

Comparison of M5 Model Tree and Nonlinear Autoregressive with eXogenous inputs (NARX) Neural Network for urban stormwater discharge modelling

EL GHAZOULI Khalid^{1,2}, EL KHATABI Jamal¹, SHAHROUR Isam¹ and SOULHI Aziz³

1 Laboratoire de Génie Civil et géo-Environnement, Université de Lille, 5900 Lille, France

2 Laboratoire d'Analyse des Systèmes, Traitement de l'Information et Management Industriel, Université Mohammed V, Rabat, Morocco

3 National Higher School of Mines, Agdal Rabat, Morocco

Abstract. This paper presents a comparative study of two data-driven modelling techniques in forecasting urban drainage stormwater discharge based on rainfall prediction. Both M5T and NARX (Nonlinear Autoregressive with eXogenous inputs) Neural Network are used for 30 minutes storm water forecasting. Data are collected from watershed area of 3315 ha, located in the city of Casablanca in Morocco. The results show that both models provide good results, but however with better performances of the NARX model.

1. Introduction

Major cities are facing urban flood issues due to urbanization and climate change. Real time control (RTC) systems allow an improvement in the performances of sewer networks through the control of the different network components and the optimization of storage within retention ponds and main collectors (Beeneken et al., 2014). The essential input parameters in RTC systems are rain and flow forecasts. Artificial intelligence techniques had shown good performances in modelling nonlinear systems with a considerable reduction in computing time compared to hydrologic and hydraulic models (Abou Rjeily et al. 2017). Two artificial intelligence techniques, namely, neural networks and M5 Model Tree were discussed in several applications such as sediment transport modeling (Bhattacharya et al. 2007), river flow forecast (Solomatine and Khada. 2003), evaporation estimation (Rahimikhoob. 2014). This paper presents an extension of this comparison to urban drainage application.

2. Methodology and material

2.1. Methodology

Artificial Neural Networks

Neural networks are commonly used in artificial intelligence techniques. They showed good performances in nonlinear system simulation and time series prediction (stock forecasts (Pang, X et al 2018) and urban flood prediction (Berkhahn et al 2019)). They are considered as black box containing the information to be learned. In the beginning, the architecture of the neural network is composed of layers and nodes without any information or knowledge of the simulated phenomenon. During the learning stage, the weights of the neurons that connect the different nodes are determined according to an optimization algorithm to minimize the error of the output of the neural network and measured data.

The Nonlinear Autoregressive with eXogenous inputs (NARX) neural network given by equation (1) is used in this research:

$$y(t) = f(y(t - 1), y(t - 2), \dots, y(t - n_y), u(t - 2), \dots, u(t - n_u)) \quad (1)$$

where $y(t)$: output time series, $u(t)$: input time series, n_y and n_u : are the time delays representing the phenomenon dynamic behavior.

Two types of architectures of NARX are generally used. The first one is efficient in forecasting a time series value one-time step ahead and it's called series-parallel or open loop. The second one is called parallel architecture or closed-loop and it is efficient for multistep-ahead prediction.

M5T

The M5 model is based on the principle of information theory (Quinlan 1992, Solomatine and Khada. 2003). It can handle nonlinear problem by splitting it into multiple linear ones. The M5T is similar to conventional decision trees, but instead of predicting classes through a classifier it predicts continuous variables through linear regression functions at the leaves. The construction of the M5 model tree is done in two stages (Solomatine and Xue. 2004, Rahimikhoob et al. 2013). First, the M5 model tree is constructed by a divide-and-conquer method that splits the multidimensional data space into several subspaces based on a splitting criterion (Figure 1) that treats the standard deviation of the class values that reach a node as a measure of the error at that node and computes the expected reduction in the error by testing each attribute at that node. The splitting operation stops when the values of the instances reaches a node with slight variation or just few instances remain (Goyal and Ojha 2014)

The formula used to calculate the standard deviation reduction is given by (Solomatine and Khada. 2003, Pal and Deswal, 2009. Goyal and Ojha. 2014, Gyanendra et al 2016):

$$SDR = sd(T) - \sum \frac{|T_i|}{|T|} sd(T_i) \quad (2)$$

where T is the set of examples that reaches the node; T_i is the subset of examples that have the i th outcome of the potential set; and sd is the standard deviation

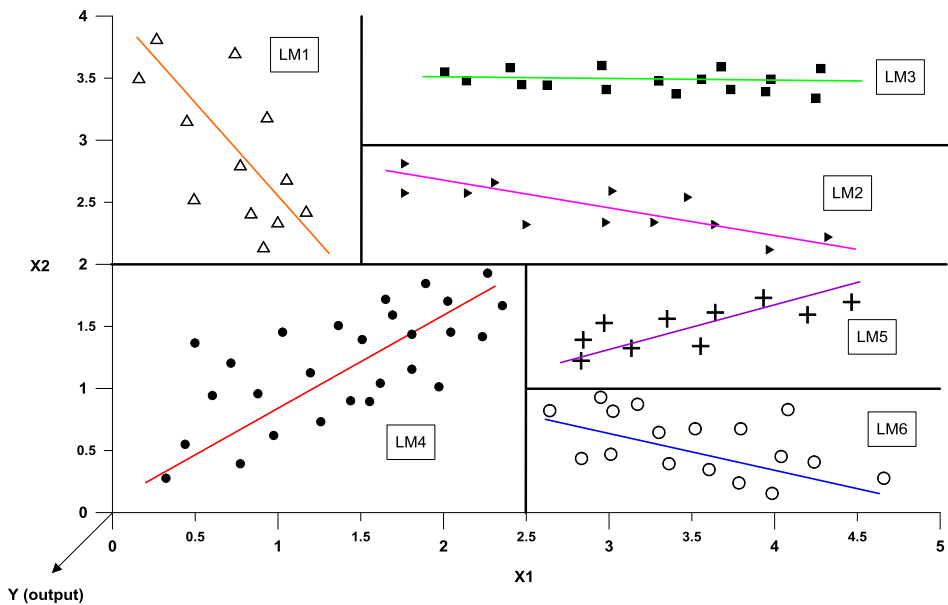


Figure 1: Splitting the input space ($X1 \times X2$) by M5 model tree algorithm

The result of the splitting is a tree with splitting rules at the nodes with expert linear models (LM) at the leaves (Figure 2). The combination of linear models can be seen as committee machine (Haykin, 1999).

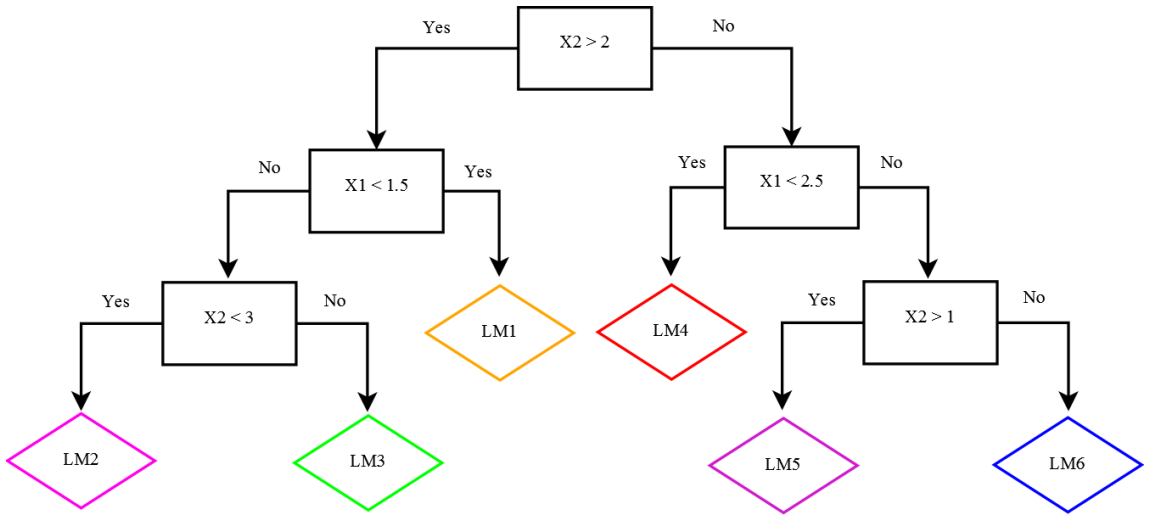


Figure 2: Diagram of the M5 model tree with six linear regression models at the leaves

The generated tree is often large and difficult to analyze and may lead to overfitting. In the second stage a pruning operation is needed to overcome this problem by replacing a subset with a leaf. In addition to this, the smoothing is performed to compensate the discontinuities that can occur between the adjacent linear models at the pruned leaves of the tree.

2.2. Site description

The study area concerns a watershed of 3315 ha located in the city of Casablanca in Morocco. The urban drainage system (UDS) in this area is composed of 485 km of both combined and separate networks (Figure 3). The UDS is equipped with a depth meter and flow meter the outlet of the watershed. The measurements for UDS are recorded at 15 minutes time interval. A rain gauge is located in the watershed to record the rain intensity at a 5 minutes time interval.

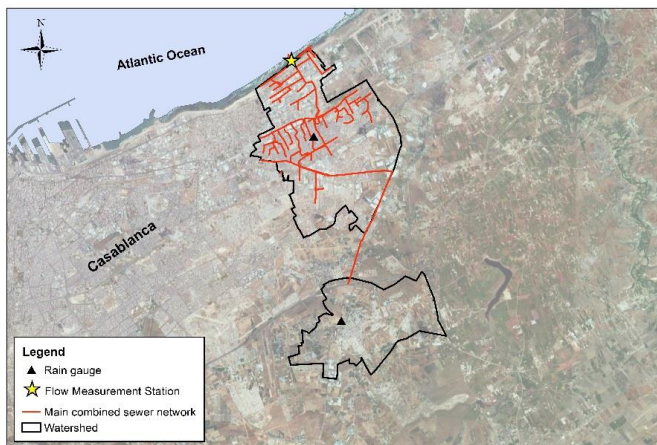


Figure 3: Area used in this study, Casablanca

2.3. Data collection

Data had been extracted from the database for a period of 4 years (January 2015 – January 2019). Since the main sewer system at the outlet of the watershed is a combined one dry weather flow had been removed to keep only data related to rain flows.

The rain dataset used in this present work is characterized by return periods summarized in table 1, which shows a dominance of low to medium rains in the dataset.

Return Period	Percentage of the sample
1 year	58%
2 year	18%
5 year	14%
10 year	2%
20 year	4%
50 year	4%

Table 1: Statistics of return periods (RP) of the rain dataset

3. Application

Both M5T and NARX were used to predict the stormwater flow at 30 minutes. Data used in the construction of the two models are the appropriate data that contribute in the stormwater flow $Q_{t+30min}$.

These Data are the previous rainfall data (Ret-80min, ..., Ret-5min, Ret) and the previous measured discharge ($Q_{t-10min}$, Q_{t-5min} and Q_t). Both models were trained on a data sample of 4 years.

In order to evaluate the performances of both models in predicting $Q_{t+30min}$ for different meteorological conditions, a sample of 4 rain events with various intensities and different return period was used.

These rainfall return periods used for testing the models are: 50YRP (Year Return Period), 2 YRP, 5YRP and 5YRP.

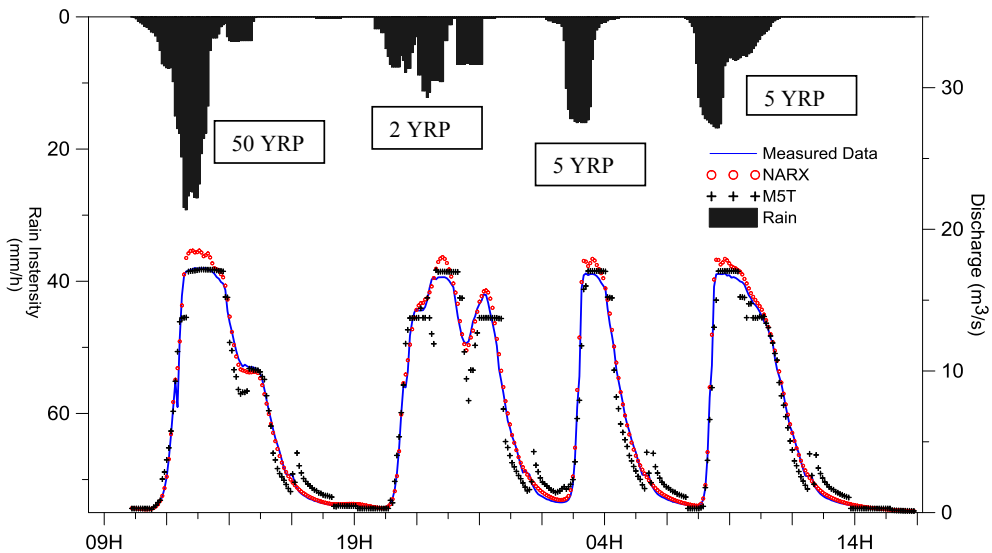


Figure 4: Prediction results for both NARX and M5T for 4 different rain events

	NARX	M5T
RMSE (m ³ s)	0.726	1.330
NSE	0.988	0.959

Table 2: Performance statistics of NARX and M5 model tree

The efficiency of the models was examined through results and measurements comparison with a Nash Sutcliffe Efficiency (NSE) and Root Mean Square Error (RMSE) calculation.

Results in table 2, show a good performance of both models with a small RMSE and a high NSE. Figure 4 shows that the NARX model represents better the flow hydrograph with a slight overestimation of the peak flow (about 10%) compared to M5T, which gives better results in predicting peak flows.

4. Conclusion

This paper presented the use of two artificial intelligence techniques, namely NARX and M5T for forecasting urban drainage stormwater discharge using rainfall predictions. Both of these models could be easily implemented for practical use. They were tested and compared on data collected from watershed area of 3315 ha, located in the city of Casablanca. Results show that both models provide good results. The NARX model shows good performances in modelling all the ranges of flow with slight overestimation of peak values compared to M5T Model, which provides good results in peak flows forecasting.

References

1. Y. Abou Rjeily, O. Abbas, M. Sadek, I. Shahrour and F. Hage Chehade (2017) Flood forecasting within urban drainage systems using NARX neural network, *Water Science & Technology*
2. Bhattacharya, B., Price, R. K., and Solomatine, D. P. (2007). "Machine learning approach to modeling sediment transport." *J. Hydraul. Eng.*, 133(4), 440–450
3. Beeneken, T., Erbe, V., Messmer, A., Reder, C., Rohlfing, R., Scheer, M., Schuetze, M., Schumacher, B., Weilandt, M. & Weyand, M. (2014). Real time control (RTC) of urban drainage systems – A discussion of the additional efforts compared to conventionally operated systems. *UrbanWater Journal* 10
4. Berkhahn S, Fuchs, L., Neuweiler I. An ensemble neural network model for real-time prediction of urban floods. (2019) *Journal of hydrology- Volume 575-Volume 575-*, Pages 743-754
5. Goyal MK, Ojha CSP (2014) Evaluation of rule and decision tree induction algorithms for generating climate change scenarios for temperature and pan evaporation on a Lake Basin. *ASCE J Hydrol Eng* 10.1061/(ASCE)HE. 1943–5584.0000615
6. Singh G, Sachdeva SN, Pal M (2016) M5 model tree based predictive modeling of road accidents on non-urban sections of highways in India. *Accid Anal Prev* 96:108–117
7. Haykin, S. (1999). *Neural networks: A comprehensive foundation*, Prentice-Hall, Englewood Cliffs, N.J.
8. Pal M, Deswal S (2009) M5 model tree based modelling of reference evapotranspiration. *Hydrol Process* 23:1437–1443
9. Pang, X.; Zhou, Y.; Wang, P.; Lin, W.; Chang, V. (2018). An innovative neural network approach for stock market prediction. *J. Supercomput.*, 1–21.
10. Quinlan, J.R., (1992). *Learning with continuous classes*. In: *Proceedings of Australian Joint Conference on Artificial Intelligence*, World Scientific Press: Singapore, pp. 343–348.
11. Rahimikhoob, A. (2014). Comparison between M5 model tree and neural networks for estimating reference evapotranspiration in an arid environment. *Water Resour. Manage.* 28, 657–669.
12. Solomatine, D. P. & Dulal, K. N. (2003) Model trees as an alternative to neural networks in rainfall–runoff modelling. *Hydrological Sciences Journal.* 48(3), 399–411.
13. Solomatine DP, Xue Y (2004) M5 model trees compared to neural networks: application to flood forecasting in the upper reach of the Huai River in China. *J Hydr Engrg* 9(6):491–501