PAPER • OPEN ACCESS

Long short term memory (LSTM) recurrent neural network (RNN) for discharge level prediction and forecast in Cimandiri river, Indonesia

To cite this article: Y Sudriani *et al* 2019 *IOP Conf. Ser.: Earth Environ. Sci.* **299** 012037

View the article online for updates and enhancements.

Long short term memory (LSTM) recurrent neural network (RNN) for discharge level prediction and forecast in Cimandiri river, Indonesia

Yuli Sudriani[1]*, Iwan Ridwansyah[1], and Hadiid Agita Rustini[1]

[1]Research Center for Limnology, Indonesian Institute of Sciences (LIPI), Cibinong Science Center - Botanical Garden, Jl. Raya Jakarta Bogor Km.46. 16911

*E-mail: yuli@limnologi.lipi.go.id

Abstract. Cimandiri watershed in Sukabumi prefecture of West Java, Indonesia, has been used for profitable activities such as power plant, rafting tourism, drinking water, and municipal, industries, agriculture, and fishery, and irrigation. More than 60% water source of PDAM, which supplies irrigation water of 1,217 Ha of rice fields and hundreds of industries, is obtained from the river. The irrigation in Cimandiri is even designed to be the model of irrigation in West Java. Low river discharge during the dry season can generate disadvantages to irrigation. This paper presents a method to predict and forecast the discharge of five days ahead to help the decision maker to control the operation of irrigation. We use a Deep Learning algorithm which involves Recurrent Long Short Term Memory (LSTM) Neural Network. Daily discharge data at two river gauge stations were analyzed. These stations are Leuwilisung (17 years data), and Tegaldatar (13 years data). The result shows that the relative errors are below 10% which is acceptable. In this study, dynamic changes of discharge level are evaluated to give a contribution to irrigation and water management control in Cimandiri River, Indonesia.

1. Introduction

Cimandiri is a 183,196 Ha watershed located between 6º42'56" - 7º8'45" SL and 106º30'45" - 107º4'50" EL in West Java ([1]). The upstream of Cimandiri watershed is located in Mt. Kendeng and Mt. Gede (3,000 masl), Mt. Halimun (1,321 masl), and Mt. Salak (2,052 masl), while its downstream is in Pelabuhan Ratu, Sukabumi Regency. The slope in this watershed varies between 0% (flat) to more than 40% (very steep). According to the Spatial and Regional Plan of Sukabumi Regency ([1]) 2012-2032, this river supplies water for irrigation of 1,217 Ha area. More importantly, the irrigation in this watershed is designed to be the model of irrigation in West Java ([2]).

During the dry season, the river discharge is very low and the reservoir could not fulfill water demand adequately ([3],[4]). Such condition can generate disadvantages, especially for model irrigation where water must be available at the right time and the right amount. Discharge prediction is one of the important information for water managers [5] to make decisions regarding water supply efficiently, effectively, and reliably.

Many flow prediction models require accurate hydrological long term flow data and extensive field measurement, which can be costly, especially when the water flows must be determined at many locations in a large basin [7]. For practical objectives, research about simulation of discharge level prediction using conventional Artificial Neural Network (ANN) has been widely used [7-20]. Limitations to ANN methods give rise to the development of Deep Learning (DL), a new methodology

of machine learning which can provide information from a sequence data to next sequence data using self-learning functions based on a layer by layer. One of DL Algorithm is Recurrent Neural Network (RNN). Long Short Term Memory (LSTM) architecture is a special kind of RNN which was introduced by Hochreiter and Schmidhuber [6] to overcome the weakness of the traditional RNN to learn long term dependencies (over ten sequences) such as shown by Bengiot et al. [21]. RNN-LSTM is a suitable method for sequence data. This method tries to give a prediction using step by step of sequence data. The advantage of LSTM-RNN is suitable for time series prediction and its robustness in dealing with nonlinearity and huge data.

Study about the model of discharge water level using RNN LSTM have been used increasingly in water resources management such as discharge level prediction [22,[27-30], water distribution data[28], for rainfall-runoff process[25], for flood prediction [27, 31-33], for environmental water quality [26]. Wu et al. [22] and Wu and Rahman [23] discuss the implementation of DL for water distribution including an electric steam power plant, rafting tourism, drinking water and municipal, industries, fishery, and irrigation. This paper aims to predict and forecast discharge level data based on RNN LSTM methodology.
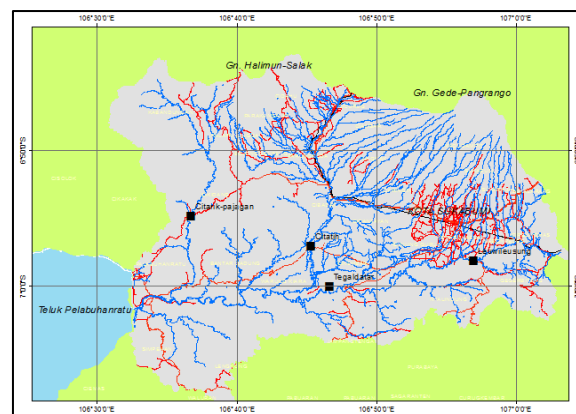
2. Methods

Figure 1. Location of Study Area for Leuwilisung and Tegaldatar Discharge Station in Cimandiri Watershed (Source: Ridwansyah, 2017).

In this study, we use daily discharge data from two stations; Leuwilisung (17 years) and Tegaldatar (13 years) which were recorded manually by the Center for Research and Development of Water Resources, the Ministry of Public Works and Housing (PUSAIR-PUPERA).

LSTM Method is related to the memory cell and how to related between old context (C_{j-1}) to a new context (C_j). Cj is a memory cell. Input gate is consist of discharge level data from Leuwilisung arena and Tegal Datar. Out$_j$(output gate) is memory cell result of out of in j (the "input gate") and then as input to the next sequence. When adding a time (t) as input memory, then the equation will be using $y^{inj}(t)$, Out$_j$, and $y^{Outj}(t)$. f^{putj} is the process of total out in sequence data. The first step to calculate the LSTM is The explicit formula of the LSTM-RNN is [6]:

$$y^{Outj}(t) = f^{outj}(net_{outj}(t)); \quad y^{inj}(t) = f_{inj}(net_{inj}(t)) \tag{1}$$

Where,

$$net_{outj}(t) = \Sigma \, \omega_{outj \, \mu} y^{\mu} \, (t-1) \tag{2}$$

$$net_{inj}(t) = \Sigma \ \omega_{inj \ \mu} y^\mu \ (t-1) \tag{3}$$

$$net_{cj}(t) = \Sigma \ \omega_{cj \ \mu} y^\mu \ (t-1) \tag{4}$$

Definition:

$y^{Outj}(t)$: Total Result Process of Output Gate of Memory Cell

f^{outj} : The result of Forget Cell Data for output Gate

net_{outj} : The Result of Output Gate

$y^{inj}(t)$: Total Result Process of input Gate

f_{inj} : The result of Forget Cell Data for input Gate

net_{inj} : The Result of input Gate

μ : Input unit, memory unit, gate unit

ω : Result of Store information.

There are two simulations in each study area: RNN- LSTM Simulation for Leuwilisung and RNN-LSTM simulation for Tegaldatar. The first step of this method is making a frame for the array of discharge level sequence data in Leuwilisung and Tegaldatar. The loaded dataset will be processed to be supervised learning. The second step is the transformation of time series to fixed data using *Tan h* which result in a constant value between -1 and 1. *Tan h* will transform the time series to scale and got trend data. The third step is prediction and forecast using *predict ()* and *forecast* syntax with the output is two dimensions with one value. The last step is the evaluation of performances for prediction and forecast using root mean square error (RMSE) and mean absolute error (MAE) as the following:

$$RMSE = \sum_{i=1}^{n} (H_i - H_i)^2 \tag{5}$$

$$MAE = \frac{1}{2}\sum_{i=1}^{n}|e_i| \tag{6}$$

Where $\widehat{H_\iota}$ and H_i are the forecasted and observed discharged level of the data, respectively. H^dan H- are the average forecasted and observed discharge level. n is the number of data points and e_i are n samples of model error.

3. Results and Discussion
After data preprocessing, a total of 10,589 records of daily data from two stations (Leuwilisung and Tegaldatar) are collected. Leuwilisung data using 17 Years data; 1993,1994,1996,1998-2001,2003-2004,2006-2013. Tegaldatar data using 13 Years data; 1999-2001,2003-2005,2007-2013 as follows:
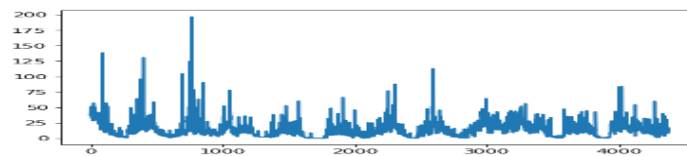

Figure 2. Maximum daily discharge in Leuwilisung.
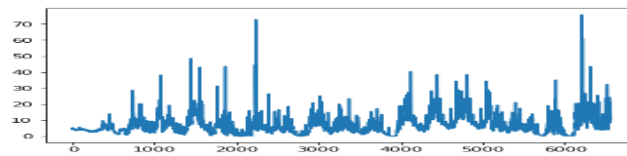
Figure 3. Maximum daily discharge in Tegal Datar.

The maximum daily discharge for Leuwilisung and TegalDatar is about 70 ml/Second and 200ml/Second.
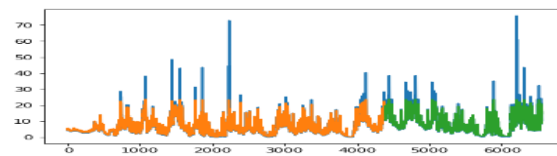

Figure 4. Discharge Level Train Data and Test Data Leuwilisung Station.
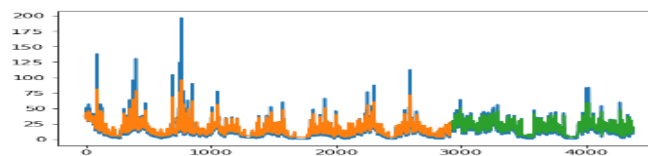

Figure 5. Discharge Level Train Data and Test Data Leuwilisung Station.

Before the process of forecast simulation, data were divided into train data and test data to simulate the initial data condition. Results showed that daily discharge level for Tegaldatar is higher than Leuwilisung.

Table 1. Validation of Error for Train Data and Test Data.

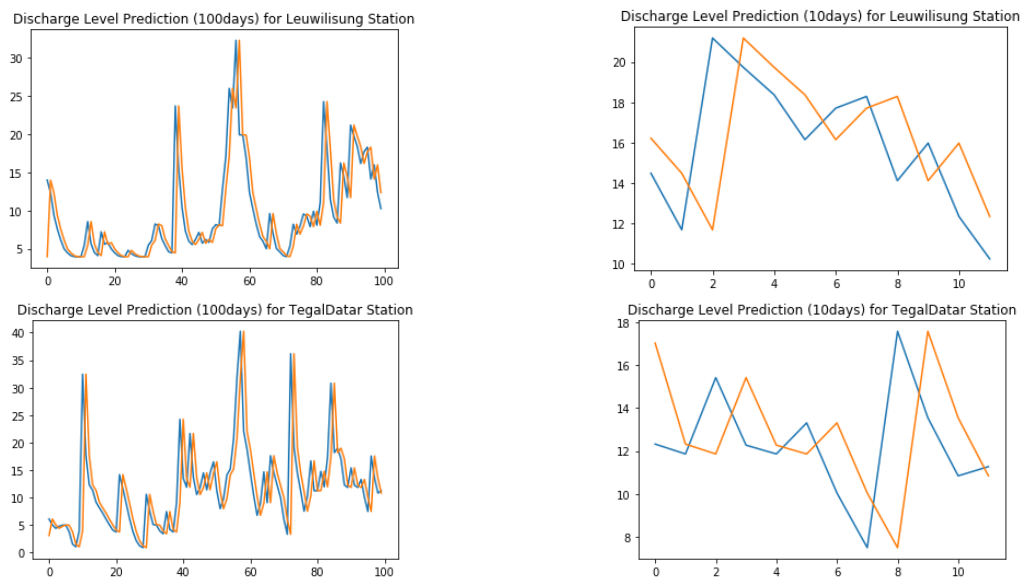| Station	Train Score Error	Test Score Error
Leuwilisung	Train Score: 2.36	Test Score: 3.27
TegalDatar	Train Score: 8.98	Test Score: 7.94

Figure 6. Prediction Data Result compared with Observed Data for 2 Station.

The algorithm separates the data into the training datasets (67% of the observations) and testing datasets (33%). Figure 5 shows the result of our model; the original dataset in **blue**, the predictions for the training dataset in **orange**, and the predictions on the unseen test dataset in **green**. The Prediction error is accumulated over time. Train Score Error and test score error were less than 10%, whereas error for Tegaldatar was higher than that of Leuwilisung. This is likely because the input data for Leuwilisung has a longer data series than Tegaldatar, as was also concluded by Bengiot et al. ([21]).

Table 2. Validation of Error for Prediction Data and Observed Data.

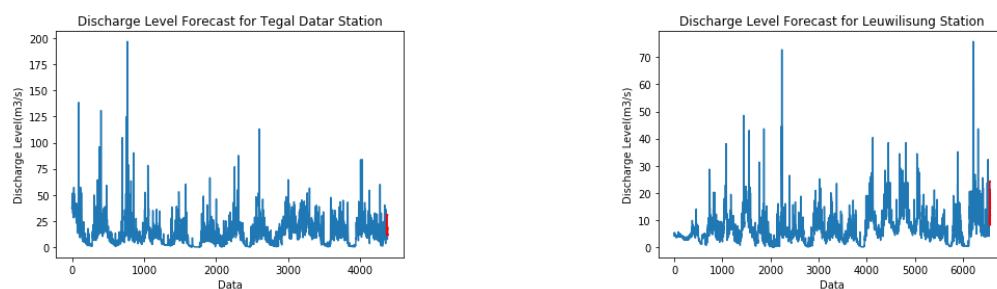| Station	Train Score Error
Leuwilisung	RMSE: 3.989
Tegal Datar	RMSE: 6.989

Figure 7. Discharge Water Level for Forecast in TegalDatar and Leuwilisung Station.

Table 3. Validation Error Of Forecast Discharge Level For 5 Days A Head

Station	RMSE (Root Mean Squared Error)	MAE (Mean Absolute Error)
Leuwilisung	t+1 RMSE: 4.182481 t+2 RMSE: 5.061993 t+3 RMSE: 4.158058 t+4 RMSE: 5.282580 t+5 RMSE: 6.583957	t+1 MAE: 2.322929 t+2 MAE: 2.837076 t+3 MAE: 2.560078 t+4 MAE: 2.308029 t+5 MAE: 2.209977
Tegal Datar	t+1 RMSE: 2.625329 t+2 RMSE: 4.360693 t+3 RMSE: 3.835111 t+4 RMSE: 4.038824 t+5 RMSE: 4.306496	t+1 MAE: 2.219459 t+2 MAE: 2.264950 t+3 MAE: 2.417023 t+4 MAE: 2.557929 t+5 MAE: 2.402291

Table 4 shows the values of the validation error of forecast discharge level for five days ahead. The Relative error for the forecast is less than 10.0%. In regards to accuracy, all the models provide good accuracy for a short time horizon forecast. However, if longer time horizons are considered (Tegaldatar) the RMSE is 2.6253 but was increasing to 1.7 points on the fifth day. This temporal limit is coherent when compared between all station's data.

4. Conclusion

The most important part of the ANN model is its ability to forecast future events. Discharge water level prediction and forecasting is an essential topic in water management. The benefit of LSTMs with advantages for long sequences is the method can learn to make a one-shot multistep forecast which useful for time series forecasting. The forecasting accuracy tendency (worse or better) depends on the complexity of data. Overall, the approaches show performances which remain satisfactory up to 5 days ahead for the watershed.

Acknowledgment

References

[1]   Peraturan Daerah Kabupaten Sukabumi No. 22 Tahun 2012 tentangRencana Tata Ruang Wilayah KabupatenSukabumiTahun 2012-2032, 136p.

[2]   Bappeda Prov. Jawa Barat, 2016, (http://bappeda.jabarprov.go.id/pemprov-jabar-buka-lahan-sawah-baru-di-pesisir-cimandiri/) accessed on Oct 20, 2018

[3]   Suherlan D, 2018, Walhi: Darurat, KekeringanTerusMeluas di Jawa Barat, (https://jabarnews.com/2018/08/walhi-darurat-kekeringan-terus-meluas-di-jawa-barat.html) accessed on Dec 10, 2018.

[4]   Rayadie A, 2017, Kekeringan di SukabumiMeluas, (http://www.pikiran-rakyat.com/jawa-barat/2017/08/18/kekeringan-di-sukabumi-meluas-407692) accessed on Dec 10, 2018.

[5]   Aroengbinang B. W, 2015, *Analisis Faktor Pendorong Perubahan Penutupan dan Penggunaan Lahan di Daerah Aliran Sungai Cimandiri dan Cibuni*, Thesis in Institut Pertanian Bogor, 79p.

[6]   Hochreiter S, Schmidhuber, 1997, Long Short-Term Memory, *Neural Computation,* 9(8):1735-1780.

[7]   Kişi Ö. (2007). Streamflow forecasting using different artificial neural network algorithms. Journal of Hydrologic Engineering, 12(5), 532-539.

[8]   Alcazar J, A. Palau, and C. Vega-Garcia, 2008, A neural net model for environmental flow

estimation at the Ebro River Basin, Spain, *Journal of Hydrology* 39(1):44-55.

[9]   Tayyab Muhammad, Jianzhong Zhou, XiaofanZeng, Rana Adnan. 2016. Discharge Forecasting by Applying Artificial Intelligence Neural Networks at TheJinsha River Basin, China. *European Scientific Journal vol.12.* ISSN: 1857-7881 (Print) e-ISSN :1857-7431. Pp: 108-126

[10] Joorabchi A, H.Zhang, M. Blumenstein. 2007. Applying of Artificial Neural Networks in Flow Discharge Prediction for Fitzroy River, Australia. *Journal of Coastal Research.*pp: 287-291

[11] Alvisi Stefano, Franchini Marco. 2011. "Fuzzy Neural Network for water level and Discharge Forecasting with Uncertainty". *Environmental Modelling and Software 26*. Pp:523-537

[12] Yawar Ali M, Panwar Sugandha. 2014. "Neural Network Model for Discharge and Water Level Prediction for Ramganga River Catchment of Ganga Basin, India". *Hydrological Sciences Journal.*

[13] Jia-hua Wei, Chu Hai-bo, Wang Rong, Jiang Yuan. 2015. Numerical Simulation of Karst Groundwater System for Discharge Prediction and Protection Design of Spring in Fangshan District, Beijing. *Journal of Groundwater Science and Engineering 3.*pp: 316-330

[14] Aichouri Imen, Azzedine Hani, Nabil Bougherira, Larbi Djabri, Hicham Chaffai, Sami Lallahem. 2015. "River Flow Model Using Artificial Neural Network". *Energy Procedia 74.* Pp:1007-1014

[15] Lee Do-Hun, Doo-Sun Kang. 2016. "The Application of Artificial Neural Network Ensemble Model for Simulating Streamflow". *Procedia Engineering 154.* Pp:1217-1224

[16] Uysal Gokcen, Ali Arda Sorman, Aynur Sensoy. 2016. "Streamflow Forecasting using Different Neural Network Models With Satellite Data for a Snow Dominated Region in Turkey".*Procedia Engineering 154.* Pp: 1185-1192

[17] Yang Beibei, Kuniong Yin, Suzanne Lacasse, Zhongqiang Liu. 2019. " Time Series Analysis and Long Short Term Memory Neural Network to Predict Landslide Displacement". *Landslide 16.* Pp: 677-694

[18] Lee Do-Hun, Doo-Sun Kang. 2016. "The Application of Artificial Neural Network Ensemble Model for Simulating Streamflow". *Procedia Engineering 154.* Pp:1217-1224

[19] AichouriImen, Azzedine Hani, Nabil Bougherira, LarbiDjabri, HichamChaffai, Sami Lallahem. 2015. "River Flow Model Using Artificial Neural Network". *Energy Procedia 74.* Pp:1007-1014

[20] Bengiot, Y, P. Simard, and P. Frasconi, 1994, Learning Long-Term Dependencies with Gradient Decent is Difficult, *IEEE Transactions on Neural Networks*, 5(2):157-166.

[21] Zhang,J, Yan,Zhu, Xiaoping Zhang, Ming ye, Jinzhong Yang. 2018. Developing a Long Short Term Memory (LSTM) based model for predicting water table depth in agricultural areas. *Journal of Hydrology*. 561(918-929)

[22] Hu Caihong, Qiang Wu, Hui Li, ShengqiJian, Nan Li, Zhengzheng Lou. 2018. "Deep Learning with a Long Short Term Memory Network Approach for Rainfall Runoff Simulation", *Water,* 10(11),1543, https://doi.org/10.3390/w10111543

[23] Zhang Jun, Xiyao Cao, JieminXie, Pangao Kou.2019. "An Improved Long Short Term Memory Model for Dam Displacement Prediction",*Mathematical Problems in Engineering,* Article ID: 6792189

[24] Kratzert Frederik, Daniel Klotz, Claire Brenner, Karsten Schulz, Mathew Herrnegger, 2018. " Rainfall-runoff modelling using Long Short Term Memory (LSTM) Networks, *Hydrology and Earth System Sciences(22),*6005-5022

[25] Kangil Kim, Kim Dong-Kyun, Noh Junhyug, Kim Minhyeok. 2018. "Stable Forecasting of Environmental Time Series via Long Short Term Memory Reccurent Neural Network". *Digital Object Identifier.* Pp:75216-75228

[26] Wu, Z.Y, El - Maghraby, M, Pathak, S, 2015, Applications of Deep Learning for Smart Water Networks", *Procedia Engineering,*119, 479-485

[27] Wu, Z.Y, Rahman, A, 2017, Optimized Deep Learning Framework for Water Distribution Data Driven Modeling, *Procedia Engineering,* 186, 261-268

[28] Pedro,D, A Few Useful Things To Know About Machine Learning, *Article in Departement Computer Science and Engineering University of Washington.*

[29] Cai H, H.H.G Savenije, C.Jiang. 2014. "Analytical Approach for Predicting Fresh Water Discharge in an Estuary Based on Tidal Water Level Observations. *Hydrology and Earth System Science 18.* Pp: 4153-4168

[30] Desmukh P Rahul, A.Ghatol. 2010. "Short Term Flood Forecasting Using Reccurent Neural Network a Comparative Study". *IACSIT International Journal of Engineering and Technology 2*. ISSN : 1793-8236. Pp: 430-434

[31] Jia-hua Wei, Chu Hai-bo, Wang Rong, Jiang Yuan. 2015. Numerical Simulation of Karst Groundwater System for Discharge Prediction and Protection Design of Spring in Fangshan District, Beijing. *Journal of Groundwater Science and Engineering 3.*pp: 316-330