GLOBAL POSITIONING SYSTEM PRECIPITABLE WATER VAPOR INTERPOLATION USING INVERSE MULTIQUADRIC, ARTIFICIAL NEURAL NETWORK AND INVERSE DISTANCE WEIGHTED

Sh. Chamankar [1]*, Y. Amerian [1], Sh. Naderi Salim [1]

[1] Faculty of Geodesy and Geomatics Engineering, K. N. Toosi University of Technology, Tehran, Iran
chamankar@email.kntu.ac.ir, amerian@kntu.ac.ir, shervin.naderi@email.kntu.ac.ir

Commission IV, WG IV/3

KEY WORDS: Global Positioning System, Precipitable Water Vapor, Artificial Neural Network, Inverse Multiquadric, Inverse Distance Weighted

ABSTRACT:

Precipitable water vapor (PWV) is one of the most critical data in many meteorological departments. This component has great spatial and temporal changes, so the global positioning system (GPS) always seeks to increase the accuracy of estimating the water vapor component in the troposphere. The waves sent by the satellites of this system are delayed due to passing through atmospheric layers such as the troposphere. In this paper, interpolation methods are used to estimate precipitable water vapor. Inverse multiquadric (IMQ) interpolation which is based on radial basis functions, artificial neural network (ANN) method, and inverse distance weighted (IDW) which are the most common method of interpolation in meteorology. A region in North America with 23 GPS stations was randomly selected. Then, the interpolation of precipitable water vapor on a summer day is done using GPS data. The root mean square error value (RMSE) for the IMQ method was the lowest compared to other methods and was equal to 2.11 mm. Finally, using the IMQ interpolation method, a dense map of Precipitable water vapor changes in the troposphere layer is developed for the study area.

1. INTRODUCTION

With the advent of GNSS and advancements in geoscience studies and weather forecasting, further analysis of the atmosphere and its parameters has become particularly important. Furthermore, it has caused the emergence of different methods to achieve higher accuracy for estimating atmospheric parameters (Yang et al., 2019). The troposphere layer is essential in atmospheric studies, for the reason that GNSS signals reach the ground in a bent and delayed manner due to refraction caused by the troposphere layer (Nilsson et al., 2013). The tropospheric delay of the GNSS signal consists of dry and wet parts. The prediction of the zenith wet delay and the interpolation of its parameters is difficult due to the rapid temporal and spatial change of water vapor compared to the dry part (Smith and Weintraub, 1953). besides, the parameter of PWV is a function of ZWD and has many uses in weather forecasting programs (Chen et al., 2018). using interpolation, the desired parameters in the troposphere can be obtained. In general, interpolation is a process by which the value of a quantity at points with known coordinates can be obtained using the value of the same quantity at other points (Davis, 1975). Therefore, to get the amount of PWV that can be precipitated in the troposphere layer, we can use the interpolation methods such as inverse multiquadric, artificial neural network, and inverse distance weighted. To do that it is necessary to have the PWV value in the desired reference stations. Therefore, first, the tropospheric delay in the zenith direction should be calculated using the precise point positioning method (PPP) from the GPS observation. To calculate ZHD, pressure observations are needed, which can be obtained from equation (1) by using the dry experimental model of saastamoinen, which is one of the most accurate global prediction models (Mendes, 1999, Saastamoinen, 1972).

$$ZHD = \frac{0.002277 \times P}{1 - 0.00266 \times \cos(2\varphi) - 0.00000028 \times h} \quad (1)$$

where P is surface pressure (hPa), φ is latitude and, h is altitude (m). The value of the ZWD is calculated using equation (2) (Bevis et al., 1992):

$$ZWD = ZTD - ZHD \quad (2)$$

Then, PWV is obtained from the following equation:

$$PWV = \frac{10^6}{\rho_w \times R \times \left(\frac{k_3}{T_m} + k_2\right)} \times ZWD \quad (3)$$

where $R = 0.4613 J.g^{-1}.K^{-1}$

$k_2 = 22.1 \pm 2.2 K.hPa^{-1}$

$k_3 = (3.739 \pm 0.012) \times 10^5 K^2.hPa^{-1}$

besides ρ_w is water density and the parameters R, k_2 and k_3 are constants. T_m is the atmospheric weighted average temperature. Using the Bevis formula do that as follows (Bevis et al., 1992):

$$T_m = 85.63 + 0.668 T_s \quad (4)$$

---

* Corresponding author

where T_s is the surface temperature is in Kelvin at the location.

2. METHODS

2.1 Radial basis functions

The radial basis function (RBF) is a universal interpolation method (Leyla, 2011). This method is defined as the weighted sum of transformations of the radially symmetric basis function. In this method, a set of distinct point data and corresponding observation values are considered, and the primary radial basis function used for interpolation is expressed as follows (Sharifi, 2016):

$$f(x) = \sum_{i=1}^{n} w_i U_i(r) \quad (5)$$

where n is the number of reference stations and $U_i(r)$ is a function of the kernel interpolator. This function depends on the Euclidean distance. r is the Euclidean distance and w_i are the unknown expansion coefficients and can be calculated with the input values of the kernel interpolator through the linear system. The generalized version of the radial basis function of Hardy, which was developed for the preparation of topographic maps, is in the form of the following equation (Leyla, 2011):

$$f(x) = \sum_{i=1}^{n} \alpha_i U_i(r) + \sum_{k=1}^{u_m} \beta_k p_k(x) \quad (6)$$

α_i and β_k are unknown polynomial coefficients The second term of the equation (6) represents the polynomial function whose coefficients are as follows:

$$p_1 = 1, \ p_2 = \varphi, \ p_3 = \lambda, \ p_4 = h \quad (7)$$

As a result of the linear interpolation system, the number of unknowns is more than the number of observations, and conditions are added for its solution. Therefore, u_m more equations are needed to form a certain linear system. Consequently, equation (8) must hold (Leyla, 2011, Sharifi, 2016).

$$\sum_{i=1}^{n} \alpha_i p_k(x_i) = 0 \quad (8)$$

Equation (8) is equal to:

$$P^T \alpha = 0 \quad (9)$$

In other words, in this article, according to the data in the 3 dimensions of longitude, latitude and altitude, the relationships $\sum_{i=1}^{n} \alpha_i = 0$, $\sum_{i=1}^{n} \alpha_i \varphi_i = 0$, $\sum_{i=1}^{n} \alpha_i \lambda_i = 0$ and $\sum_{i=1}^{n} \alpha_i h_i = 0$ should be established.
The combination of interpolation conditions and additional equations forms a linear system of equations, expressed as follows (Leyla, 2011).

$$\begin{pmatrix} A & P \\ P^T & 0 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad (10)$$

where A is the design matrix and contains the kernel of the radial basis function, is a square matrix according to the number of reference stations and is defined as follows:

$$A = \begin{bmatrix} U(r_{11}) & U(r_{12}) & \cdots & U(r_{1n}) \\ U(r_{21}) & U(r_{22}) & \cdots & U(r_{2n}) \\ \vdots & \vdots & \ddots & \vdots \\ U(r_{n1}) & U(r_{n2}) & \cdots & U(r_{nn}) \end{bmatrix} \quad (11)$$

And the matrix P is defined as follows:

$$P = \begin{bmatrix} 1 & \varphi_1 & \lambda_1 & h_1 \\ 1 & \varphi_2 & \lambda_2 & h_2 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & \varphi_n & \lambda_n & h_n \end{bmatrix} \quad (12)$$

There are different kernels for radial basis function interpolation, in this paper we select the inverse multiquadric (IMQ) interpolator kernel.

2.1.1 Inverse Multiquadric: IMQ method is based on radial basis functions. For n distinct data, The IMQ is expressed as follows:

$$f(\varphi,\lambda,h) = \sum_{i=1}^{n} \alpha_i U_i(\varphi_i,\lambda_i,h_i,\varphi,\lambda,h) + \sum_{k=1}^{4} \beta_k p_k(\varphi,\lambda,h) \quad (13)$$

(φ,λ,h) is the coordinates of the station where the interpolation is done, and $(\varphi_i,\lambda_i,h_i)$ is the coordinates of the reference stations.

The IMQ interpolator kernel is defined as (Sturgill, 2009):

$$U(r) = \frac{1}{\sqrt{r^2 + c^2}} \quad (14)$$

where r is the Euclidean distance and c is the shape parameter determining the interpolation accuracy. The shape parameter is obtained by the cross-validation method (Sturgill, 2009).

Interpolation based on the IMQ method requires discrete data, therefore, the following equation must hold (Leyla, 2011):

$$r = \sqrt{(x_i - x)^2 + (y_i - y)^2 + (z_i - z)^2} > 1000m \quad (15)$$

In this paper calculation is based on having the geodetic coordinates of the points from equation (16):

$$r_{s_0,s} = 2R \sin^{-1}\left(\sqrt{A+B}\right) \quad (16)$$

where $\quad A = \sin^2\left(\dfrac{\varphi_s - \varphi_{s_0}}{2}\right)$

$$B = \cos\varphi_{s_0}\cos\varphi_s\sin^2(\dfrac{\lambda_s - \lambda_{s_0}}{2})$$

besides R is half the greatest length of the earth and s are the data points.

The position of the control points, which in this paper is in the form of longitude, latitude, and altitude, and the information, is expressed as follows:

$$D^T = \begin{bmatrix} PWV_1 & \cdots & PWV_n & 0 & 0 & 0 & 0 \end{bmatrix} \quad (17)$$

The unknown coefficients α, β of equation (13) are calculated via the following equation:

$$\begin{bmatrix} 0 & U(r_{12}) & \cdots & U(r_{1n}) & 1 & \varphi_1 & \lambda_1 & h_1 \\ U(r_{21}) & 0 & \cdots & U(r_{2n}) & 1 & \varphi_2 & \lambda_2 & h_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots \\ U(r_{n1}) & U(r_{n2}) & \cdots & 0 & 1 & \varphi_n & \lambda_n & h_n \\ 1 & 1 & \cdots & 1 & 0 & 0 & 0 & 0 \\ \varphi_1 & \varphi_2 & \cdots & \varphi_n & 0 & 0 & 0 & 0 \\ \lambda_1 & \lambda_2 & \cdots & \lambda_n & 0 & 0 & 0 & 0 \\ h_1 & h_2 & \cdots & h_n & 0 & 0 & 0 & 0 \end{bmatrix}^{-1} \begin{bmatrix} PWV_1 \\ PWV_2 \\ \vdots \\ PWV_n \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

$$= \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix} \quad (18)$$

Then by using equation (13), PWV values in other places can be obtained.

2.2 Artificial neural network

ANN is an information processing system that mimics the brain's biological neural network by connecting many artificial neurons. Artificial neural networks can train and generalize. The work of a neural network is to create a specific output pattern according to a learning process based on the input pattern provided to the system. There are different types of neural networks. Perceptron neural network is used in this paper. because it is mainly used for interpolation (Zhang et al., 1998).

2.2.1 Perceptron neural network: Multilayer perceptron neural networks (MLPNNs) is the most common type of neural network due to their fast performance, simple implementation process, and small training set requirement (Kocyigit et al., 2008). According to Figure (1), MLPNN consists of three consecutive layers: input, hidden, and output layers. The hidden layer processes the input information and sends it to the output layer.
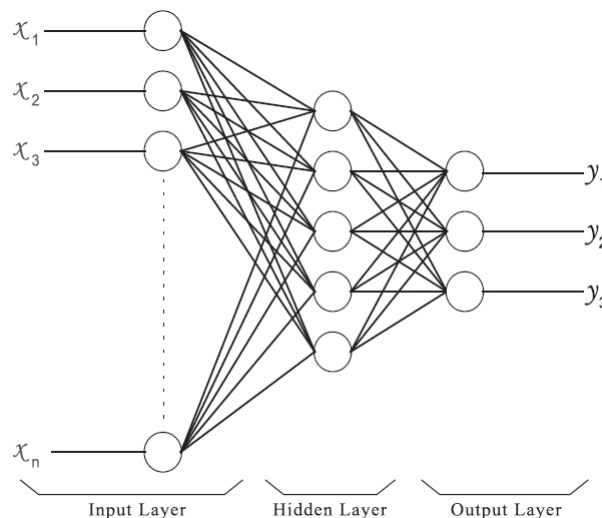

Figure 1. The structure of the MLPNN model (Orhan et al., 2011)

There is no way to determine the number of neurons in the hidden layer. Therefore, the number of neurons is selected only by trial and error. If the network does not converge to the desired value, the number of hidden layer neurons are increased (Kocyigit et al., 2008, Subasi, 2005). Perceptron neural network training is based on the backpropagation algorithm. The weights in perceptron neural network training are adjusted in a way that the error between the desired output and the output is reduced. The values of the main inputs are actual values, where our inputs are the three variables of latitude, longitude, and geodetic height. According to the importance the entries are weighted. Each neuron j in the hidden layer sums its input signals xi impinging onto it after multiplying them by their respective connection weights wji. The output of each neuron is described as follows (Orhan et al., 2011, Sharifi, 2016):

$$y_j = f(\sum w_{ji}x_i) \quad (19)$$

where f is the activation function using the weighted sum of inputs, ReLU activation function is used in this paper, is as defined as by Eq (20). and is of the form shown by Figure (2). ReLU is a non-linear function. In addition, this function is derivable and can be used for backpropagation, with few calculations (Dubey and Jain, 2019).

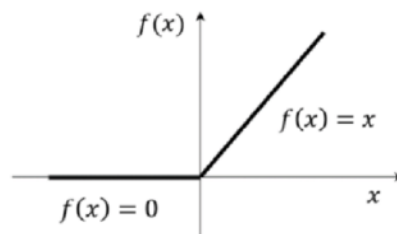$$f(x) = \begin{cases} 0 & x < 0 \\ x & x \geq 0 \end{cases} \quad (20)$$

Figure 2. ReLU activation function (Dubey and Jain, 2019)

2.3 Inverse distance weighted (IDW)

One of the most common spatial interpolation methods in earth sciences is the IDW method. In this method, the position of each point is considered separately, and the relative position of the points is not considered. In this method, the weight is a function of the inverse of the distance, and the closer points have a more significant effect on the estimation of the unknown point so that the points with the same distance from the point with unknown values are given the same weight. The closest points are given highest weight, and it follows as: (Shepard, 1968, Sarma, 2010).

$$f(\varphi, \lambda, h) = \sum_{i=1}^{n} w_i f(\varphi_i, \lambda_i, h_i) \qquad (21)$$

where w_i are the weights and are obtained from the following equation:

$$w_i = \frac{r_i^{-2}}{\sum_{i=1}^{n} r_i^{-2}} \qquad (22)$$

where r is the Euclidean distance between the sample stations and the location to be interpolated. And n is the number of reference stations.

3. STUDY AREA AND DATA

An area of 15448 square kilometers was selected in the state of Washington, USA. In this paper, 23 GPS stations of the UNAVCO network, a research institute related to earth sciences, have been used. The location and distribution of the stations are shown in figure (3).
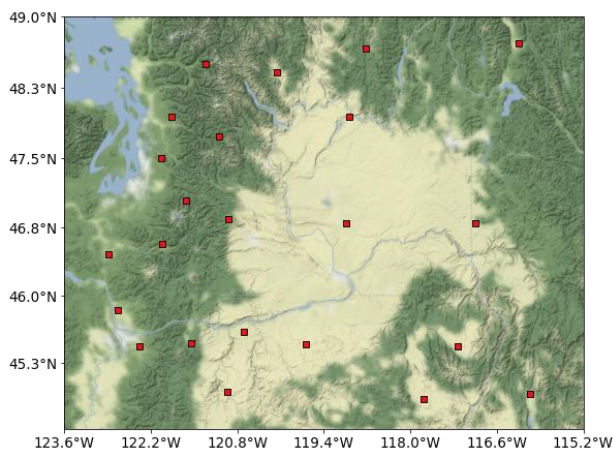

Figure 3. Map of the study area and the satellite positioning reference stations

In addition, four stations of this network are considered as control stations to check the accuracy of interpolation. The details of these stations are shown in table (1). Besides, in this paper, the reanalyzed data of ERA-5 is used for accurate estimation of surface pressure and surface temperature for PWV calculation. The temporal resolution of ERA-5 data is 1 hour, and its spatial resolution is about 30 kilometers and provides data in 37 pressure layers.

Station name	Latitude N(°)	Longitude W(°)	Altitude (m)
OBSR	46.899797	121.815367	2380.03
P700	46.17813	122.217349	1538.67
SC00	46.950925	120.724609	1197.74
P450	45.953335	119.544194	183.25

Table 1. Names and coordinates of control stations

4. PROCESSES AND RESULTS

In Bernese software, the data of GPS stations have been processed by the PPP method, and ZTD has been calculated in each station with a time resolution of 1 hour. Besides ZHD has been calculated according to the latitude and atmospheric pressure obtained from ERA-5 data at the stations using the Saastamoinen dry model. Then, with the difference of ZHD from ZTD, the troposphere delay is estimated, and using equation (3), PWV has been obtained for each station in millimeters. For interpolation by IMQ, using the cross-validation method, the optimal shape has been determined so that RMSE be having the smallest value. RMSE values according to different shape parameters are shown in Figure (4). The shape parameter 5000 has the lowest RMSE value and is selected as the optimal shape parameter. Then, the Euclidean distance between the reference stations and the control stations is calculated. Using the design matrix formed by the IMQ interpolator kernel, the unknown coefficients in equation (11) have been obtained. Finally, the interpolation has been done by the IMQ method for the control stations.
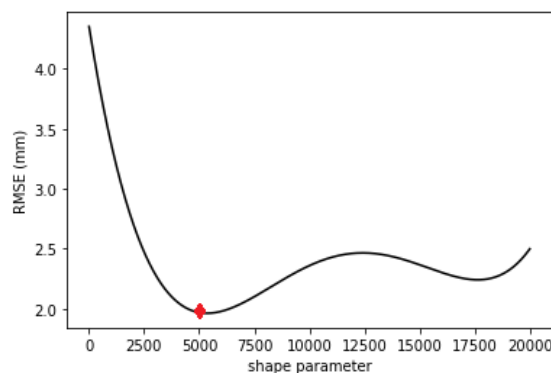

Figure 4. Determination of the shape parameter

By changing the number of hidden layers and trial and error, the number of hidden layers with a good answer is obtained. The number of neurons is also selected only by trial and error. Having too few neurons can reduce the stability of the system. Using more hidden layers, instead of increasing the number of hidden neurons, increases the processing power of the system. So, in this paper, an input layer, two hidden layers, and an output layer are used for perceptron neural network interpolation. Furthermore, the activation function of both layers is considered as the ReLU function. Using this function, modeling time is reduced. The first hidden layer has twelve neurons and the second hidden layer has eight neurons. 30% of

the data have been selected as test data, and 70% of the data have been selected as training data. The inputs, which include the latitude, longitude and altitude of the stations, are multiplied by their corresponding weight, and their sum is calculated. In the training step, the difference between the actual PWV values and the desired PWV is calculated. In this step, the weights are modified, and the network is trained. Finally, the latitude and longitude, and height of the control stations are entered into the system to obtain the PWV values. In the next step, the interpolation is done by the IDW using the data weighting based on the Euclidean distance between the reference stations and the control stations and using the mentioned equation (21).

RMSE and relative root mean square value (RRMSE) are the evaluation criteria and comparison between the interpolation methods performed in this paper as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(PWV_i - PWV_i^R)^2} \qquad (23)$$

$$RRMSE = \frac{RMSE}{\frac{1}{n}\sum_{i=1}^{n}PWV_i^R} \times 100 \qquad (24)$$

where PWV_i^R is the real value and PWV the interpolated value at each control station. And n is the number of stations. Table (2) shows the comparison of the results.

Method	RMSE (mm)	RRMSE (%)
IMQ	2.11	16.92
ANN	3.02	24.25
IDW	4.21	33.79

Table 2. Statistical comparison according to the results obtained

The results obtained from RMSE of the interpolated models with a resolution of 1 hour are shown in Figure (5). Also, the comparison between RMSE between PWV obtained from GPS and PWV obtained from the IMQ interpolation method in each control station is shown in figure (6).
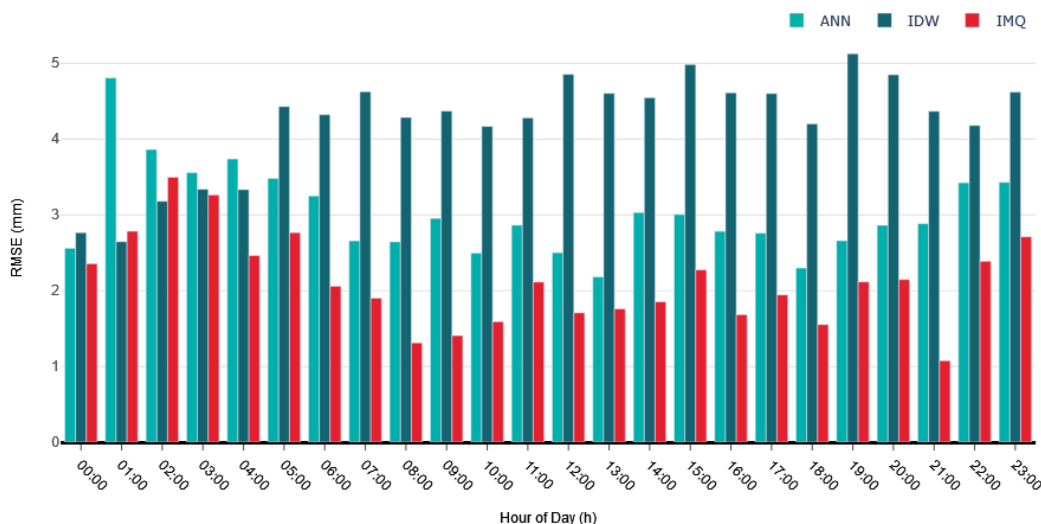

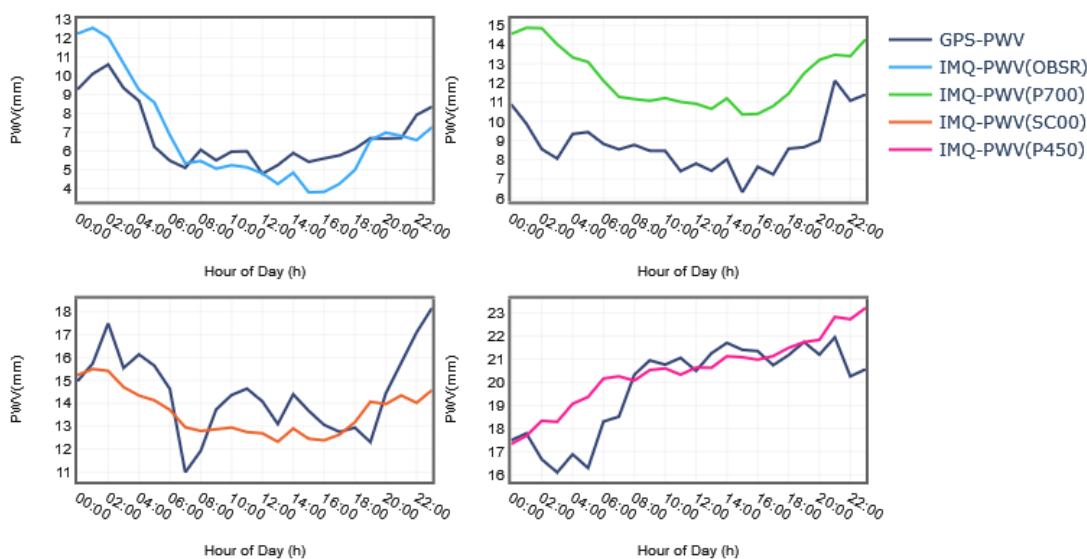**Figure 5.** RMSE of different interpolation schemes (in hour)

Figure 6. Comparison between PWV obtained from GPS signals and PWV obtained from IMQ interpolation of the four control stations

In the last step, a dense PWV map is prepared using the IMQ method in the region. For this purpose, interpolation points have been selected on a regular grid in the area with proper distribution and a distance of about 6 km from each other.

Then, interpolation using the IMQ method has been used to estimate PWV values at forecast points. Dense maps of PWV to investigate PWV changes with the best accuracy in 3 different hours of the day are displayed in Figure (7).
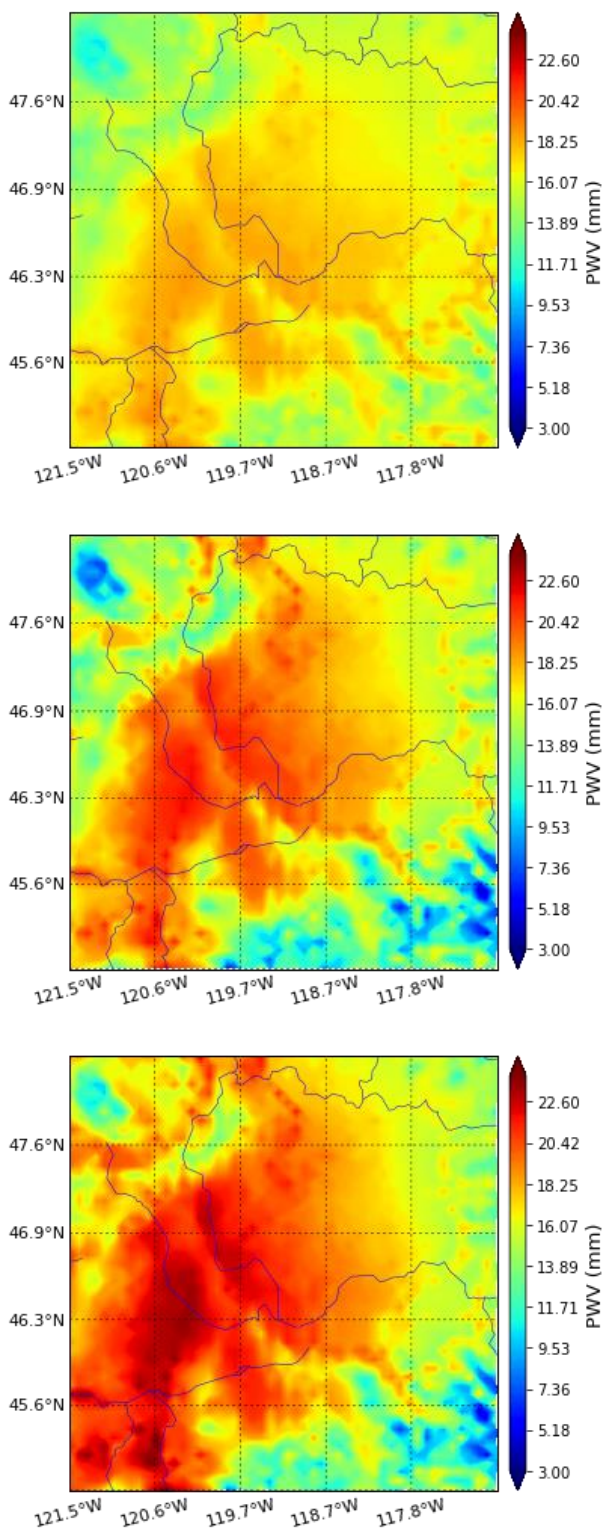

Figure 7. Dense maps of PWV values from top to bottom at 00:02, 14:00, and 22:00 using IMQ method

The amount of PWV is changing during the day. According to the figure (7), at 00:02, the amount of PWV is from 10.1 to 19.1 mm, at 14:00 from 3.2 to 21.2 mm, and at 22:00 from 3.7 to 24 mm.

5. CONCLUSION

In this paper, to improve the accuracy of PWV estimation in the troposphere, The IMQ method was used and compared with ANN and IDW methods. Four stations among the points in the area were considered as control stations. The RMSE value obtained for the IMQ method is 2.11 mm less than other methods, which means that the IMQ method has better accuracy than other methods. The RRMSE value obtained for the IMQ method is 16/9% mm, which is 9% better than the ANN method and 17% better than the IDW. As a result, the IMQ interpolation method is more suitable for sparse data, and the dense map created by the IMQ method shows the PWV changes very well.

REFERENCES

BEVIS, M., BUSINGER, S., HERRING, T. A., ROCKEN, C., ANTHES, R. A. & WARE, R. H. 1992. GPS meteorology: Remote sensing of atmospheric water vapor using the global positioning system. *Journal of Geophysical Research: Atmospheres,* 97**,** 15787-15801.

CHEN, B., DAI, W., LIU, Z., WU, L. & XIA, P. 2018. Assessments of GMI-derived precipitable water vapor products over the south and east China seas using radiosonde and GNSS. *Advances in Meteorology,* 2018.

DAVIS, P. J. 1975. *Interpolation and approximation*, Courier Corporation.

DUBEY, A. K. & JAIN, V. 2019. Comparative study of convolution neural network's relu and leaky-relu activation functions. *Applications of Computing, Automation and Wireless Systems in Electrical Engineering.* Springer.

KOCYIGIT, Y., ALKAN, A. & EROL, H. 2008. Classification of EEG recordings by using fast independent component analysis and artificial neural network. *Journal of Medical Systems,* 32**,** 17-20.

LEYLA, C. 2011. Modified quadratic shepard and radial basis functions interpolations for determination of orthometric heights from global positioning system (GPS) heights. *International Journal of Physical Sciences,* 6**,** 8183-8187.

MENDES, V. 1999. Modeling the neutral-atmospheric propagation delay in radiometric space techniques. *UNB geodesy and geomatics engineering technical report*.

NILSSON, T., BöHM, J., WIJAYA, D. D., TRESCH, A., NAFISI, V. & SCHUH, H. 2013. Path delays in the neutral atmosphere. *Atmospheric effects in space geodesy.* Springer.

ORHAN, U., HEKIM, M. & OZER, M. 2011. EEG signals classification using the K-means clustering and a multilayer perceptron neural network model. *Expert Systems with Applications,* 38**,** 13475-13481.

SAASTAMOINEN, J. 1972. Atmospheric correction for the troposphere and stratosphere in radio ranging satellites. *The use of artificial satellites for geodesy,* 15**,** 247-251.

SARMA, D. D. 2010. *Geostatistics with applications in earth sciences*, Springer Science & Business Media.

SHARIFI, M. A. 2016. Comparison of the geodetic height correcting surface determination methods: A case study for Tehran. *Iranian Journal of Geophysics,* 10**,** 40-52.

SHEPARD, D. A two-dimensional interpolation function for irregularly-spaced data. Proceedings of the 1968 23rd ACM national conference, 1968. 517-524.

SMITH, E. K. & WEINTRAUB, S. 1953. The constants in the equation for atmospheric refractive index at radio frequencies. *Proceedings of the IRE,* 41**,** 1035-1037.

STURGILL, D. 2009. Variable shape parameter strategies in radial basis function methods.

SUBASI, A. 2005. Automatic recognition of alertness level from EEG by using neural network and wavelet coefficients. *Expert systems with applications,* 28**,** 701-711.

YANG, F., GUO, J., MENG, X., SHI, J. & ZHOU, L. 2019. Establishment and assessment of a new GNSS precipitable water vapor interpolation scheme based on the GPT2w model. *Remote Sensing,* 11**,** 1127.

ZHANG, G., PATUWO, B. E. & HU, M. Y. 1998. Forecasting with artificial neural networks:: The state of the art. *International journal of forecasting,* 14**,** 35-62.