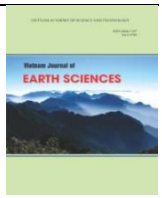
Vietnam Journal of Earth Sciences, 44(4), 521-534, https://doi.org/10.15625/2615-9783/17373

Vietnam Academy of Science and Technology
Vietnam Journal of Earth Sciences
http://www.vjs.ac.vn/index.php/jse



Assessment of machine learning techniques for prediction of integrated water vapor using meteorological data

Nirmala Bai Jadala[1,2], Miriyala Sridhar[1*], D. Venkata Ratnam[1], Gopa Dutta[2]

[1]*Department of ECE, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur Dt, 522302, A.P, India*
[2]*Department of ECE, Vignana Bharathi Institute of Technology, Hyderabad, India*

Received 06 April 2022; Received in revised form 19 July 2022; Accepted 29 July 2022

ABSTRACT

Weather and Climatological studies are critical in assessing atmospheric conditions like storms and cyclones. Integrated water vapor (IWV) is an important greenhouse gas in the atmosphere responsible for the Earth's radiative balance. Global Positioning System (GPS) observations have been used for monitoring the IWV variability. The IWV estimations are carried out using ground-based GPS observations at Hyderabad (17.4°N, 78.46°E), India using GAMIT software. GAMIT is GPS analysis software developed by MIT, USA. It takes input as GPS observation data containing pseudo ranges, navigation data containing ephemeris, clock errors, g-files with orbital information, and meteorological data like pressure, temperature, and relative humidity to calculate IWV. However, estimating IWV for forecasting applications is impossible with a GPS system. This paper introduces a methodology to predict IWV during normal days and severe cyclonic events using machine learning (ML) techniques. Rational quadratic Gaussian process regression (RQ-GPR) and neural network (NN) algorithms are considered for identifying suitable ML prediction algorithms over tropical conditions. Meteorological surface data like Pressure, Temperature, and relative humidity are given as input to the machine learning models. The IWV values computed from GPS are compared with the model's predicted values. RQ-GPR model is showing good accuracy with the IWV values calculated from GPS against the NN model. The correlation coefficient (ρ) achieved for RQ-GPR is 0.93, and 0.86 is obtained for the NN model.

The RMSE (Root Mean Square Error) of the predicted IWV value with RQ-GPR is better than the NN model. We have obtained mean square error (MSE) and mean absolute error (MAE) as 18.146 kg/m^2 and 3.0762 kg/m^2 for RQ-GPR and 27.509 kg/m^2 and 3.9102 kg/m^2 for the NN model which is showing RQ-GPR is a suitable model for forecasting applications. The HUDHUD cyclonic event that occurred in October 2014 is considered for testing the proposed ML algorithms. RQ-GPR model has better results in the Prediction of IWV than the NN model. The RMSE value obtained is 2.837 kg/m^2 for RQ-GPR and 3.327 kg/m^2 obtained from the NN model. The results indicate that the RQ-GPR model has more accuracy than the other IWV prediction models. The prediction results are helpful for meteorology, weather, and climatology studies and useful to improve the accuracy of the regional numerical weather prediction models.

Keywords: Integrated water vapor, machine learning, prediction, rational quadratic Gaussian process regression, neural networks.

1. Introduction

While traveling through the Earth's atmosphere, radio waves get delayed due to refraction (Lanyi, 1984). Water vapor, an essential constituent of the troposphere, is responsible for the delay of the signals. The

*Corresponding author, Email: sridhar.m@kluniversity.in

study of detailed analysis of water vapor in the atmosphere results in a better estimate of precipitation, leading to improved climate change analysis (Duan et al., 1995).

The presence of water vapor in the atmosphere strongly affects the Earth's radiation budget and temperature. Global mean temperature increases with CO_2 and other greenhouse gases in the atmosphere (Dai et al., 2001). Warming becomes double the available water vapor if the CO_2 doubles in the atmosphere (Houghton et al., 2001). However, in comparison with other greenhouse gases, water vapor is highly variable (Held and Soden, 2000; Dai et al., 2000; Trenberth et al., 2003). Integrated water vapor (IWV) is the total atmospheric water vapor in a vertical column from the Earth's surface to the top of the atmosphere. IWV is expressed in kg/m2 (Ruckstuhl et al., 2007).

It is well known that Integrated Water Vapor (IWV) influences weather and climate significantly and is responsible for global warming, changes in the water cycle, etc. (Kiehl and Trenberth, 1997). The evolution of the Global positioning system (GPS) and Global Navigation Satellite System (GNSS) receivers improved the accuracy in measuring integrated water vapor in the atmosphere compared to conventional methods like water vapor radiometers, radiosondes, etc. Many researchers have used GPS/GNSS receivers to estimate IWV (Bevis et al., 1992). Hong et al. (2015) studied the distribution of precipitable water vapor using 952 GNSS stations spread over China. Diurnal and semi-diurnal harmonic variations of IWV using 16 GPS stations over Russia were studied by Kalinkov and Khutorova (2017). The study of integrated water vapor (IWV) using observations of a GPS receiver located at VBIT, Hyderabad, India, has been analyzed by Nirmala Bai et al. (2019).

Machine learning techniques were effectively implemented to develop models catering to the needs of predicting various meteorological phenomena that occur in nonlinear combinations of several processes in the Earth's atmosphere (Bengio et al., 2013).

Ghosh (2010) proposed an optimization model, a machine learning technique called Support Vector Machine (SVM), coupled with a Probabilistic Global Search Algorithm (PGSL) for statistical downscaling and predicted monsoon rainfall of North-Eastern India. A gradient boosting algorithm is implemented to correct the satellite-derived column water vapor error (Just et al., 2020). Jain et al. (2020) researched precipitable water vapor using the LSTMs algorithm and found an RMSE of 0.098 mm. Senkal et al. (2012) investigated the modeling of precipitable water vapor using the ANN algorithm over Turkey. They achieved a correlation coefficient between predicted and actual values of 94% for training data and 91.84% for testing data. Benevides et al. (2019) worked on predicting rainfall using GNSS precipitable water vapor by applying a neural network algorithm and found a better correlation from 63% to 72%. In contrast, the inaccurate positive rate is 36%, down to 21%.

Compared to ANN, Gaussian Process Regression (GPR) technique was an effective tool for handling nonlinear, complex classification and regression problems and, hence, attracted significant attention in the machine learning domain in recent years (Hong et al., 2015).

Moreover, the GPR model was proved to be the better method when compared to the particle filter (P.F.) method for bearing remaining life prediction (Hong et al., 2015). As an alternative to GPS, Suparta and Alhasa (2016) developed a model using an Adaptive Neuro-Fuzzy Inference System (ANFIS) technique and estimated precipitable water vapor (PWV) with a 99% confidence level over peninsular Malaysia. Gao et al. (2017)

reported that the GPR squared exponential model showed 94.8% accuracy, while other GPR models were less accurate in estimating rock fragmentation (Gao et al., 2017). In addition, studies of six-year rainfall data in South Tengerang, Indonesia, Suparta, and Samah (2020) predicted rainfall using the ANFIS time-series technique with 80% data validity.

2. Data and Methodology

A NOVATEL dual-frequency GPS receiver GSV4004VB is available at Vignana Bharathi Institute of Technology (VBIT), Hyderabad (https://www.navtechgps.com/novatel_gpstation_6_series_receiver/). This works in L- band. Two GPS signals are transmitted on L1 = 1575.42 MHz and L2 = 1227.60 MHz frequencies. The raw data from the receiver is converted into Receiver Independent Exchange (RINEX) format using Convert4 software. The receiver worked from the year 2013 to 2016. The 2014 year's GPS data is considered for this work. A Mini boundary layer mast (MBLM) is available at VBIT. Hourly values of Pressure (P), Temperature (T), and Relative Humidity (R.H.) are taken from the MBLM. The data description is explained in Nirmala Bai et al. (2019). The meteorological data between December 2013 and November 2014 were obtained from Mini Boundary Layer Mast (MBLM), operating on the same campus. The P, T, and R.H. are used as input to the models, and IWV is the target output. GPS analysis at MIT (GAMIT) software version 10.61 estimates IWV values http://geoweb.mit.edu/gg/. GAMIT software consists of a collection of programs that works on phase data to estimate three-dimensional relative positions of ground stations and satellite orbits, atmospheric zenith delays, and earth orientation parameters. The software is designed to run under UNIX operating system. We have used LINUX 14.04 LTS operating system for GAMIT software. GAMIT software inputs GPS and meteorological data and calculates tropospheric dry, wet delays, and P.W values. IWV values are calculated in Nirmala Bai et al. (2019).

Tropospheric delay of the signal is caused due to the changing refractive index along the signal's path, which depends on the meteorological parameters. The transmission delay in the zenith path, Zenith Total Delay (ZTD) present in the troposphere contains two parts: dry or hydrostatic delay (ZHD) is of all atmospheric constituents except water vapor, and wet delay is produced with the water vapor present in the lower troposphere (ZWD). ZTD is obtained from the GPS data by using the Vienna mapping function. ZHD is calculated using three models, e.g., the Saastamoinen model, the Hopefield model, and the Black model (Saastamoinen, 1973; Hopfield, 1971; Black, 1978). GAMIT software incorporates the most widely used Saastamoinen model for data processing. (Bai and Feng, 2003) has suggested the following equations for the delay calculation using this model (Ds).

$$D_S = 0.2277 \frac{P}{F(\varphi,H)} \quad (1)$$

$$F(\varphi,H) = 1 - 0.0026Cos(2\varphi) - 0.00028H \quad (2)$$

where the latitude of the station (radian) is represented by φ, H is the height from the mean sea level (km), and the pressure is represented by P (hPa). Using ZTD, the wet delay (ZWD) is calculated from the following equation,

$$ZWD = ZTD - ZHD \quad (3)$$

From equation (3) PWV can be calculated as

$$PWV = \prod(T_m)ZWD \quad (4)$$

where, $\prod(T_M)$ is a dimensionless quantity. T_m is the weighted mean temperature of the atmosphere which is helpful for the computation of $\prod(T_m)$. Error in the surface pressure measurements influences PWV

values computed with the Saastamoinen model and is ~1.01 mm (Bai and Feng, 2003).

A default 10° angle has been taken for the data processing in the GAMIT software for the zenith cut-off angle. If the angle is more than 10° gives dry bias in the PWV calculation (Emardson et al., 1998). Following equation (5) the *IWV* values are obtained from the *PWV* as,

$$IWV = PWV \times d \quad (5)$$

where *d* is the density of liquid water.

The obtained IWV values can then be compared with the prediction's RQ-GPR model and NN model.

3. Gaussian Process Regression (GPR)

A neural network is a group of neurons in many layers. A neural network is also called an artificial neural network (ANN). ANN is an artificial adaptive system. They can think like human brains. ANN networks can change their operation based on the required application. Here we have chosen a medium neural network that works with increasing first layer setting.

Neural network-based algorithms are widely used in engineering to improve the prediction problem's solutions (Wang and Alrueyemi, 2021). However, the main drawback of using this algorithm is overfitting which can be solved with weight adjustments (Williams, 1998; Uusitalo, 2007).

The Bayesian network could be the better option for the neural network solution, which uses the Bayesian interface to evaluate the probability (Recknow, 1999). But this method shows uncertainty.

Gaussian process regression shows a better solution to this uncertainty. Advantages of the RQ-GPR method over Bayesian are simplicity, nonlinearity, and straightforward generalization (Wang and Alruyemi, 2021). Gaussian process (G.P.) is a random process containing arbitrary parameters with fixed numbers that carry a joint Gaussian distribution. A GP is described by mean function μ(x) and covariance function *C(x, x')*. The equations used in this study are taken from the following (Hong et al., 2015).

$$f(x) \sim GP(\mu(x), C(x, x')) \quad (6)$$

$$\mu(x) = E[f(x)] \quad (7)$$

$$C(x, x') = E[(f(x) - \mu(x))(f(x') - \mu(x'))] \quad (8)$$

where $(x,\ x') \in X$ are random variables. X represents the input parameters like Pressure, Temperature, Relative Humidity, IWV, etc., applied to the machine learning algorithms.

Given priori information related to the GP and a sequence of training points

$D = \{(x_i, f(x_i)) \mid i = 1, 2, \ldots\ldots, n\}$, $x_i \in R^d$ (*n* is dataset number. Rassmusen and Williams (2006) have given the posterior distribution by applying the limitation on joint priori distribution. A GP model, along with noise,

$$y_i = f(x_i) + \varepsilon \quad (9)$$

where e, ε is called additive Gaussian noise *N* $(0, \sigma^2)$.

A new test input *x**, training output *y* and test output *y** can be created with Joint Gaussian priori distribution as

$$\begin{bmatrix} y \\ y^* \end{bmatrix} \sim N\left(0, \begin{bmatrix} C(X,X) + \sigma_n^2 I & C(X,x^*) \\ C(X,x^*) & C(x^*,x^*) \end{bmatrix}\right), \quad (10)$$

where, *C(X, X)* is n - order symmetric positive definite covariance matrix, *C(X, x*)* is the *n* x 1 covariance matrix of the test input *x**, and the training input *X* and covariance *C(x*, x*)* is the covariance matrix of the test input *x**.

From the posterior probability formula for a given input *x** and the training set *D*, the *GP* can evaluate output *y**

$$y^* | x^*, D \sim N\left(\mu_{y^*}, \sigma_{y^*}^2\right) \quad (11)$$

$$\mu_y^* = C(x^*,\ X)\,(C(X,X) + \sigma_n^2 I)^{-1}\, y = \sum_{i=1}^{n} \alpha_i\, C\left(x^i,\ x^*\right) \quad (12)$$

where, μ_{y^*}, σ_{y^*} are expectation and variance of *y** and $\alpha = (C + \sigma^2{}_n I)^{-1} y$, where *I* is the identity matrix. A covariance function is the main parameter responsible for creating the functions used to provide the relationship between the data.

There are many covariance functions like Squared Exponential (S.Q.), Rational Quadratic (R.Q.), and Matern class of covariance (Rassmussen and Williams, 2006; Liu et al., 2009). In the present study, R.Q. covariance function is used. It is described below

$$C_{RQ}(x^i, x^j) = \sigma_f^2 \left(1 + \frac{(x^i - x^j)^2}{2\alpha l^2}\right)^{-\alpha} \quad (13)$$

where α, l,σ_f^2, M are hyperparameters, i and j represent the i^{th} and j^{th} vector in the input matrix [X].

The performance of the selected methods can be estimated by using Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R-Squared (R^2). Formulas are described below, following (Alghamdi et al., 2020).

$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \quad (14)$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{N}} \quad (15)$$

$$\text{MAE} = \frac{\sum_{i=1}^{N}|\hat{y}_i - y_i|}{N} \quad (16)$$

$$\text{MAPE} = \frac{1}{N}\sum_{i-1}^{N}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100\ \% \quad (17)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2} \quad (18)$$

where y_i is the observed IWV, $\hat{y}_i$ is the predicted IWV, $\bar{y}_i$ is the mean IWV, and N is the number of data samples.

The covariance matrix C (x, x') contains the IWV data derived from the GPS measurements containing no. of hours of the day and day of the year. The mean function $\mu(x)$ represents the average value of the taken IWV data and covariance function C (x, x') have been derived using IWV data. The covariance function derives the IWV points that are the same as predicted IWV value and have similar outputs.

4. Results and Discussion

We have used the Rational Quadratic GPR (RQ-GPR) model and the Medium Neural network (NN) to predict IWV. Meteorological data of Hyderabad station is given as input and output as IWV values to the models as mentioned above. The met data includes Year, DOY, Hour, Pressure, Temperature, Relative Humidity, and IWV. The cross-validation data splitting method is adopted in our analysis to validate the proposed model accurately and to overcome the shortcomings of the bootstrap method (Burman (1989); Kohavi (1995); Kim (2009)). The data samples considered as input to the proposed model during 2014 (January to December) are split into training and testing data samples using the cross-validation method, respectively, with 85% and 15%. The methodology implemented with different algorithms is illustrated in the flowchart, as shown in Fig. 1.

Figure 2 shows the details of the input data and output data. GPS data from the GPS receiver and meteorological data from the MBLM are considered to calculate IWV values at the VBIT station. The met data used for the processing of IWV are Pressure, Temperature, and Relative Humidity have been taken as input. Fig. 2 contains (a) Time series plot of the pressure data from the MBLM for 2014. (b) Time series plot of the Temperature data from the MBLM data for 2014. (c) Time series plot of the Relative Humidity for the year 2014. (d) GPS-derived IWV values with meteorological data for The year 2014. MBLM gives pressure, temperature, and relative humidity data every hour. Pressure data is low during summer time and high during winter. Temperature is high during pre-summer and summer. Low relative humidity is observed during summer. IWV values were observed high during summer compared to other seasons.

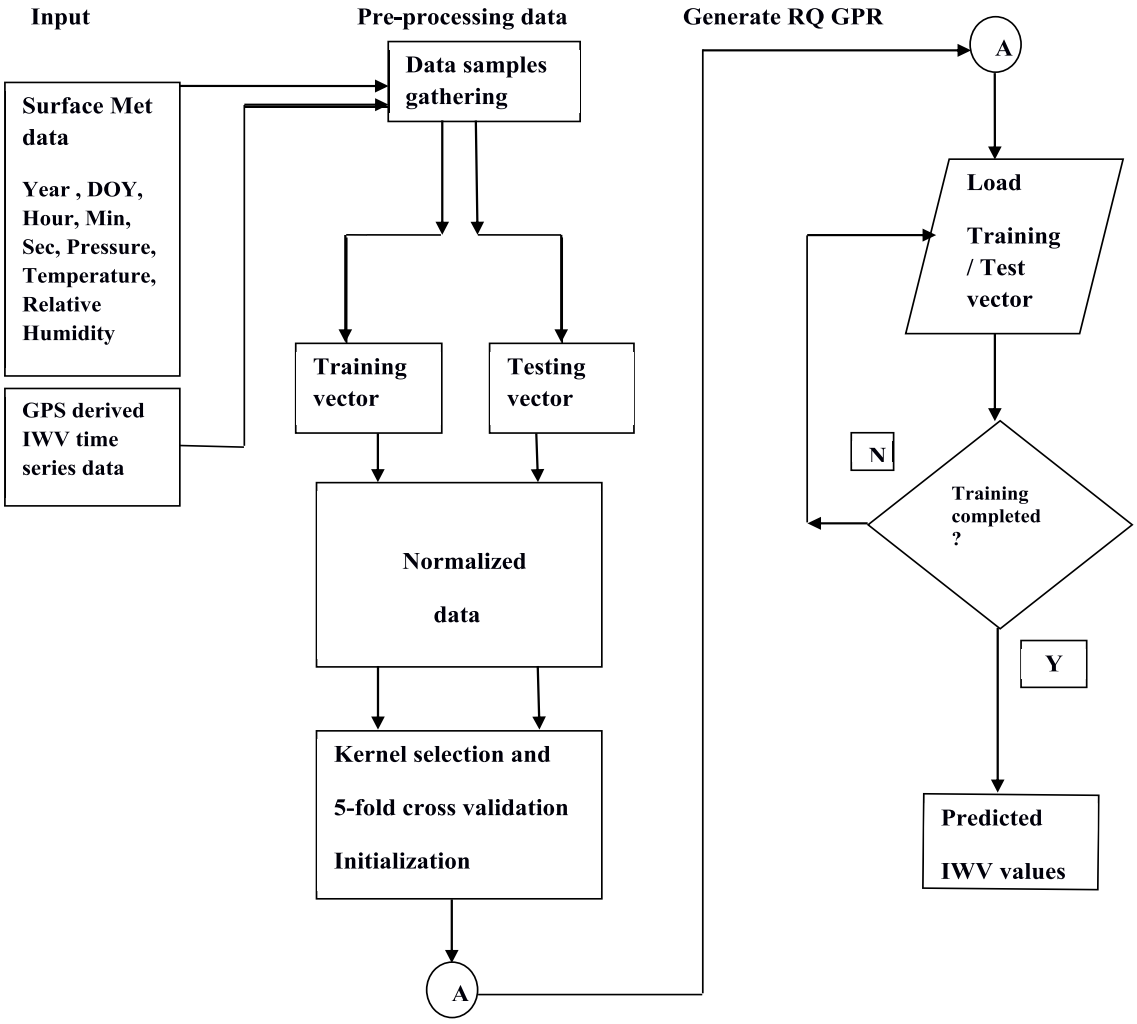


Figure 1. Flowchart for estimation of PWV using RQ-GPR and NN algorithm

Figure 3 shows the predicted IWV values plotted with the RQ-GPR and NN models. IWV values predicted from the RQ-GPR model show minor deviation from the NN model-derived IWV values. Green color gives RQ-GPR predicted IWV values, and the NN model predicted IWV values are represented with black. Both the models show high IWV values during the summer compared to other seasons.

Residual is described as the error between the predicted data and the actual data. Figure 4 shows the residuals calculated by taking the difference between each actual and estimated IWV value. The variation of the RQ-GPR model shows between +12% to -23%, whereas, for the NN model, the variation is +15% to -29%. Thus, the RQ-GPR model is showing minor errors compared to the NN model.

In Fig. 4, the two models showed a low positive residual, indicating that the observed value is more than the predicted value. In contrast, the negative residual indicates that the observed value is less than the predicted value. Thus, the RQ-GPR model has fewer errors than the NN model.

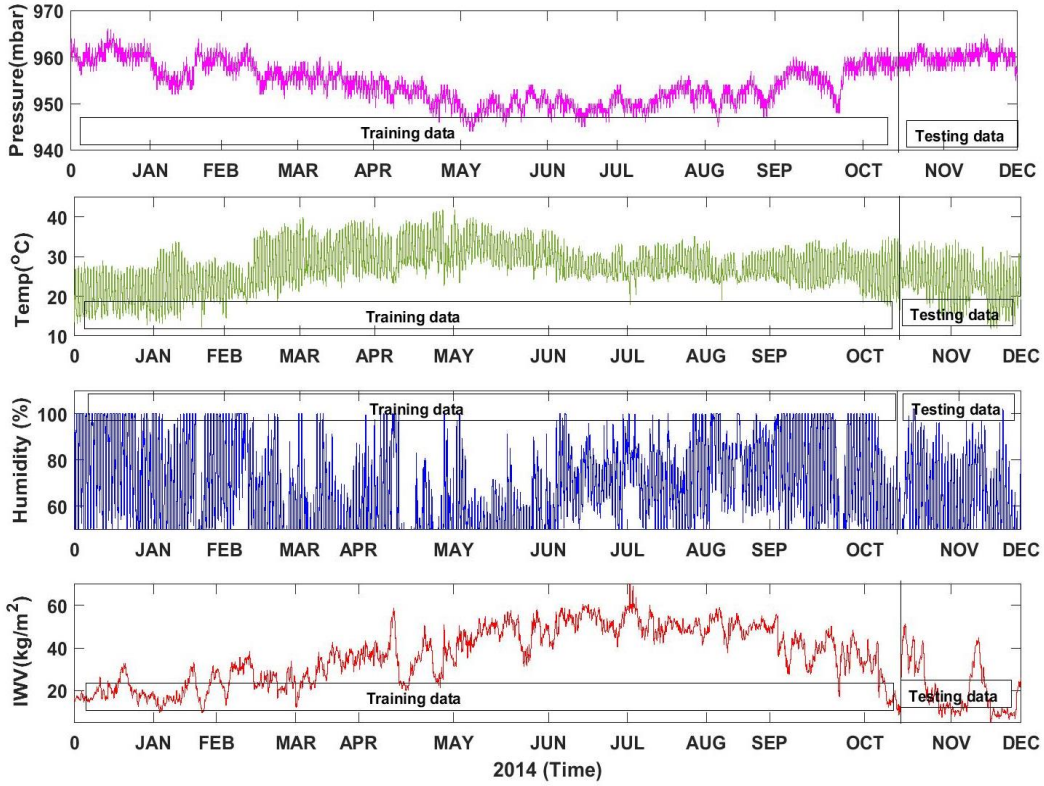


Figure 2. Time series plot of a) Pressure (Pink color), b) Temperature (Green color), c) Relative Humidity (Blue color), and d) GPS-derived IWV (Red), for the year 2014, including Training and Testing period. Data Available at VBIT, Hyderabad

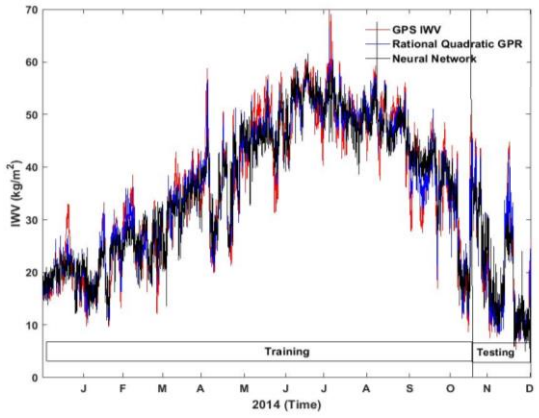


Figure 3. The time series plot of RQ-GPR predicted IWV (Blue) and NN model (Black) for 2014 at VBIT, Hyderabad

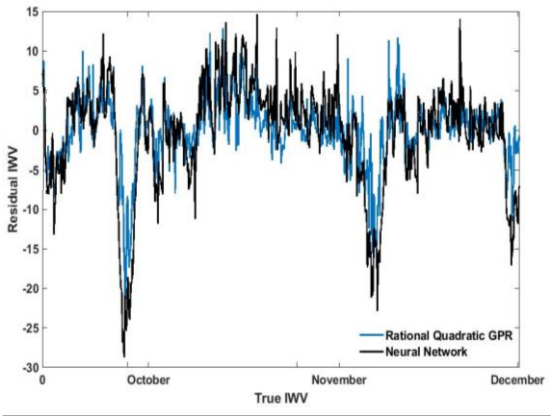


Figure 4. Residual plot of RQ-GPR model (Blue) and NN mode is with (Black) for the testing period only i.e., from October, November and December 2014 at VBIT Hyderabad

Fig. 5 shows the time series plot of GPS-derived IWV (Red), RQ-GPR-predicted IWV (Blue), and NN model predicted (Black) for the year 2014 that including the training and

testing period. Data from January to December 2014 were divided into training and testing periods. The training period is 85%, and the testing is 15%. Fig. 5 shows the predicted IWV values plotted with the RQ-GPR model and NN model against GPS-derived IWV values. Figure 5 indicates the variation of IWV values during the testing period. IWV values predicted from the RQ-GPR model are followed by GPS-derived IWV values.

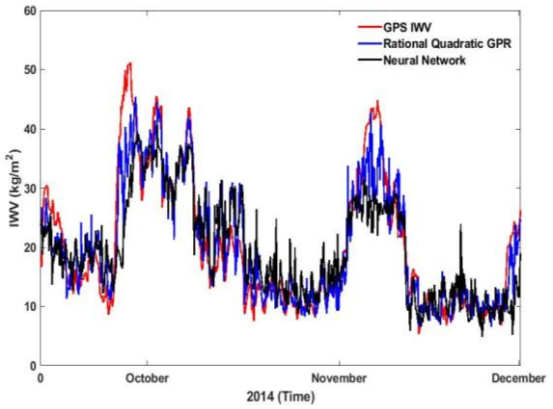


Figure 5. Time series plot of GPS derived IWV (Red), RQ-GPR predicted IWV (Blue), and NN model predicted (Black) during the testing period, i.e., during October, November, and December 2014 at VBIT, Hyderabad

NN model predicted IWV values show some deviation from the actual IWV values. The blue line represents the GPS-derived IWV values, whereas the green and red colors give RQ-GPR predicted and NN model predicted IWV values, respectively. It is concluded that the RQ-GPR model is the most suitable IWV prediction model using GPS observations.

The validation of these models can also be seen from the correlation plots in Fig. 6(a) and Fig. 6(b). The details of the two models are given in Table 1. The scatter plots of IWV values were predicted using RQ-GPR and NN models, as illustrated in Fig. 6. The IWV values estimated with the RQ-GPR model match the actual data. On the other hand, the NN model shows quite a deviation from the actual data compared with the RQ-GPR model. The correlation coefficient (ρ) obtained with the RQ-GPR and NN model for the testing set is 0.93 and 0.86, respectively.

The performance of the proposed model is evaluated using the statistical analysis metrics functions such as mean average error (MAE), mean absolute percentage error (MAPE), root mean squared error (RMSE), and R^2 (Gao et al. (2017); Alghamdi et al. (2020); Wang and Alruyemi (2021)). The MAE shows how models overestimate or underestimate the measured values. The RMSE describes the average difference between the predicted value and measured value, the MAPE defines the accuracy of the models by error percentage, and the coefficient of determination R^2 describes the degree of association between the predicted and the measured values.

The mean difference between the predicted and measured values is described as RMSE following (Gao et al., 2017). The RQ-GPR model calculated RMSE value is 4.2598 kg/m^2, less than the NN model with 5.2449 kg/m^2. Estimation of the measured values is given by Mean Absolute Error (MAE). The MAE value obtained from RQ-GPR is less than the NN model. The Mean Square Error (MSE), MAPE, and R-squared (R^2) also give better accuracy than the NN Model. The details of the parameters are described in the Table 1.

Table. 1 IWV prediction using RQ-GPS and NN models

MODELS	MSE (kg/m^2)	RMSE (kg/m^2)	R^2	MAE (kg/m^2)	MAPE (%)
NN MODEL	27.509	5.2449	0.86	3.9109	13.45
Rational Quadratic GPR (RQ-GPR)	18.146	4.2598	0.93	3.0762	8.97

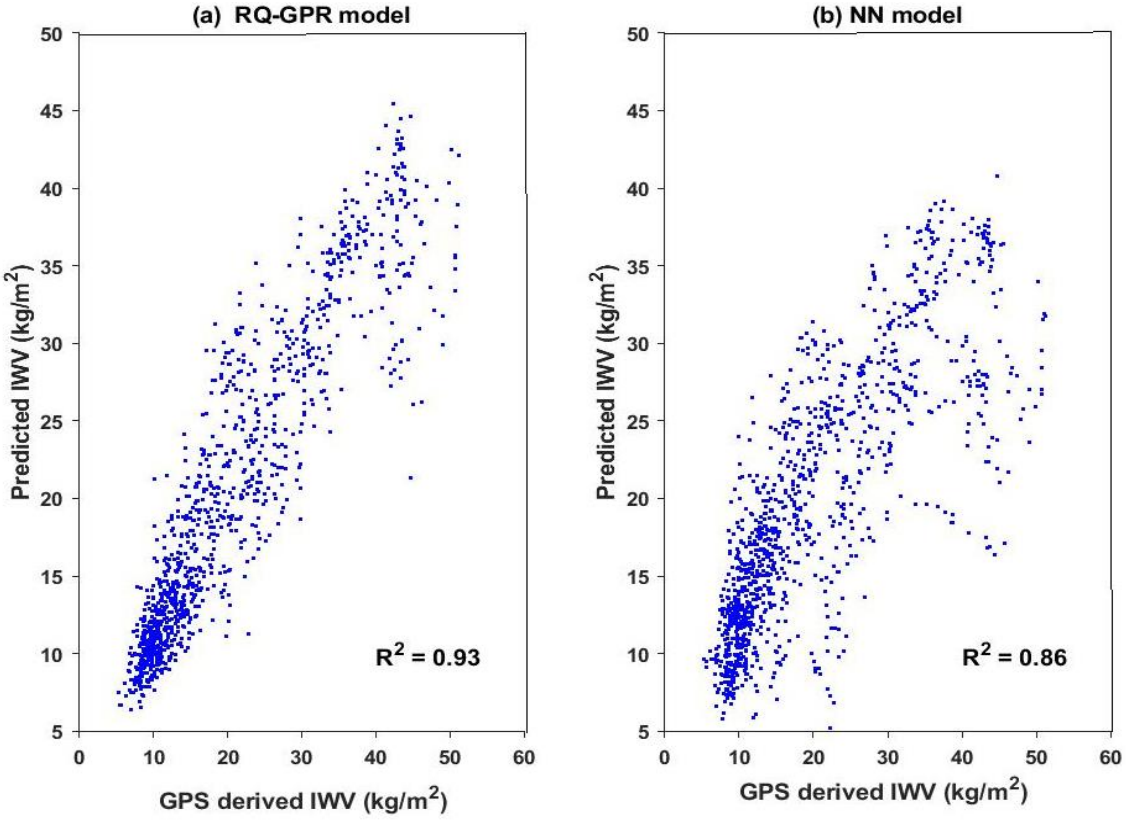


Figure 6. Scatter plots of IWV values predicted using RQ-GPR and NN model for 2014 at VBIT, Hyderabad

4.1. Prediction of IWV during HUDHUD cyclonic storm

Short-term changes in the IWV cause severe effects in the atmosphere, leading to rainstorms, snowstorms, cyclonic storms, and depression. We have considered data on HUDHUD cyclone that occurred in the year 2014. Hudhud was a severe cyclonic storm that developed over the Bay of Bengal from 7th-14th October 2014. On 7th October, it started with depression with a wind speed of 25-30 kt, and it changed its state to Cyclonic Storm on 8th October with a wind speed of 40-45 kt. Then, the Severe Cyclonic Storm (SCS) was observed with a 45-60 kt surface wind speed on 9th October and Very Severe Cyclonic Storm (VSCS) on 10th October and 11th October with a surface wind of 60-100kt. Finally, it reached landfall on 12th October 2014, at Vishakhapatnam, with a surface wind of 100-40kt. A detailed description of the cyclone is described in Table 2.

HUDHUD cyclonic storm has been analyzed by taking GPS and Met data for August, September, and October 2014. August and September data is considered training data, and October data is testing. Fig. 7 shows the relationship between IWV and wind values, indicating that the IWV and wind values are against every day during the event. The variation indicates that the cyclone started on 7th October 2014, with Deep depression (DD). The IWV values varied slowly, and there was an increment in the wind data.

Table 2. Hudhud cyclonic storm description

DAY	Cyclone Intensity Number	Surface Wind (kt)	Cyclone Category
07th October,2014	1.5-2	25-30	D-DD
08th October, 2014	2-3	30-45	DD-CS
09th October, 2014	3-3.5	45-55	CS-SCS
10th October, 2014	3.5-4	60-75	SCS-VSCS
11th October, 2014	4-5	75-100	VSCS
12th October, 2014	------	100-40	VSCS-CS
13th October, 2014	------	30-25	DD-D
14th October, 2014	------	25-20	D-low pressure

Deep Depression (DD) turned into a Severe Cyclone (SC) and severe cyclone storm (SCS). The wind data also increased, whereas IWV values decreased during the Severe Cyclonic Storm (SCS). IWV values were shallow during the Very Severe Cyclonic Storm (VSSC), and the wind values were tremendous. From the figure, it can be concluded that the IWV and wind values show opposite trends during cyclonic storms. From Fig. 7, RQ-GPR model-derived IWV values show good matching with the IWV data computed from GPS against NN derived model.

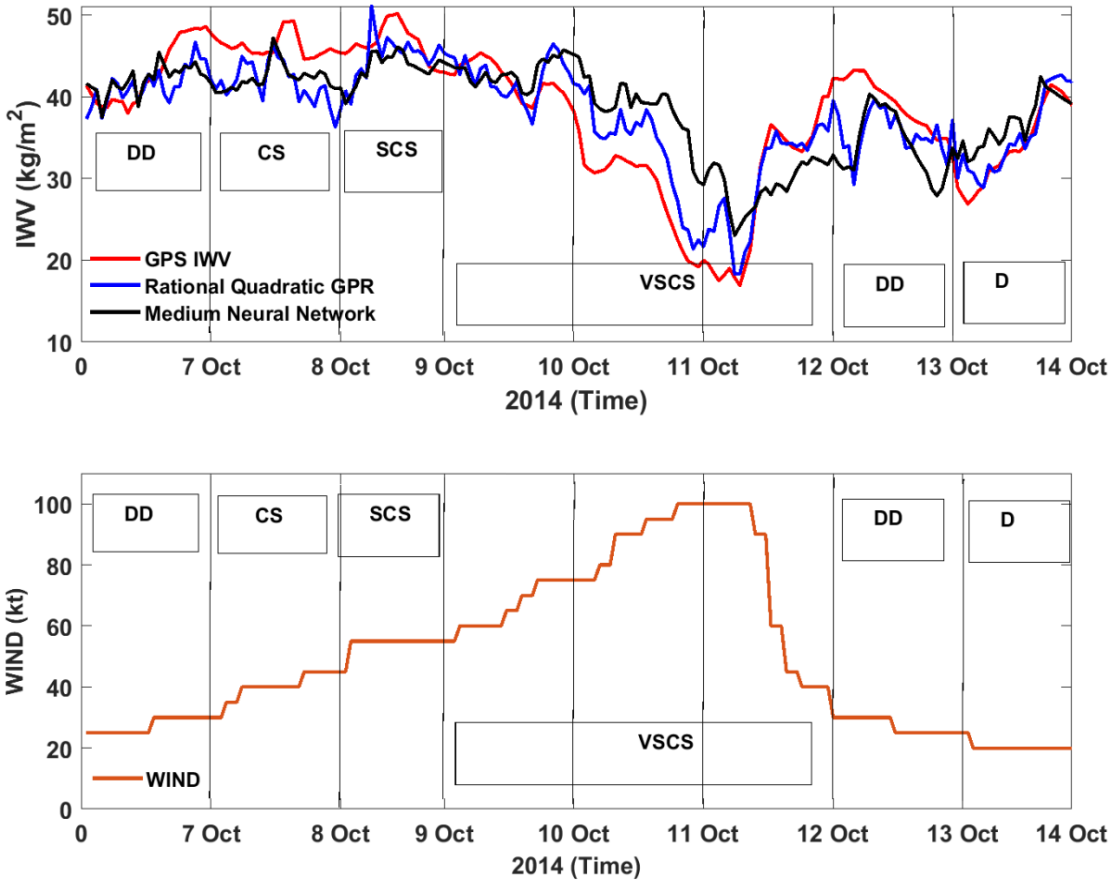


Figure 7. a) Time series plot of HUDHUD cyclone analysis from 7 October to 14 October over Hyderabad using GPS calculated IWV (Red), RQ-GPR model predicted IWV (Blue), and NN model IWV (Black). b) Wind values plotted during the HUDHUD cyclone

Figure 8 shows the time series plot of all three data sets. IWV data calculated from the GPS is the actual data calculated using GAMIT software used as the reference data. August and September month IWV variations are quiet. During the cyclone, the variation is more before the peak moment of the cyclone. When it reached the Very Severe Cyclonic Storm (VSCS) due to the heavy rainfall, the IWV values decreased drastically. Again, during the depression and landfall time, the IWV values have increased to the typical values. Fig. 9 shows the scatter plot of the variation of IWV during the HUDHUD cyclonic storm. The data from August 2014 to October 2014 were considered for the analysis. 85% of the data is taken for training, while the last 15% is considered for testing. With this criterion, the HUDHUD cyclone has been studied. The GPS-derived IWV values correlate with the RQ-GPR model in Fig. 9 (a) whereas GPS-derived IWV values correlate with the NN model in Fig. 9 (b).

The correlation coefficients (ρ) obtained for RQ-GPR and NN models) is 80% and 60%, respectively. With RQ-GPR, a maximum matching can be seen with GPS-derived IWV data during the cyclone period, whereas only 60% correlation matching is noticed for the NN model. RMSE error for the RQ-GPR model is also less than the NN model, as mentioned in Table 3.

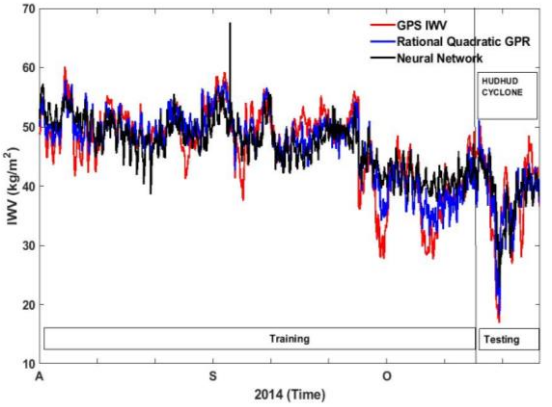


Figure 8. Time series plot of HUDHUD cyclone analysis by taking 3 months data prior to the cyclone and on the cyclone time (August, September and October) over Hyderabad using GPS calculated IWV (Red), RQ-GPR model predicted IWV (Blue) and NN model IWV (Black)

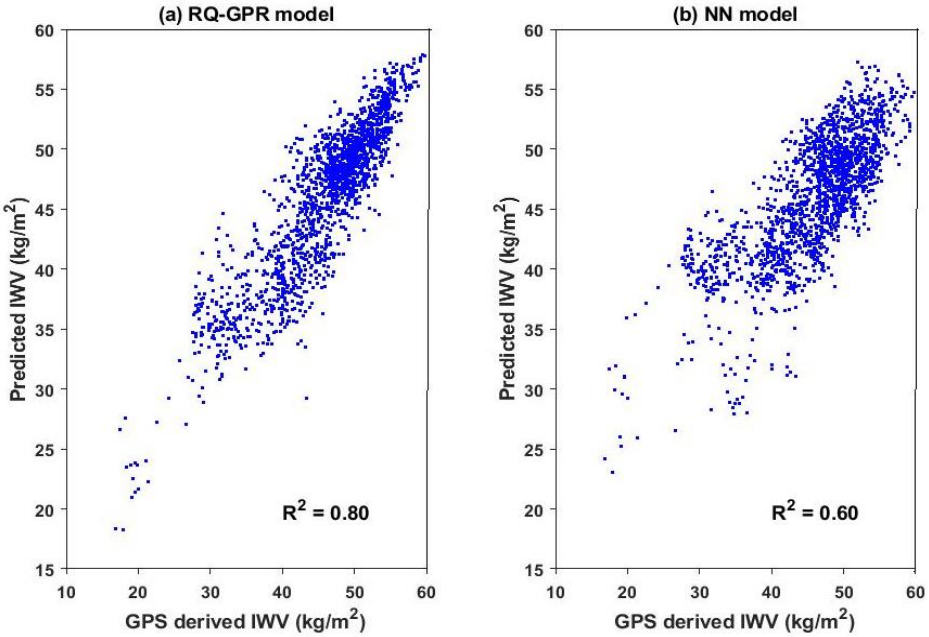


Figure 9. is the scatter plots of GPS-derived IWV values predicted by using the RQ-GPR and NN model during HUDHUD cyclone

Table 3. Prediction of IWV during HUDHUD cyclonic event

MODELS	RMSE (kg/m^2)	R^2
Rational Quadratic GPR (RQ-GPR)	2.837	0.80
NN MODEL	3.327	0.60

5. Conclusions

This study has shown the efficacy of the GPR technique as a new model for forecasting in climate and weather forecasting applications, especially in predicting IWV. This model is also helpful for the regions where the weather is similar to us. We have considered the Rational Quadratic GPR (RQ-GPR) model, which correlates with the NN model for predicting IWV. In the present work, the analysis has been considered for the year 2014. Data for January to December months. The models' accuracy is evaluated using performance prediction metrics like MAE, MSE, RMSE, and R^2. It is found that the RQ-GPR model accuracy is better than the NN model. Analysis shows that the RQ-GPR model obtained an R^2 of 0.93, much higher than the NN model's R^2 of 0.86.

RQ-GPR predicted values are very much closer to actual data. Using the RQ-GPR and NN model, we have taken the HUDHUD cyclonic event to predict the IWV values. RQ-GPR-derived IWV values show more accuracy than the GPS-derived IWV values, whereas; the NN model shows more deviation. Correlation studies have been done during HUDHUD cyclonic events. RQ-GPR algorithm shows a good match of 80% correlation with GPS-derived IWV. NN model gives only a 60% correlation, which is unsuitable for forecasting. Many researchers have used other models like ANFIS model, decision tree, and random forest, but they have drawbacks when selecting the parameters for better accuracy. In future work, more atmospheric events will be investigated for predicting IWV. The prediction results would help to investigate the climate and weather conditions for providing necessary weather alerts. If GPS data is unavailable in remote areas, this model is constructive for estimating IWV.

Acknowledgements

Two authors, Ms. Nirmala Bai Jadala and Dr. Gopa Dutta thank the National Atmospheric Research Laboratory (NARL), Department of Space, India, for installing a GPS/GNSS receiver at VBIT. The computational work for the research has been carried out in the computational laboratory developed by DST's FIST grant. We thank the India meteorological department, Hyderabad, for providing cyclone information.

References

Ahmed S. Alghamdi, Polat K., Alghoson A., Alshdadi A.A., Ahmed A., EL -Latif E.L.A., 2020. Gaussian Process Regression (GPR) based non-invasive continuous blood pressure prediction method from cuff oscillometric signal. Applied Acoustics., 164, 107256.

Aiguo Dai, Meehl G.A., Washington W.M., Wigley T.M.L., Arblaster J.M., 2001. Ensemble simulation of twenty-first century climate changes: Business-as-usual versus CO2 stabilization. Bull. Am. Meteorol. Soc., 82(11), 2377-2388. https://www.jstor.org/stable/26215289.

Aiguo Dai, Wang J.H., Ware R.H., Van Hove T., 2002. Diurnal variation in water vapor over North America and its implications for sampling errors in radiosonde humidity. J. Geophys. Res., 107(D10), 4090. https://doi.org/10.1029/2001JD000642.

Allan C. Just., Liu Y., Sorek-Hamer M., Rush J., Dorman M., Chatfield R., Wang Y., Lyapustin A., Kloog I., 2020. Gradient boosting machine learning to improve satellite-derived column water vapor measurement error. Atmos. Meas. Tech., 13, 4669-4681. https://doi.org/10.5194/amt-13-4669-2020.

Black H.D., 1978. An easily implemented algorithm for the tropospheric range correction. J. Geophys. Res., 83(B4), 1825-1828.

Burman P., 1989. A comparative study ordinary cross-validation, v-fold cross validation and the repeated

learning-testing methods. Biometrika., 83(B4), 1825-1828.

C.E. Rasmussen, Williams C.K.I., 2006. Gaussian processes for machine learning. The MIT press, Cambridge, MA. ISBN 026218253X.

C.K.I. Williams, 1998. Prediction with Gaussian process: from linear regression to linear Prediction and beyond, in learning in graphical models. Springer.

Christian Ruckstuhl, Philipona R., Morland J., Ohmura A., 2007. Observed relationship between surface specific humidity, integrated water vapor, and longwave downward radiation at different altitudes. J. Geophys. Res., 112, D03302. Doi: 10.1029/2006JD007850.

G. Lanyi, 1984. Tropospheric delay effects in radio interferometry. TDA progress report., 42-78, 152-159.

H.S. Hopefield, 1971. Tropospheric effect on electromagnetically measured range: prediction from surface weather data. Radio. Sci., 6(3), 357-367. https://doi.org/10.1029/RS006i003p00357.

Hong Liang, Cao Y., Wan X., Xu Z., Wang H., Hu H., 2015. Meteorological applications of precipitable water vapor measurements retrieved by the national GNSS network of China. Geodesy and Geodynamics, 2, 135-142. http://doi:org/10.1016/j.geog.2015.03.001.

Isaac M. Held, Soden B.J., 2000. Water vapor feedback and global warming. Annu. Rev. Energy Environ., 25, 441-475. https://doi.org/10.1146/annurev.energy.25.1.441.

J.K. Kiehl, Trenberth K.E., 1997. Earth's Annual Global Mean energy budget. Bull. Am. Meteor. Soc., 78, 197-208. https://doi.org/10.1175/1520-0477(1997)078%3C0197:EAGMEB%3E2.0.CO;2.

J. Saastamoinen, 1973. Contributions to the theory of atmospheric refraction Part II, Refraction corrections in satellite geodesy. Bull. Geodes, 25, 1935-1948.

J.T. Houghton, Ding Y., Griggs D.J., Noguer M., Vander Winden P.J., Dai X., Maskell K., Johnson C.A., 2001. Climate Change 2001. The Scientific Basis. Contribution of Working Group 1 to the Third Assessment Report. Cambridge Univ. Press, New York, 881p.

Jingping Duan, Bevis M., Fang P., Bock Y., Chiswell S., Businger S., Rocken C., Solheim F., Hove T.V., Ware R., Mcclusky S., Herring T.A., King R.W., 1995. GPS meteorology: Direct of the absolute value of precipitable water. J. Applied. Meteorology, 35, 830- 838. https://www.jstor.org/stable/26187904.

K. Liu, Liu B., Xu C., 2009. Intelligent analysis model of slope nonlinear displacement time series based on genetic-gaussian process regression algorithm of combined kernel function. Chinese Journal of Rock Mechanics and Engineering, 10, 2128-2134.

Kenneth H. Recknow, 1999. Water quality prediction and probability network models. Canadian Journal of Fisheries and Aquatic Sciences, 56, 1150-1158. https://doi.org/10.1139/f99-040.

Kevin E. Trenberth, Dai A.G., Rasmussen R.M., Parsons D.B., 2003. The changing character of precipitation. Bull. Am. Meteorol. Soc., 84(9), 1205-1217. https://doi.org/10.1175/BAMS-84-9-1205.

Kim J.H., 2009. Estimating classification error rate: Repeated cross-validation, Repeated hold-out and bootstrap. Computational statistics and data Analysis., 53, 3735-3745.

Kohavi R., 1995. A study of cross validation and bootstrap for accuracy estimation and model selection. International joint conference on artificial intelligence, 14(2), 1137-1145.

Lara Uusitalo, 2007. Advantages and challenges of bayesian networks in environmental modelling. Ecological Modelling, 203, 312-318.

Mayank Jain, Manandhar S., Lee Y.H., Winkler S., Dev S., 2020. Forecasting Precipitable Water Vapor using LSTMs. IEEE USNC-CNC-URSI North American Radio Science Meeting, 147-148. http://doi:10.23919/USNC/URSI49741.2020.9321614

Michael Bevis, 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. J. Geophys. Res. Atmos., 97, 15787-15801. https://doi.org/10.1029/92JD01517.

N.B. Jadala, Sridhar M., Dashora N., Dutta G., 2020. Annual, seasonal and diurnal variations of integrated water vapor using GPS observations over Hyderabad, a tropical station. Adv. Space Res., 65, 529-540. http://doi.org/10.1016/j.asr.2019.10.002.

Nguyen Ngoc L., Coleman R., 2021. Assessment and reduction of zenith path delay biases due to day boundary effect. Vietnam Journal of Earth Sciences, 43(4), 546-554. https://doi.org/10.15625/2615-9783/16587.

Pedro Benevides, Catalao J., Nico G., 2019. Neural Network approach to forecast hourly intense rainfall using GNSS precipitable water vapour and meteorological sensors. Remote Sens., 11(8), 966.

Senkal O., et al., 2012. Precipitable Water vapour modelling using Artificial neural network in Cukurova region. Environ Monit Assess (Springer), 184, 141-147. http://doi.org/10.1007/s10661-011-1953-6.

Sheng Hong., Zhou Z., Lu C., Wang B., Zhao T., 2015. Bearing remaining life prediction using Gaussian process regression with composite kernel functions. J. Vibro Engineering, 75, 695-704.

Subimal Ghosh, 2010. SVM-PGSL coupled approach for statistical downscaling to predict rainfall from GCM output. J. Geophys. Res., 115, D22102.

T. Ragne Emardson, Elgered G., Johansson J.M., 1998. Three months of continuous monitoring of atmospheric water vapor with a network of Global positioning system receivers. J. Geophys. Res., 103, 1807-1920.

Vladislav V. Kalinnikov, Khutorova O.G., 2017. Diurnal variations in integrated water vapor derived from a GPS ground network in the Volga-Ural region of Russia. Ann. Geophys., 35, 453-464,

Wang B., Alruyemi I., 2021. Comprehensive Modeling in Predicting Biodiesel Density Using Gaussian Process Regression Approach. BioMed Research International, 1-13. https://doi.org/10.1155/2021/6069010.

Wayan Suparta., Alhasa K.M., 2015. Modeling of zenith path delay over Antarctica using an Adaptive neuro fuzzy inference system technique. Expert Systems with Applications, 42, 1050-1064. https://doi.org/10.1016/j.eswa.2014.09.029.

Wayan Suparta, Samah A.A., 2020. Rainfall prediction by using ANFIS time series technique in South Tengerang, Indonesia. Geodesy and Geodynamics, 11, 411-417, http://doi.org/10.1016/j.geog.2020.08.001.

Wei Gao., Karbasi M., Hasanipanah M., Zhang X., Guo J., 2017. Developing GPR model for forecasting the rock fragmentation in surface mines. Engineering with Computers Springer, 34(3). http://doi.org/10.1007/s00366-017-0544-8.

Yoshua Bengio, Courville A., Vincent P., 2013. Representation learning: A review and new perspectives. IEEE Trans Pattern Anal Mach Intell, 35, 1798-1828. https://doi:10.1109/TPAMI.2013.50.

Zhengdone Bai, Feng Y., 2003. GPS water vapour estimation using interpolated surface meteorological data from Australian automatic weather stations. J. Global. Pos. Sys., 2, 83-89.