

PREDICTION OF PRECIPITABLE WATER VAPOR WITH A NEURAL NETWORK FROM THE ECUADORIAN GNSS AND METEOROLOGICAL DATA

RICARDO ROMERO¹; CRISTIAN PILAPANTA¹; LUIS PORRAS¹; ALFONSO TIERRA².

¹ INSTITUTO GEOGRÁFICO MILITAR, QUITO, ECUADOR. Seniergues E4-676 y Gral. Telmo Paz y Miño, Quito, Ecuador. E-mail: luis.porras@igm.gob.ec

² GRUPO DE INVESTIGACIÓN GEOESPACIAL, UNIVERSIDAD DE LAS FUERZAS ARMADAS ESPE. Av. General Rumiñahui S/N, Sangolquí, Ecuador. E-mail: artierra@espe.edu.ec

Recibido: 24 de agosto de 2018 / Aceptado: 15 de noviembre de 2018

ABSTRACT

The Global Positioning System (GPS) consists of a constellation of satellites that transmit of radio signals to large numbers of users engaged in navigation, time transfer, and relative positioning. These L-band radio signals are delayed by atmospheric water vapor while they travel from GPS satellites to ground GPS receivers. Water vapor plays a crucial role in atmospheric processes that act over a wide range of temporal and spatial scales, from global climate to micrometeorology. Atmospheric scientists have developed a variety of means to measure the vertical and horizontal distribution of water vapor. In this work, was studied a model for the prediction of precipitable water content from GNSS data using a neural network. In this case was set as the initial parameters: Pressure, temperature, positions and zenith total delay and for the neural network was one of type a radial basis function neural network (RBFNN) with three layers. Results demonstrated that the RBFNN achieved to predict precipitable water vapor with a RMS until to 2 mm in overall GNSS network, it means, this methodology is a valuable alternative to establish a modeling for these conditions and parameters.

Keywords: Atmospheric Water Vapor, Global Navigation Satellite System, Precipitable Water Vapor, Radial Basis Function Neural Network.

INTRODUCTION

Space geodesy refers to observations that are transmitted or received by natural or artificial objects outside the lower portions of the atmosphere, i.e., in space where the density of the atmosphere is sufficiently small to allow stable satellite orbits. Space-geodetic techniques like VLBI, GNSS, SLR, or DORIS are used to observe Earth rotation variations, a large part of which is caused by atmospheric effects (Böhm and Schuh, 2013).

The atmosphere has been modeled for the purpose of analysis, short-term weather forecasts and climate projections (Nilsson et al., 2013). Such models, which are produced and used by the world's major research and weather forecast centers are based on the principles of atmospheric physics which frame an analysis of the weather and climate system.

The Global Positioning System (GPS) consist of a constellation of satellites that transmit of radio signals to large numbers of users engaged in navigation, time transfer,

and relative positioning (Bevis et al, 1994). These L-band radio signals are delayed by atmospheric water vapor as they travel from GPS satellites to ground receivers. Water vapor plays a crucial role in atmospheric processes that act over a wide range of temporal and spatial scales, from global climate to micro-meteorology. The neutral atmosphere is a mixture of dry gases and water vapor (Bevis et al, 1992), furthermore, the water vapor is unique in this mixture because it is the only constituent which possesses a dipole moment contribution to its refractivity.

The amount of precipitable water vapor (PWV) contained in the neutral atmosphere can be inferred from the propagation delay of Global Positioning System (GPS) signals passing through the troposphere (Nistor and Buda, 2015). Throughout most of the troposphere the dipole component of the refractivity is about 20 times larger than the nondipole component (Bevis et al, 1992). For this reason it has become common to treat the dipole component of the water vapor refractivity separately from the nondipole components of the refractivity of the water vapor and other constituents in the atmosphere. These two components are referred to as the wet (ZWD) and hydrostatic (ZHD) delays which both are constituents of the zenith total delay (ZTD).

Since the observables of space geodesy typically are measurements of the travel time of the signals, the absorption is typically not important since it does not affect the propagation delay (Böhm and Schuh, 2013). Of course, absorption will affect the delay measurements by increasing the noise; higher attenuation will cause the signal-to-noise-ratio to be lower, and thus the accuracy of the measured delay will be worse (in the worst case the signal cannot be detected).

Meteorologists and GPS specialists working together should be able to design procedures that can be used to characterize the troposphere in more detail. Mathematical techniques have been developed to map the delay at any elevation and the removal of the tropospheric delay by estimation has become an integral part of precise VLBI and GPS analyses (Niell, 1996; Herring, 1992; MacMillan and Ma, 1994).

In space geodesy normally the travel time (or difference in travel time) between a source in space (a satellite or a quasar) to a receiver on the surface of the Earth is measured. If the variations in the refractivity over the distance of one wavelength is negligible, we can use the geometric optics approximation (Nilsson et al., 2013). This means that the propagation of an electromagnetic wave can be described as a ray. When calculating the propagation time of the electromagnetic wave we thus only have to consider the refractivity along the ray path. Similarly, for the microwaves, the refractivity can also be divided into a hydrostatic and a non-hydrostatic (wet) part (Nilsson et al., 2013). It is common to refer the slant delays to the delays in the zenith direction. The zenith hydrostatic delay is subtracted from the zenith neutral delay to determine the zenith wet delay, which is then transformed into an estimate of precipitable water. By incorporating a few remote global tracking stations (and thus long baselines) into the geodetic analysis

of a regional GPS network, it is possible to resolve the absolute (not merely the relative) value of the zenith neutral delay at each station in the augmented network. The integrated water vapor in zenith direction can also be provided as precipitable water vapor (PWV) which corresponds to the height of the equivalent water column above the station following by Nilsson et al, 2013.

Water vapor is an extremely important variable for severe weather forecasting and operational numerical weather prediction, because moisture distribution is directly related to the formation of clouds and precipitation (Kuo et al, 1996). The aim of this work is to establish a model for prediction of precipitable water content using the GNSS data processing with a spatio-temporary-meteorological neural network. This model can provide recommendations to make decisions related with the analysis of atmospheric parameters.

ARTIFICIAL NEURAL NETWORK (ANN)

In this work was used a radial basis function neural network (RBFNN) by its interpolation approach in a space high-dimensionally (Broomhead and Lowe, 1988). RBFNN are supervised, feedforward and fully connected networks in which the learning process consists in calculate weights and bias in a defined number of iterations until reaching a close response to desired output.

The RBFNN typically have three layers of which hidden layer performances non-linear modeling with free parameters and it maps the input space onto a new space. Thus, output layer implements a linear combination on this new space to adjust linear parameters (weights) (Chen et al, 1991). The fundamentals of RBFNN derives from the theory of approximation, where function f is intended to look for:

$$f(x) = \sum_{i=1}^K b_i \diamond((x - c_i)) \quad (1)$$

where j is the radial basis function, c_i are K centers (weights) which have to be calculated and b_i is the bias vector for each hidden neuron.

The performance of training of an Artificial Neural Network (ANN) could be measured by various statistical methods, following different evaluation criteria such as the coefficient of correlation (COR), the root mean squared error (RMSE), the mean absolute error (MAE) and the Willmott's index of agreement (Ladlani et al, 2012).

In order to implement the process with the RBFNN was necessary to choose a set of data to the learning process and another set to validate the parameters calculated in the earlier procedure. Input data was selected according certain precedents, for example, the

derivation of PWV from ZTD (zenith wet delay) (Bevis et al, 1994; Bevis et al, 1992; Hagemann et al, 2003), which it demonstrates an interdependence for each variable. Although the potential use of GNSS data comes given by its high temporal resolution, the corresponding time derivative of PWV highlights changes in the atmospheric water content on the different time scales (Bordi et al, 2015). Additionally, the atmospheric delays are highly correlated with the elevation of the site (Xu et al, 2011).

Furthermore, Senkal (2015) shows a high degree of correlation between meteorological parameters and location with the PWV value, which it suggests that a process with an Artificial Neural Network is more cheaper and faster in relation with others meteorological methods for the estimation of the PW content. However, PWV time series have been proved a poor correlation with precipitation, which that affects the process of the Precipitable Water availability in the middle atmosphere (Bordi et al, 2015).

METHODOLOGY

The strategy of GNSS data processing gave as a result values of PWV defined by meteorological data from global models available in a geodetic software processing, and which for this study it was chosen the GPT model (Global Pressure and Temperature) in Gamit/Globk (Herring et al, 2015). For covering the majority of the continental territory, all the Ecuadorian GNSS Permanent Network formed by 34 meteorological stations (model 4A) was processed (Figure 1). The data of processing was selected for all the year 2015 with an interval of record of 1 second and precisions for Pressure: better than + 0.08 hPa, Temperature: better than + 0.5 °C and for the relative humidity: better than + 2% a 25°C.

With this background, the variables selected were location (latitude, longitude and ellipsoidal height), the meteorological parameters (temperature and pressure), the ZTD and the epoch of occurrence, all of them with the purpose of predicting the values of PWV.

The processing of data was made of the following form: a) Download and quality control of the rinex data obtained from each of the continuous monitoring GNSS stations, b) Data processing using GAMIT/GLOBK, c) Quality control of raw files and d) Calculation of atmospheric parameters using GAMIT metutil tool. The final files contain the following precisions:

A standard established for training an ANN is about of the selection of training, generalization and validation of data sets, where the first one is intended for to calculate the parameters of the learning, which the conditions to assign this data set for this study was based on spatial criteria. The second is for to determine the influence of smoothness of function with respect to different inputs and in the last case, this is the validation data set is not taken in account in training process because the purpose of this data is to

measure at the end the capability of learning of the ANN.

In this study were selected only two data sets for training and for test, the training and generalization data were merged into the software. In figure 1 is shown the distribution of training and test data of the Ecuadorian GNSS Permanent Network.

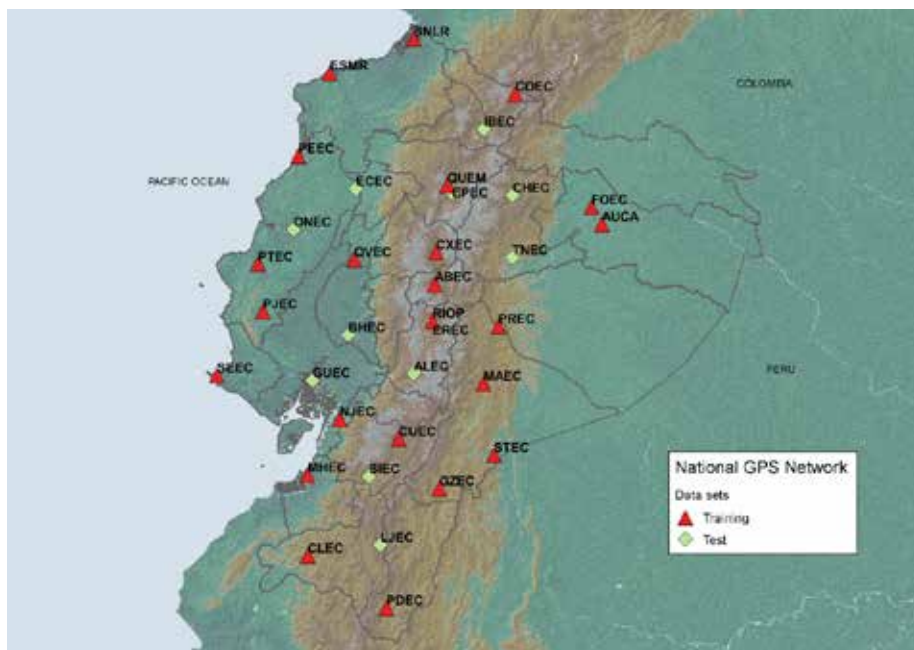


Figure 1: Distribution of meteorological stations in Ecuador and the distribution of data set for training RBFNN. Fuente: Instituto Geográfico Militar, 2018.

For the analysis, 22 stations were selected for the training and generalization functions of the ANN where these stations are located in zones with lowest and highest altitudes covering all Continental Ecuadorian territory, however the data set were chosen randomly. The next step was to implement the RBFNN with inputs and outputs normalized and where the objective was to perform the learning. After that, the initialization function was started for to calculate the weights of the connections according to the input patterns.

Firstly, the procedure selects centers c_i randomly from the input patterns and assigns them to the links between input and hidden layer. Afterward the bias b_i is set to a value in all hidden neurons and finally the links (w_{jk}, b_2) between the hidden and output layers are calculated (Bergmeir and Benitez, 2012) following the equation 2 and where the activation function was a Gaussian function (ϕ_1) and the output function (ϕ_2) was a linear regression.

$$y_j = \phi_2 \left[b_2 + w_{jk} \sum_{i=1}^K \phi_1 (\|x - c_i\|) b_1 \right] \quad (2)$$

The implementation was achieved in R software with models provided by the RSNNS package (Bergmeir and Benitez, 2012; Zell et al, 1998) with parameters for Gaussian functions. The training process was made with the framework listed in the table 1.

Table 1. Fundamental parameters/functions implemented in RBFNN to predict PWV in Ecuador.

Parameter/Function	Value/Name
Activation Function	Gaussian Function
Output Function	Linear Function
Normalization	Scaling Between 0 And 1
Number of Hidden Neurons	50
Max. Iterations	50000
Spread	0,63
Learning Rate	1 E -8
Error	1e-3
Momentum	0,9

RESULTS AND DISCUSSION

The first analysis carried out in this study was a comparison between selected inputs and the PWV from training data established to define some type of correlation inwardly. This fact is showed in the figure 2 in which is illustrated the affinity with the PWV.

Six input data sets were studied, that were latitude, longitude, ellipsoidal height, zenith, total delay, pressure and temperature. Only the epoch was not analyzed because that it determines a temporal characteristic of the neural network.

In the Figures 2a. and 2b. did not show an explicit correlation, which implies that PWV values do not depend of the spatial location, however in the graph 2b is easy to identify that PWV has an increase when the longitudes are for the coast region (-80° , -81°) and for the amazon region (-77° , $-77,5^{\circ}$). This fact is probably because of the dense cloudiness in these zones.

On the other hand, in relation to the Ellipsoidal Height, in the Figure 2c, showed an inverse relation with the PWV values, the reason is due to low values of relative humidity in highlands of Ecuador. This fact is similar for the ZTD (Figure 3d) because PWV values are derivable of the ZWD values that it is a component of ZTD. Finally, the meteorological parameters as pressure and temperature suggest a direct relation with the PWV values. To confirm these approximations, a matrix of correlation was calculated in order to prove earlier statements. The product of the matrix appears in the table 2.

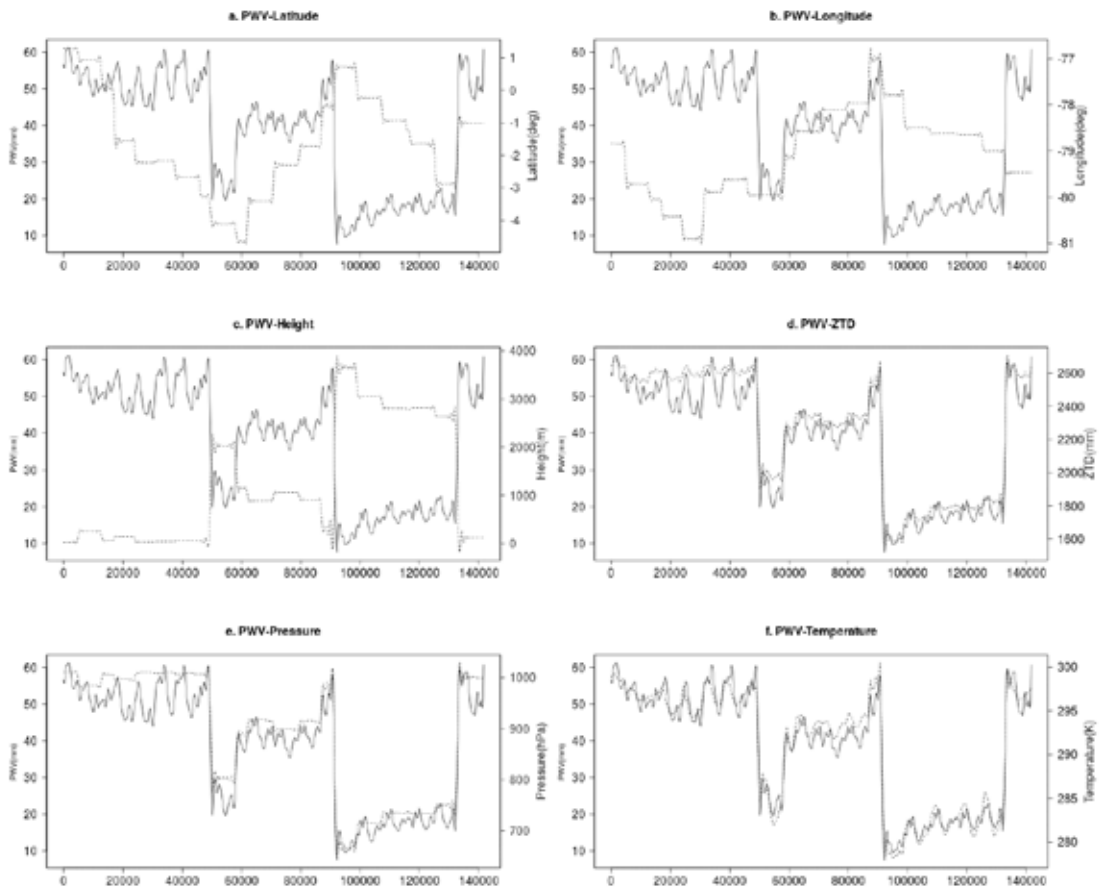


Figure 2. Graphical correlation between PWV and input variables of RBFNN. Continuous lines represent the PWV values and the dash lines represent the individual inputs, (a) Graphical correlation between PWV and the Ellipsoidal Latitude, (b) Graphical correlation between PWV and the Ellipsoidal Longitude, (c) Graphical correlation between PWV and the Ellipsoidal Height, (d) Graphical correlation between PWV and the ZTD, (e) Graphical correlation between PWV and the Atmospheric Pressure, and (f) Graphical correlation between PWV and the Temperature. Fuente propia, 2018

Table 2. Correlation between the inputs parameters of the RBFNN and the PWV values.

	Latitude	Longitude	Height	ZTD	Pressure	Temperature
PWV	0,002	-0,41	-0,96	0,98	0,96	0,97

As it mentioned before, the neural network was trained with RSNNS package in R, with all input patterns, given by a multiple variations one another. The Figure 3 shows the interaction of inputs in a heat map.

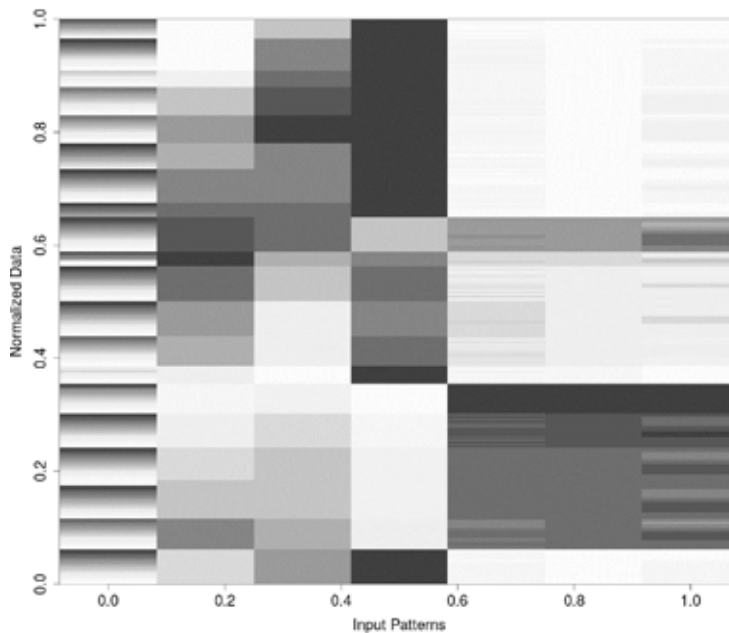


Figure 3. Mapping of Distribution and interaction among inputs represented in a heat map.

In the figure 3 is represented all inputs in columns (epoch, latitude, longitude, height, ZTD, pressure and temperature) in a heat map for to display the way of activation of each normalized parameter. Black color represents low values that is degraded to white color that represents high values. This technique of representation allows the supervisor interpreting heterogeneity of data sets therefore the conduct of inputs.

The training process was executed with parameters indicated in the Table 1. The process is evidenced in the Figure 4a, reaching an error by the sum of squares of 22.32 from 141362 data patterns. The training curve shows a considerable decrease in first stages of learning, after the time, it has demonstrated a stabilization of error close to optimum minimum.

The outcomes of the RBFNN were compared with a desired output in order to establish the coefficient of determination (R^2), which indicates the response of fit of the ANN with normalized real values. It was obtained one R^2 of 0.9971 indicating a good training.

For the validation process, was used a desnormalized data to decide if the generalization of the RBFNN can be implemented to predict confident spatio-temporary-meteorological values of PWV. In this case was obtained a value of 0,98 for R^2 (this fact denotes that the ANN achieves the main purpose of learning). The statistical fundament of this model was based principally on the correlation coefficient (COR), the root mean squared error (RMSE), the mean absolute error (MAE) and the Willmott's index of agreement; all of them are displayed in the table 3.

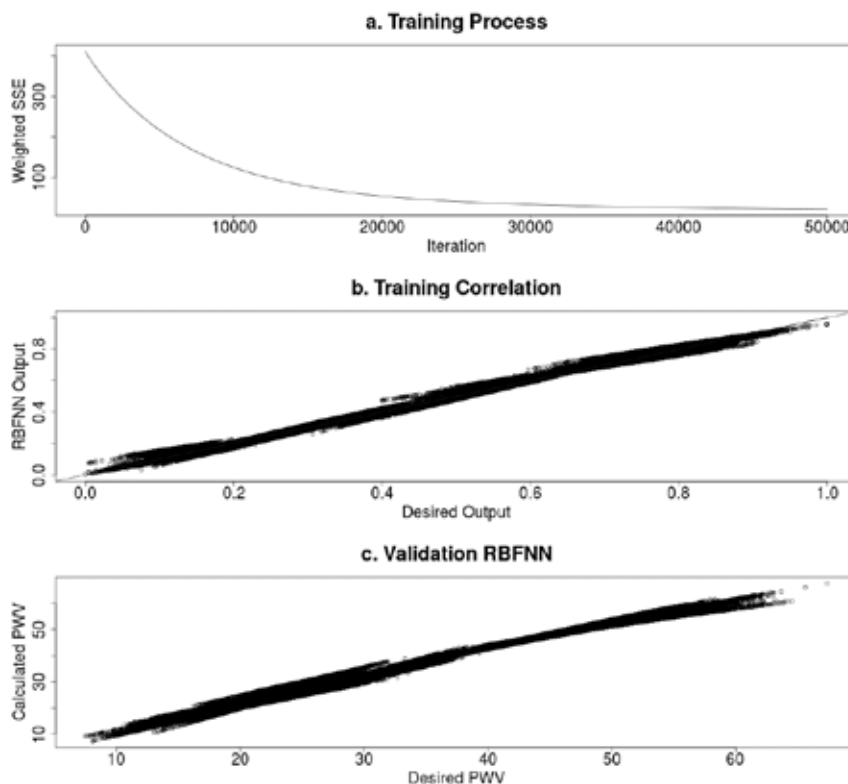


Figure 4. The learning process of RBFNN to predict the PWV with inputs related to epoch, location, altitude, ZTD and meteorological parameters. (a) The training process to adjust the error. (b) Training correlation where it indicates how well ANN response fit with normalized real values. (c) Validation process between the calculated values by RFBNN and desired values of PWV.

Table 3. Main Statistical Parameters for the Validation Process.

Parameter	Values
COR	0,99
RMSE	2,33 mm
MAE	1,95 mm
Index of Agreement	0,99
Standard Deviation	1,33 mm

These results can be looked at the Figure 5, where a double y-axis plot illustrates in light grey line the real values of PWV and in black line the values calculated by RBFNN model. Additionally, the dash line represents the error in mm of each calculated value.

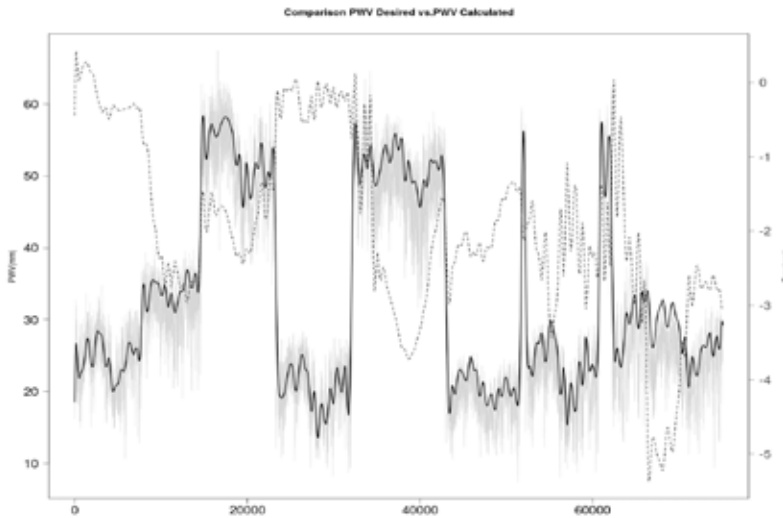


Figure 5. Comparison between the desired PWV (Light grey line) and the calculated values by the RBFNN (Black line) and its corresponding error (Dash line).

For the validation of model is used the MODIS sensor and the results were showed in the table 4. In this case there is a correlation in stations near of the coast but is not the same for the others regions in Ecuador. It's important to know the fact of the climate in Ecuador is very complex.

Table 4. Model Validation Process with the MODIS sensor.

Month	Sensor	Meteorological Station				
		GUEC	PTEC	SEEC	ALEC	MAEC
Jan	IGM	55,58	52,65	51,16	25,31	28,92
	MODIS	46,99	49,95	52,2	36,99	16,34
Feb	IGM	53,77	51,33	49,67	26,82	39,73
	MODIS	47,03	50,6	51,74	39,4	16,08
Mar	IGM	57,08	54,84	52,96	27,99	41,78
	MODIS	50,45	53,67	53,71	42,13	17,7
Apr	IGM	59,06	56,55	55,21	27,76	42,03
	MODIS	51,26	53,87	54,99	43,04	18,53
May	IGM	59,49	56,48	55,55	28,19	41,52
	MODIS	51,23	53,78	55,6	40,86	18,35
Jul	IGM	51,87	50,73	48,98	23,78	38,69
	MODIS	45,82	50,19	49,82	40,17	16,83
Aug	IGM	48,56	46,25	45,7	21,41	36,7
	MODIS	44,28	47,07	45,16	38,99	16,66
Sep	IGM	48,21	47,99	45,85	26,36	37,05
	MODIS	46,62	47,43	47,17	37,88	16,34

Oct	IGM	54,1	49,08	51,27	25,76	40,9
	MODIS	47,32	50,5	50,19	40,78	18,45
Nov	IGM	53,86	50,56	52,37	25,5	42,66
	MODIS	45,17	48,85	47,76	29,7	17,94
Dic	IGM	57,19	58,69	53,73	25,81	40,08
	MODIS	49,42	52,45	54,58	40,71	16,91

CONCLUSIONS

The model for predicting Precipitable Water Vapor (PWV) values based on a Radial Basis Function Neural Network was calculated with a data from the Ecuadorian GNSS Permanent Network and with an accuracy obtained around 2 mm for stations near of coast.

The input pattern was formed with parameters such as epoch, position, altitude, Zenit Wet Delay and meteorological parameters, in which some of them confirmed a degree of correlation with PWV values. For this reason it is important to take in account spatio-temporary-meteorological values for the prediction values and where this assertion have been ratified with the graphical relationships between inputs and PWV values.

The methodology used for the ANN has demonstrated to be a valuable alternative to establish a model for predicting of PWV values, considering the heterogeneity of the input data set and the low correlation with a few parameters.

REFERENCES:

- Bevis M., Businger S., Chiswell S., Herring T.A., Anthes R.A., Rocken C. and Ware R.H. (1994). GPS meteorology: Mapping zenith wet delays onto precipitable water. *Journal of applied meteorology*, 33(3), 379-386.
- Bevis M., Businger S., Herring T., Rocken C., Anthes R. and Ware R. (1992). GPS meteorology: Remote sensing of atmospheric water vapor using the Global Positioning System. *Journal of Geophysical Research*, 97(D14), 15787-15801.
- Böhm J., and Schuh H. (2013). *Atmospheric effects in space geodesy* (pp. 1-33). Berlin, Germany: Springer
- Bordi I., Raziei T., Pereira L.S., and Sutera A. (2015). Ground-based GPS measurements of precipitable water vapor and their usefulness for hydrological applications. *Water Resources Management*, 29(2), 471-486.
- Bordi I., Zhu X., and Fraedrich K. (2015). Precipitable water vapor and its relationship with the Standardized Precipitation Index: ground-based GPS measurements and reanalysis data. *Theoretical and Applied Climatology*, 1-13.
- Broomhead D., and Lowe D. (1988). Multi-variable functional interpolation and adaptive networks. *Complex Systems*, 2, 321-355.
- Chen S., Cowan C.F., and Grant P.M. (1991). Orthogonal least squares learning algorithm for radial basis function networks. *Neural Networks, IEEE Transactions on*, 2(2), 302-309.

- Hagemann S., Bengtsson L., and Gendt G. (2003). On the determination of atmospheric water vapor from GPS measurements. *Journal of Geophysical Research: Atmospheres* (1984–2012), 108(D21).
- Herring T.A., King R.W., Floyd M., and McClusky S. (2015). Introduction to GAMIT/GLOBK, Release 10.6. Available in Department of Earth, Atmospheric and Planetary Science at MIT. <http://www-gpsg.mit.edu/>.
- Herring T.A. (1992). Modeling atmospheric delays in the analysis of space geodetic data. *Proceedings of Refraction of Transatmospheric signals in Geodesy*, eds. JC De Munck and TA Spoelstra, Netherlands Geodetic Commission Publications on Geodesy, 36.
- Kuo, Y. H., Zou, X., and Guo, Y. R. (1996). Variational assimilation of precipitable water using a nonhydrostatic mesoscale adjoint model. Part I: Moisture retrieval and sensitivity experiments. *Monthly Weather Review*, 124(1), 122.
- Ladlani I., Houichi L., Djemili L., Heddami S., and Belouz K. (2012). Modeling daily reference evapotranspiration (ET₀) in the north of Algeria using generalized regression neural networks (GRNN) and radial basis function neural networks (RBFNN): a comparative study. *Meteorology and Atmospheric Physics*, 118(3-4), 163-178.
- MacMillan D.S., Ma C. (1994): Evaluation of very long baseline interferometry atmospheric modeling improvements. *Journal of Geophysical Research: Solid Earth* (1978–2012), 99(B1), 637-651.
- Niell A.E. (1996). Global mapping functions for the atmosphere delay at radio wavelengths. *Journal of Geophysical Research: Solid Earth* (1978–2012), 101(B2), 3227-3246.
- Nilsson T., Böhm J., Wijaya D.D., Tresch A., Nafisi V., and Schuh H. (2013). Path delays in the neutral atmosphere. In *Atmospheric Effects in Space Geodesy* (pp. 73-136). Springer Berlin Heidelberg.
- Nistor S., and Buda A.S. (2015). Using Different Mapping Function In GPS Processing For Remote Sensing The Atmosphere. *Journal of Applied Engineering Sciences*, 5(2), 73-80.
- Senkal O. (2015). Solar radiation and precipitable water modeling for Turkey using artificial neural networks. *Meteorology and Atmospheric Physics*, 1-8.
- Xu W.B., Li Z.W., Ding X.L., and Zhu J.J. (2011). Interpolating atmospheric water vapor delay by incorporating terrain elevation information. *Journal of Geodesy*, 85(9), 555-564.
- Zell A., Mache N., Hübner R., Mamier G., Vogt M., Schmalzl M., and Herrmann K.U. (1994). SNNs (Stuttgart Neural Network Simulator). In *Neural Network Simulation Environments* (pp. 165-186). Springer US.