Ecevit University. Her research interests are GNSS, tropospheric parameter estimation by academic processing softwares, tropospheric delay, GPS Meteorology and strain analysis.

CONTACTS

Asst. Prof. Dr. İlke Deniz
Zonguldak Bülent Ecevit University
Engineering Faculty, Geomatics Eng. Department
67100, Zonguldak
TURKEY
Tel. +903722912576
Email: ideniz@beun.edu.tr
Web site: https://akademikcv.beun.edu.tr/cv/ideniz.html