Journal of the Korean Society of Surveying, Geodesy, Photogrammetry and Cartography
Vol. 35, No. 5, 423-430, 2017
https://doi.org/10.7848/ksgpc.2017.35.5.423

ISSN 1598-4850(Print)
ISSN 2288-260X(Online)
Original article

Preliminary Study of Deep Learning-based Precipitation Prediction

Kim, Hee-Un[1] · Bae, Tae-Suk[2]

Abstract

Recently, data analysis research has been carried out using the deep learning technique in various fields such as image interpretation and/or classification. Various types of algorithms are being developed for many applications. In this paper, we propose a precipitation prediction algorithm based on deep learning with high accuracy in order to take care of the possible severe damage caused by climate change. Since the geographical and seasonal characteristics of Korea are clearly distinct, the meteorological factors have repetitive patterns in a time series. Since the LSTM (Long Short-Term Memory) is a powerful algorithm for consecutive data, it was used to predict precipitation in this study. For the numerical test, we calculated the PWV (Precipitable Water Vapor) based on the tropospheric delay of the GNSS (Global Navigation Satellite System) signals, and then applied the deep learning technique to the precipitation prediction. The GNSS data was processed by scientific software with the troposphere model of Saastamoinen and the Niell mapping function. The RMSE (Root Mean Squared Error) of the precipitation prediction based on LSTM performs better than that of ANN (Artificial Neural Network). By adding GNSS-based PWV as a feature, the over-fitting that is a latent problem of deep learning was prevented considerably as discussed in this study.

Keywords : Deep Learning, Global Navigation Satellite System, Precipitable Water Vapor, Meteorological Factors, Precipitation Prediction

1. Introduction

Due to a continually changing climate, severe disasters affecting human life and property are increasing year after year. In many countries there have been studies, including GNSS (Global Navigation Satellite System), for better weather forecasting to minimize the damage caused by natural disasters. For example, NOAA (National Oceanographic and Atmospheric Administration), MAGIC (Meteorological Applications of GPS Integrated Column) in Europe, and GEONET (GPS Earth Observing NETwork) of Japan operate the GNSS network for weather forecasting (Kim, 2016).

For accurate measurements, the data from advanced equipment such as WVRs (Water Vapor Radiometer) and radiosondes was used in this study. However, these systems are very expensive to use when acquiring data at a high temporal resolution, resulting in the potential degradation of performance depending on weather conditions. In particular, radiosondes are very sparsely installed in Korea, and in fact do not provide coverage over the entire peninsula. The spatial resolution is relatively low, and furthermore it is limited to only two to four times a day; therefore, continuous information cannot be provided (Kim, 2016). On the other hand, over 100 GNSS CORSs (Continuously Operating Reference Station) are configured across the country, which operate 24 hours a day without interruption. Thus, they can provide information with high spatial and temporal resolutions at lower cost than radiosondes.

In this study, we used GNSS PWV (Precipitable Water

Received 2017. 10. 10, Revised 2017. 10. 12, Accepted 2017. 10. 16
1) Member, Dept. of Geoinformation Engineering, Sejong University (E-mail: khu0716@gmail.com)
2) Corresponding Author, Member, Dept. of Geoinformation Engineering, Sejong University (E-mail: baezae@sejong.ac.kr)

Vapor) for effective weather prediction. For the calculation of PWV, it is required to estimate the amount of ZTD (Zenith Total Delay) in the troposphere and the average temperature of the atmosphere. We processed the GNSS data through scientific GNSS processing software; that is, Bernese GNSS Software V5.2 (hereafter, Bernese). The weighted mean temperature model was estimated for high accuracy of the estimated PWV at the corresponding station. The meteorological parameters were also used by interpolating observed values at nearby AWSs (Automatic Weather Station). By analyzing the relationship between weather behavior and GNSS PWV, it can be clearly seen that GNSS PWV and precipitation are more or less directly connected to each other in terms of increasing/decreasing patterns (Kim, 2016). Therefore, GNSS PWV can be an appropriate factor for weather forecasting, which is used in a deep learning technique.

Deep learning techniques have recently been used in different fields. It is a technique used to find the most suitable model with a large amount of complex data. Therefore, it can be effectively applied to predict the meteorological phenomena occurring in nonlinear combinations through various factors (Bengio *et al*., 2013).

So far, a precipitation prediction was mostly conducted by a machine learning technique such as an ANN (Artificial Neural Network) (Kuligowski and Barros, 1998). Many studies show that an ANN provides greater accuracy than analytical models (Kang and Lee, 2008), but the results are still not accurate enough to predict precipitation.

Precipitation is, in practice, a natural phenomenon that occurs in a complex structure, thus it is better suited to an RNN (Recurrent Neural Network) because it is essentially a form of time series data (Lee and Lee, 2016; Tran and Song, 2017). Therefore, we predicted precipitation through an RNN that is more complicated than an ANN due to the number of hidden layers in the model.

2. Methodology

2.1 Estimation of GNSS PWV

The Earth's atmosphere is divided into two parts, the troposphere and the ionosphere, depending on signal propagation conditions. The troposphere is the lower part of the atmosphere that extends from the surface of the Earth to an altitude of about 20 km. Signals propagated from GNSS satellites are distinguished by drying delays due to dry gases such as carbon dioxide and nitrogen in the troposphere, and wetting delays caused by water vapor (Davis *et al*., 1985).

The ZWD (Zenith Wet Delay) is difficult to calculate accurately due to the nature of the wet component that requires weather information at the GNSS station. Therefore, it is generally calculated by subtracting the ZHD (Zenith Hydrostatic Delay) from the ZTD (Baek *et al*., 2007; Hopfield, 1969). The Saastamoinen model was used to estimate ZHD as given by Eq. (1), which is the same model used to process GNSS data (Saastamoinen, 1972; Elgered *et al*., 1991).

$$ZHD = \frac{(2.2779 \pm 0.0024) \cdot P}{1 - 0.00266 \cdot \cos 2\phi - 0.00026 \cdot h} \quad , \tag{1}$$

where ZHD is the zenith hydrostatic delay, P is the surface pressure, ϕ is the latitude at the observing station, and h is the ellipsoidal height.

The ZWD can be calculated by subtracting the ZHD from the ZTD. Then the ultimate goal of GNSS PWV can be calculated by multiplying ZWD by the conversion factor k which varies according to the weather conditions near the station.

$$GNSS\ PWV\ [mm] = k \cdot ZWD\ [mm] \tag{2}$$

The weighted mean temperature is required to obtain the conversion coefficient k, which is given by Eq. (3) (Kim, 2016).

$$k = \frac{10^6}{\rho \cdot R_v \cdot \left(\frac{k_3}{T_m} + k_2' \right)} \quad , \tag{3}$$

where ρ is the density of water $[kg/m^3]$, and R_v is the gas constant of water vapor $[(m^3 \cdot hPa)/(kg \cdot K)]$, k_2 is 17 ± 10 $[K/hPa]$ and k_3 is $(3.776 \pm 0.004) \times 10^5$ $[K^2/hPa]$.

In this study, we collected meteorological data from seven radiosonde observatories as well as AWS and ASOS in order to determine the weighted mean temperature at the

experimental region (Gangneung, Gangwon-do).

The resulting weighted mean temperature equation was determined by applying linear regression to the calculated values (Kim, 2016; Bevis *et al*., 1992; Ha *et al*., 2016). Eq. (4) represents the weighted mean temperature determined in this study, which is consequently applied to GNSS PWV.

$$T_m = 0.9042 \cdot T_s + 19.77 \quad , \tag{4}$$

where T_s represents the surface temperature.

2.2 Prediction model

A neural network consists of several types of neurons followed by mathematical neuronal circuits. Neurons performing the same role are generally called a layer.

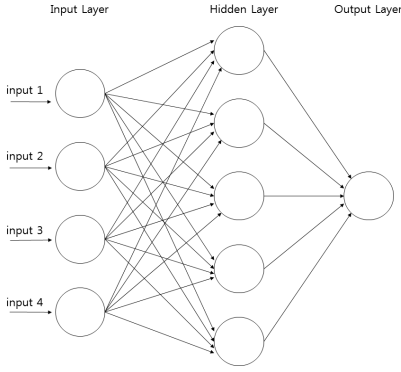

Fig. 1. Neural network algorithm

In Fig. 1, each circle is called a node which indicates the state of the neuron and the input value. The input neuron has an impact on the output value through the activation function.

One example of an activation function used in an RNN is a sigmoid function.

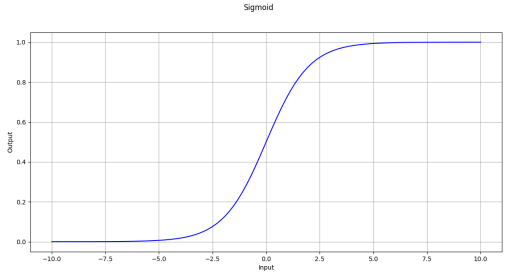

Fig. 2. Sigmoid function

The sigmoid function is a continuous one that is infinitely close to ± 1 when x gets closer to $\pm\infty$, and has a value of 0.5 when x is 0 (see Fig. 2). Similarly to the sigmoid function, there is a hyperbolic tangent function, although this is 0 at $x = 0$.

When data is input in chronological order, for example, time series data such as speech recognition, translation, video, etc., it can be learned and used to predict future values based on past data. Therefore, an RNN is more appropriate for this type of data.

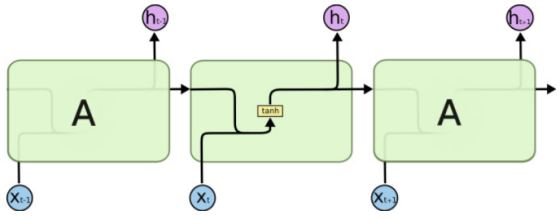

Fig. 3. An RNN (Olah, 2015)

The neural network at stage t reflects the value of the input data x_t and outputs h_t (see Fig. 3). Then it transmits the information to the next stage through a loop. The RNN can link the previous information with the current data in this way. However, if the distance gets longer, it is difficult to connect the information due to increased data.

To improve this process, a modified algorithm called LSTM (Long Short-Term Memory) was suggested. It is a recurrent neural network that can learn from preserved long-term data.

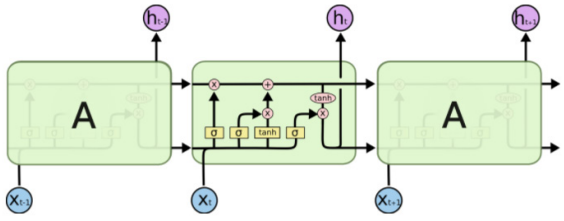

Fig. 4. An LSTM (Olah, 2015)

The LSTM algorithm is composed of a forget layer (Eq. (5)), input layers (Eqs. (6) to (8)), and output layers (Eqs. (9) and (10)). The terms in the forget layer, h_{t-1} and x_t, are determined by the sigmoid function and these decide which information to delete (Olah, 2015).

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad , \tag{5}$$

where f_t is the forget layer, σ is the sigmoid function, W_f is the weight value, h_{t-1} is the value at t-1 of the hidden layer, x_t is input value at t, and b_f is the bias.

Next, the input layers determine which information of the past to store in the cell state through a sigmoid function, and these values can be added to the new cell state through the hyperbolic tangent function.

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{6}$$

$$\widetilde{C_t} = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \tag{7}$$

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \widetilde{C_t} \quad , \tag{8}$$

where i_t is the input layer, $\widetilde{C_t}$ is the vector of new candidate values that can be added to the cell state, C_t is the hyperbolic tangent function, is the updated cell, and b_i, b_C are biases.

Finally, the sigmoid function determines what values to output in the cell state (Eq. (9)). The output value can be determined by multiplying the hyperbolic tangent in the next cell (see Eq. (10)).

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{9}$$

$$h_t = o_t \cdot \tanh(C_t) \quad , \tag{10}$$

where o_t is the output layer based on the sigmoid function, b_o is the bias, and h_t is the determined output value.

2.3 Data processing

As mentioned above, we used the LSTM algorithm, which is suitable for processing long-term time series data to solve the precipitation prediction problem. Precipitation is generally caused by the flow of air that occurs in response to temperature and pressure. Successful precipitration prediction requires various information, including temperature, water vapor pressure, wind speed, humidity, and sea surface pressure. These meteorological factors can be obtained every hour over the web from the relevant meteorological office.

GNSS PWV is the amount of water vapor in the atmosphere, which causes signal delay as they pass through the Earth's atmosphere. We used the CORS data from near the test area to determine the amount of water vapor. Fig. 5 shows the almost synchronized increasing/decreasing pattern between GNSS PWV and precipitation, although the scale of the two factors is slightly different.

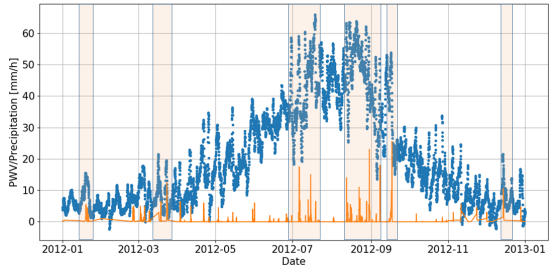

Fig. 5. Comparison between GNSS PWV and precipitation

2.3.1 Procedures

All processing, including deep learning and accompanying data manipulation, was conducted with Python. This program provides many useful packages such as Numpy and Pandas for fast calculation of multi-dimensional arrays and vectorization operations. A data set suitable for the learning model was prepared using Pandas to fit the input data format. The precipitation prediction is basically a supervised learning problem, and thus it creates a model with the correct data (label) to make predictions (see Fig. 6).

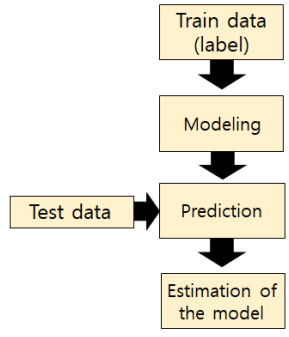

Fig. 6. Processing diagram of supervised learning

A preprocessing step is necessary to generate the input data for the learning model. All features are normalized to scale down for optimal performance of the model. We evaluated the final model by dividing the entire data into two groups, one for training and the other for verification (Fig. 7).

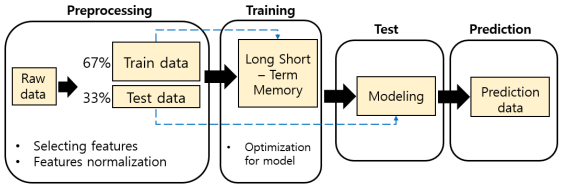

Fig. 7. Flow chart to predict precipitation

2.3.2 Design model

An open source library, Keras, was used to model the precipitation prediction. Since it was first announced in 2015, it has become an intuitive and popular API (Application Program Interface) that makes a neural network simple to implement.

A total of 10 hidden layers were set up for the LSTM in this study, and one output layer is used for precipitation prediction. The precipitation prediction model consists of supervised data. The precipitation at current time (t) can be predicted by taking into account the precipitation and weather conditions at the previous time step (t-1).

	rain(t-1)	temperature(t-1)	wind(t-1)	sea_press(t-1)	vap(t-1)	hum(t-1)	pwv(t-1)	rain(t)
1	0.05454545468	0.17582419515	0.61904716492	0.10196079314	0.01090909075	0.463768095	0.077258632	0.05454545
2	0.05454545468	0.15384617448	0.61255455017	0.10784313083	0.00727272732	0.456521749	0.085198358	0.05454545
3	0.05454545468	0.13186812401	0.61688423157	0.11176471412	0.00727272732	0.376811564	0.09116067	0.05454545
4	0.05454545468	0.13186812401	0.63419914246	0.10980392992	0.00727272732	0.362318814	0.097086087	0.05454545
5	0.05454545468	0.13186812401	0.61471748352	0.11372549832	0.00727272732	0.260869563	0.091173649	0.05454545
6	0.05454545468	0.09890110457	0.60173225403	0.12941177189	0.00363636366	0.311594218	0.099222705	0.05454545
7	0.05454545468	0.13186812401	0.61038970947	0.13725490868	0.01090909075	0.376811564	0.103263624	0.05454545
8	0.05454545468	0.16483518481	0.62121200562	0.13333334029	0.01454545464	0.253623188	0.101241797	0.05454545
9	0.05454545468	0.18681320548	0.61688423157	0.13333334029	0.01454545464	0.347826093	0.111180127	0.05454545

Fig. 8. Example of raw data

Two thirds of the raw data was used to train the model, and the rest of the data was reserved to verify the model. The input data is prepared as LSTM 3D formats - that is, samples, time steps, and features - to define and tailor the model. Each set of input data consists of 7 features along with one time step (Fig. 8). We used the MSE (Mean Squared Error) as the loss function, and the Adam Optimizer for efficient processing.

A total of 20 iterative training sessions were conducted for fitting the precipitation model with a batch size of 15. The difference between the predicted and real precipitation was calculated to evaluate model performance.

3. Results and Analysis

3.1 Comparison with artificial neural network

An LSTM consists of several hidden layers between the input and output layers while an ANN has only one hidden layer. Since the existing ANN does not reflect events that occurred before a specific event, these events cannot contribute to the prediction of future events. Precipitation is usually affected by various factors and occurs under complicated situations. Therefore, the time-series data, such as precipitation, can be more effectively processed with a loop like LSTM. The meteorological features of Korea can be clearly classified due to the distinctive seasonal variation (Sachan, 2014), and thus LSTM works better than a simple ANN in this case. Table 1 shows the RMSE (Root Mean Squared Error) of the precipitation prediction by the two methods. As discussed above, LSTM is more appropriate for predicting precipitation than the machine learning based technique ANN.

Table 1. The RMSE comparison of the two methods

	ANN [mm]	LSTM [mm]
RMSE	0.786	0.668

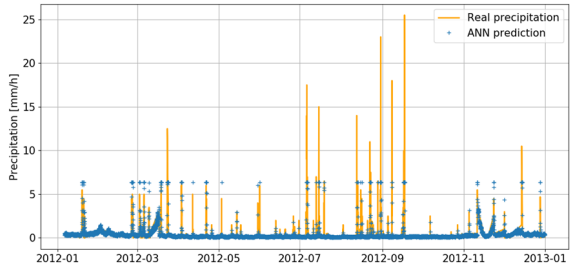

Fig. 9. Artificial Neural Network

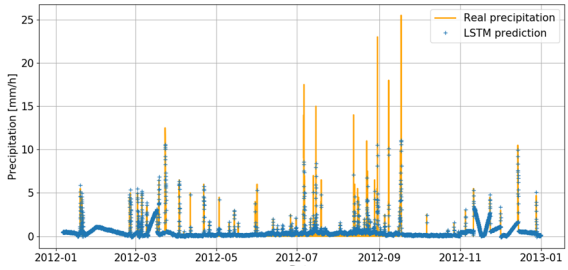

Fig. 10. Long Short Term Memory

Figs. 9 and 10 represent the prediction results by ANN and LSTM. As can be seen in the figures, both methods follow the dominant trends, although they have difficulty in explaining the detailed peaks. Despite the scale of predicted precipitation, LSTM generally reflected actual precipitation better than ANN.

3.2 Effect of adding GNSS PWV

GNSS PWV is the estimated value of water vapor in the atmosphere, mostly wet components, over the vertical path of the station. Therefore, PWV is certainly related to the occurrence of precipitation. We added PWV as a factor of the model to improve the prediction accuracy in this study.

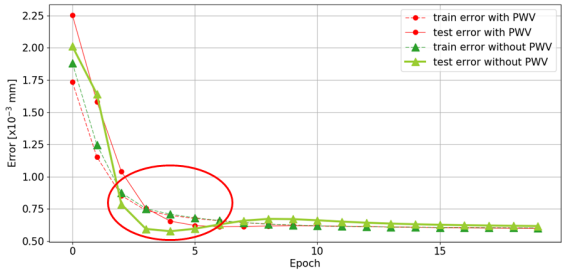

Fig. 11. Loss function with and without GNSS PWV

Fig. 11 clearly shows the performance of the model for both training and testing data sets. The errors represent the mean square error of the difference between the predicted and the actual values. The train error is defined by how many errors occur during the supervised learning process. For a smaller train error, it is necessary to increase the number of updates of the parameters and to lower the slope of the learning curve. In order to verify the performance of the model, we need to check the error of the test data set, called test error in this study. When GNSS PWV was added into the model as a feature, two errors show more closer behavior, and thus we can assume that the learning process of the model works well. On the other hand, however, the learning curves of the train and test errors are significantly different without GNSS PWV. In this case, it is considered as an over-fitting of the process, and the learning is not performed well even though there was little difference in RMSE.

4. Summary and Conclusion

This study proposed a precipitation prediction method based on the deep learning algorithm. Many studies have created a precipitation prediction using a complex mathematical weather model or ANN. However, since precipitation and meteorological data occur in a time series, predictions can be performed better through the LSTM algorithm in which past data can affect future estimations. GNSS PWV was added to the prediction model as a feature in order to estimate the amount of water vapor quantitatively. Based on the analysis of this study, we can conclude the following:

1. The experimental prediction results show that the RMSE of ANN and LSTM are 0.786 mm and 0.668 mm, respectively. The LSTM can predict precipitation more accurately than ANN (an improvement of about 15% was obtained). Therefore, it is suitable for calculating the relationship between complex factors and analyzing the data in a time series.

2. GNSS PWV was added to the precipitation prediction model in this study. As mentioned above, over-fitting of the model occurred without GNSS PWV. It means that this model is only appropriate for the used training data, and cannot predict future precipitation.

3. Localized heavy rain mainly occurs during summer in Korea. For future studies, we need to investigate the estimation accuracy of precipitation for specific seasons.

Acknowledgment

This research was supported by a grant (17RDRP-B076564-04) from Regional Development Research Program funded by Ministry of Land, Infrastructure and Transport of Korean government.

References

Baek, J.H., Lee, J.W., Choe, B.G., and Cho, J.H. (2007), Processing strategy for near real time GPS precipitable water vapor retrieval, *The Korean Space Science Society*, Vol. 24, No. 4, pp. 275-284. (in Korean with English abstract)

Bengio, Y., Courville, A., and Vincent, P. (2013), Representation learning: A review and new perspectives, *IEEE transactions on pattern analysis and machine intelligence*, Vol. 35, No. 8, pp. 1798-1828.

Bevis, M., Businger, S., Herring, T.A., Rocken, C., Anthes, R.A., and Ware, R.H. (1992), GPS meteorology: Remote sensing of atmospheric water vapor using the global positioning system, *Journal of Geophysical Research*, Vol. 97, No. 14, pp. 15787-15801.

Davis, J.L., Herring, T.A., Sharpiro, I.I., Rogers, A.E.E., and Elgered, G. (1985), Geodesy by radio interferometry: effects of atmospheric modeling errors on estimates of baseline length, *Radio Science*, Vol. 20, No. 6, pp. 1593-1607.

Elgered, G., Davis, J.L., Herring, T.A., and Shapiro, I.I. (1991), Geodesy by radio interferometry: Water vapor radiometry for estimation of the wet delay, *Journal of Geophysical Research*, Vol. 96, No. B4, pp. 6541-6555.

Ha, J.H., Lee, Y.H., and Kim, Y.H. (2016), Forecasting the precipitation of the next day using deep learning, *Journal of The Korean Institute of Intelligent Systems*, Vol. 26, No. 2, pp. 93-98. (in Korean with English abstract)

Hopfield, H.S. (1969), Two-quartic tropospheric refractivity profile for correcting satellite data, *Journal of Geophysical Research*, Vol. 74, No. 18, pp. 4487-4499.

Kang, B.S. and Lee, B.K. (2008), Predicting probability of precipitation using artificial neural network and mesoscale numerical weather prediction, *Journal of The Korean Society of Civil Engineers*, Vol. 28, No. 5B, pp. 485-493. (in Korean with English abstract)

Kim, J.S. (2016), *Enhancement of GNSS Precipitable Water Vapor Estimation and its Application to Weather Prediction*, Master's thesis, Sejong University, Seoul, Korea, 144p.

Kuligowski, R.J. and Barros, A.P. (1998), Localized precipitation forecasts from a numerical weather prediction model using artificial neural networks, *Wea. Forecasting*, Vol. 13, No. 4, pp. 1194-1204.

Lee, S.H. and Lee, J.H. (2016), Customer churn prediction using RNN, *Proceedings* of the *Korean Society of Computer Information Conference*, Vol. 24, No. 2 pp. 45-48.

Olah, C. (2015), Understanding LSTM networks, *Colah's Blog*, http://colah.github.io/posts/2015-08-Understanding-LSTMs/ (last date accessed: 27 August 2017).

Saastamoinen, J. (1972), Atmospheric correction for troposphere and stratosphere in radio ranging of satellites, *The Use of Artificial Satellites for Geodesy*, pp. 247-252.

Sachan, A. (2014), Forecasting of rainfall using ANN, GPS and meteorological data, *International Conference for Convergence for Technology-2014*, 6-8 April, Pune, India, pp. 1-4.

Tran, Q.K. and Song, S.K. (2017), Water level forecasting based on deep learning: A use case of Trinity River-Texas-The United States, *Journal of KIISE*, Vol. 44, No. 6, pp. 607-612.