

Intermittent Reservoir Daily Inflow Prediction Using Stochastic and Model Tree Techniques

Deepali More¹ · R. B. Magar^{1,2} · V. Jothiprakash¹ 

Received: 12 May 2017 / Accepted: 5 February 2019 / Published online: 9 March 2019
© The Institution of Engineers (India) 2019

Abstract The present study depicts the applicability of model tree (MT) technique to a large data set having large number of zero values. It is also aimed to develop a model that results in simple equations as that of stochastic models. The performance of MT is compared with conventional autoregressive integrated moving average (ARIMA) models. Forty-nine years of daily inflow data from Koyna Reservoir located in Maharashtra, India, are used for developing and testing the models. In this case study of developed MT models, the number of inputs is selected by trial and error and is varied from one lag to eight lags. Numerous MT models were developed by considering the model formulations of pruning and smoothing, whereas in ARIMA model, the number of inputs required for proper modeling is selected from autocorrelation function and partial autocorrelation function plots as well as through trial-and-error procedure. The performances of the developed models were evaluated using various statistical measures. On comparing the daily time step MT and ARIMA models, it is found that un-pruned and un-smoothed MT models performed better than ARIMA

models. Even though the number of leaves (local linear equations with nonlinear way of finding them) is slightly larger, the low and peak values of the reservoir inflow are predicted better by MT model. From the results, it is concluded that for better modeling and to have a set of linear applicable equations for smaller time step reservoir inflow, MT technique can be a better choice than ARIMA model.

Keywords Model Tree · Intermittent reservoir · Inflow prediction · Artificial intelligence · Koyna Reservoir · Maharashtra, India

Introduction

Reservoir inflow prediction at smaller time step is very much required for reservoir planning, operation and maintenance especially during floods and droughts [10]. Even though several artificial intelligent (AI) models have been applied to predict daily and hourly runoff, those models are difficult to be implemented in real life as they need a more sophisticated way to extract the knowledge gained through modeling.

Fairly large numbers of studies have been reported on reservoir inflow prediction, such as conventional stochastic models, which are based on statistical properties of the historical data, and AI models based on data-driven techniques [23]. ARIMA models are one such stochastic model widely applied in water resources applications, especially for prediction of monthly inflow into reservoir, streamflow, runoff, etc. [1, 15, 16, 29]. However, the drawback of using stochastic models for daily inflow prediction is its basic assumption of linearity and normality, which is uncertain in hydrological processes, especially when the time step is

✉ V. Jothiprakash
vprakash@iitb.ac.in

Deepali More
more.dipa@gmail.com

R. B. Magar
rajendramagar69@gmail.com

¹ Department of Civil Engineering, Indian Institute of Technology Bombay, Powai, Mumbai 400076, India

² Present Address: Department of Civil Engineering, School of Engineering and Technology, Anjuman-I-Islam's Kalsekar Technical Campus (AIKTC), Plot 2 and 3, Sector-16, Near Thana Naka, Khandagaon, New Panvel, Navi Mumbai 410206, India

smaller such as hourly and daily, therefore not suitable for accurate prediction. Also when there are large continuous zero values in the data series, stochastic models are not suitable [11]. Thus, researchers are seeking alternative models to forecast future values in terms of high and low flows especially for smaller time step time series.

There are fairly good number of AI techniques available in the literature for hydrologic time series modeling, including artificial neural networks (ANN), artificial neuro-fuzzy inference system (ANFIS), model tree (MT) and genetic programming (GP). Very few works have been reported on application of MT for an intermittent reservoir system having large zero values. MT results in a simple linear set of equations developed using nonlinear method and can be a better alternative to linear stochastic models because of easy model setting, faster training and understandable results. Few works have been reported on application of MT in prediction of rainfall, water level discharge, flood forecasting and prediction of sedimentation and crop evapotranspiration [4, 7, 8, 14, 18, 20, 24].

Very few applications are available on application of MT in reservoir inflow prediction in spite of its advantages like less input from modeler (thus human error is less), easy understanding and implementation in field. The two data-driven techniques ANN and MT were investigated by Solomatine and Dulal [24] for rainfall–runoff transformation, and it was concluded that MT is a promising alternative to ANN. The M5 algorithm which is popular for MT was used for generating results of effective predictions, which showed that results were almost matching with results of ANN. Solomatine and Xue [26] applied M5 MT model for a flood forecasting problem in China and compared the results with conventional ANN model. It was concluded that MT models predicted the high floods accurately. Bhattacharya et al. [5] used MT and ANN for predicting the bed-load and sediment-load transport time series. It was reported that MT models outperformed ANN models. It was also reported that MT models are found to be easier to transfer the knowledge from laboratory to field.

Bhattacharya and Solomatine [4] applied MT and ANN for modeling water level discharge relationship, and it was concluded that MT is transparent and gives very simple demonstrable model output. Štravs and Brilly [27] applied M5 MT to analyze the recorded streamflow data, and results showed that appreciable accuracy was achieved. Pal and Deswal [18] applied MT to model daily reference crop evapotranspiration. Kote and Jothiprakash [7, 8] studied the effect of pruning and smoothing while using MT model to predict the inflow into Pawana reservoir and concluded that MT models with un-pruned and un-smoothed should be used while modeling reservoir inflow prediction especially to achieve better peak inflow values. Mandal and Jothiprakash [14] applied MT to predict 1-day ahead daily

rainfall in the Koyna Reservoir catchment area. Onyari and Ilunga [17] applied MT and multilayer perceptron artificial neural network (MLP-ANN) to predict streamflow in Luvuvhu catchment, South Africa, and concluded that MT-pruned model is better than MLP-ANN. It is reported that MT is more sensitive toward data splitting. Sattari et al. [22] applied MT and ANN to predict evapotranspiration in Ankara, Turkey, and concluded that both the techniques performed well, and added the major advantage of MT, availability of simple liner equation in predicting evapotranspiration.

All the above studies indicate that MT is a promising technique to model hydrological processes; however, very few works have been reported on application of MT for reservoir inflow prediction especially for an intermittent reservoir having large number of zeros. The above works are the motivation to apply the powerful MT model to predict daily reservoir inflow data into Koyna Reservoir, which is not reported so far. This research is a case study application of MT models to predict daily reservoir inflow into Koyna Reservoir in Maharashtra, India. It is also aimed to develop hydrologic models that results in simple and powerful equations like ARMA/ARIMA to be applied in real life especially for smaller time step to use it at dam site. The results of MT models are then compared with ARIMA models to evaluate the performance of MT models. ARIMA models are developed using log-transformed data to follow normality. The model development in the present study and its application in reservoir daily inflow prediction are explained in the following sections.

Study Area and Methodology

The multipurpose Koyna Reservoir having a global coordinate latitude of 17°00'N–17°59'N and longitude of 73°02'E–73°35'E located (Fig. 1) in west coast of Maharashtra, India, is taken up as a case study [6, 13]. The average annual inflow into the reservoir is about $3809.21 \times 10^6 \text{ m}^3$ (during the period of 1961–2009) with an average annual rainfall of 4654 mm from an area of 891.78 km². The observed daily inflow time series at Koyna Reservoir shows that the river is an intermittent river and during non-monsoon periods, the inflow is zero [12]. There are difficulties in developing a conventional model for prediction of inflow, particularly, for an intermittent river with large number of zero values as input (Magar and Jothiprakash [12]). In view of this, daily inflow forecasting is very much needed for flood warning and daily reservoir operations [3]. The variation of daily statistical analysis of the observed time series is shown in Table 1.

Fig. 1 Location map of Koyna Reservoir (modified after [13])

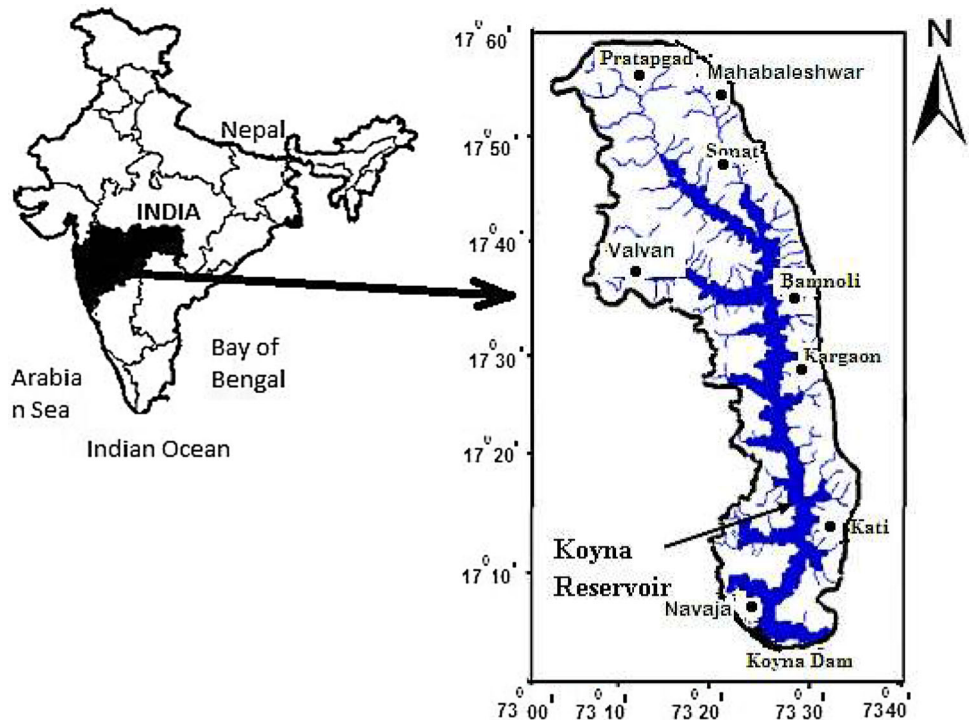


Table 1 Statistical properties of Koyna Reservoir daily inflow (from 1961 to 2009) (modified after [9])

S. no.	Statistical property	Full year daily data values
1	Average $\times 10^6 \text{ m}^3$	10.43
2	Std deviation $\times 10^6 \text{ m}^3$	25.92
3	Skewness	4.08
4	Kurtosis	21.33
5	Minimum daily flow $\times 10^6 \text{ m}^3$	0
6	Maximum daily flow $\times 10^6 \text{ m}^3$	328.58
7	Coefficient of variation	2.49
8	No. of data points	17,897
9	No. of zeros	15,088
10	% of the zeros	84.31

Table 1 shows that observed flow is positively skewed with high peakedness. Thus, maximum values are crucial in filling the reservoir and have to be predicted accurately. The percentage of zeros in the time series is 84.31%, due to zero inflows during non-monsoon period (November–May). The observed data show high coefficient of variance indicating that transformation of the data has to be carried out to follow normal distribution. The number of zero inflows and higher variation in coefficient of variation possess great difficulty in using conventional stochastic techniques to predict future values especially with this small daily time step prediction.

Model Tree (MT) Model Development

The M5 model tree technique as suggested by Solomatine and Xue [26] is developed and applied. The leaves have linear regression function which in order represents various attributes of the series for its prediction. MT process is based on information theory principle [19, 25] according to which the multi-dimensional parameter space can be split into different models which belongs to different attributes by automatic generation models based on the overall quality criterion. This theory assumes that the functional dependency of the whole domain is not constant and in fact, it can be approximated on small domains. The model leaves are linear functions separated into pieces; thus, MT is also known to be piecewise linear model and is the intermediary between linear and nonlinear models [28]. The principle of splitting the series is according to the attribute which is most suitable to divide that portion ‘*T*’ of the training data, that reaches a particular node. For this split, the measure of error is taken as standard deviation of class values of ‘*T*’ at that node by computing the expected error at each node, i.e., testing at each attributes. The equation used is as follows:

$$SDR = sd(T) - \sum_i \frac{|T_i|}{|T|} sd(T_i) \tag{1}$$

sd—standard deviation, *T*—instances reached the node, *T_i*—the subset of instances having *i*th outcome of the potential set. The stopping criteria of splitting are when the

sd of all T_i that reached the node with just a small fraction (i.e., less than 5%) as compared to the original instance and when instances remaining are only few [10]. In model tree process, the leaf means the set of some attributes that will make one linear model and the tree will be followed by the leaf by making use of attributes of instances at each node. Once a leaf is formed, it is evaluated for the test instances for rough prediction of next value. With growth of tree, there are several steps involved such as (1) error calculation, (2) simplification and (3) pruning and smoothing. The division of models often produces over-elaborated structures which in turn need to be pruned back, and as a result, the number of models gets reduced in numbers. This pruning of the leaves will last till the expected estimated error decreases; data set of expected error estimate is used in pruning. The leaf which predicts raw test instances for next values has sharp discontinuities which can be removed by means of model smoothing. MT model developed in the present study used observed daily inflow of previous time step as input and predicted inflow of next day as output. The data length is divided into 70% for training and 30% for testing. The leaves containing linear models are developed and then simplified by pruning and smoothing process. Also each input is checked for four conditions of model tree type, i.e., pruned and smoothed models (PS), un-pruned and un-smoothed models (UPUS), un-pruned and smoothed (UPS) and pruned and un-smoothed (PUS). The number of inputs given for the present study is the antecedent inflows up to eight lags.

Performance Evaluation of the Models

Apart from graphical evaluation using time series plots and scatter plots between observed and predicted values, the performance of the models is measured with well-defined statistical performance measures such as mean square error (MSE) which measures the goodness of fit relevant to high flows; mean absolute error (MAE); coefficient of correlation (R); Nash–Sutcliffe efficiency (E); Akaike information

criterion (AIC); and Bayesian information criterion (BIC) [2, 21].

Results and Discussions

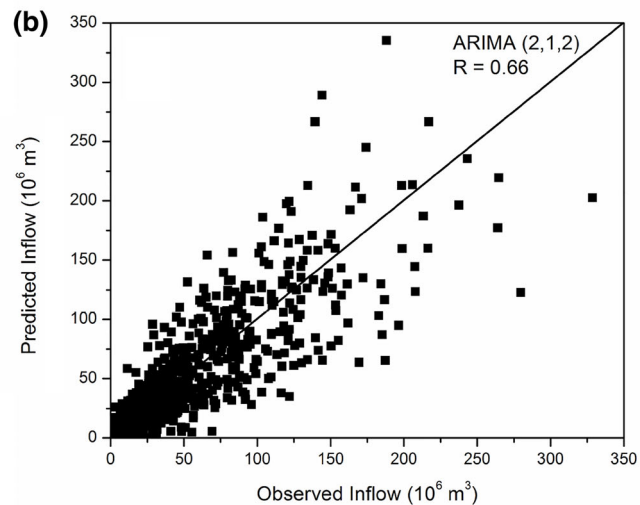
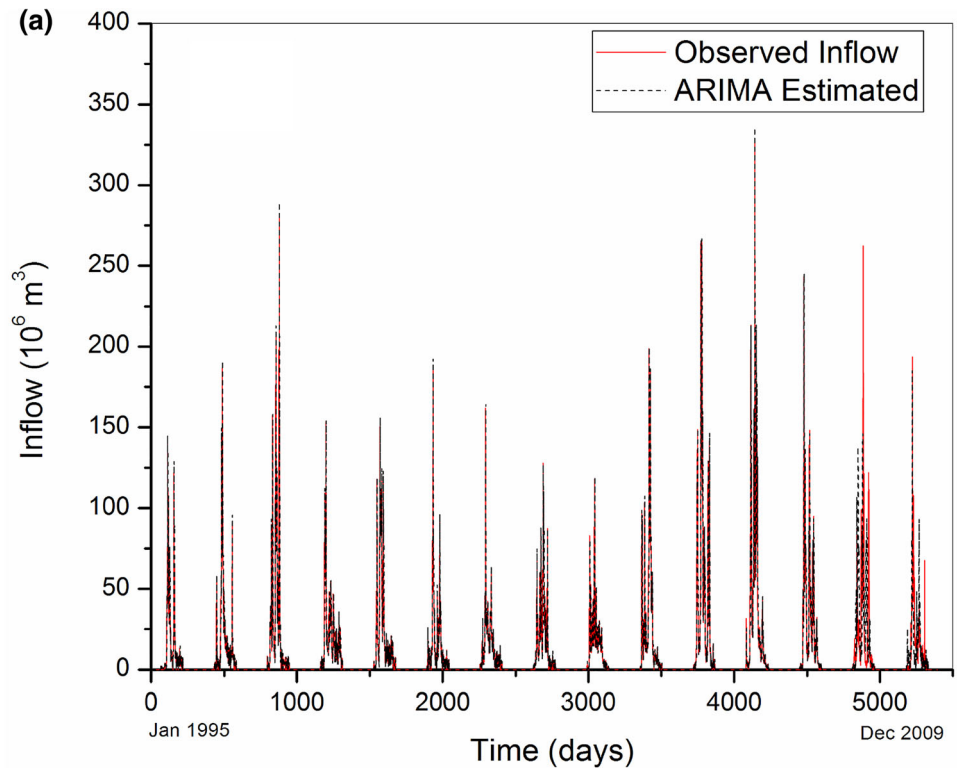
The daily inflow time series (1961–2009) observed over a period of 49 years was used in developing and evaluating the daily time step ARIMA and MT models. From the data analysis, it is found that the observed data are not following a normal distribution [12] and hence, the stochastic models are developed using the log-transformed series. It is to be remembered that data series has very large zero values, while \log transformation of the observed data the zero values are taken as 0.0001 an insignificant inflow. All the models are developed using 34 years (70%) of length of data and remaining 15 years (30%) is used for testing. After predicting the next time step inflow, the values are back transformed to arrive at the inflow values. The performance measures of ARIMA model for various model forms during development and testing stages while predicting the daily reservoir inflow series are given in Table 2.

Table 2 shows the performance by the stochastic models for various p, d, q combinations out of which ARIMA (2-1-2), model can be considered as the best model. This ARIMA (2-1-2) (marked * in Table 2) stochastic model is selected as a best stochastic model based on the performance in terms of minimum AIC and BIC, showing as parsimonious model. Other statistical performances like RMSE and E show poor performance of ARIMA model. The correlation coefficient is around 0.66 during testing, indicating that the ARIMA model could not predict this daily time step inflow values accurately and secondly the value 0.66 may be due to large data set used while testing (5475 points). The time series as well as scatter plot of the ARIMA (2-1-2) model's predicted and observed inflows during testing period is shown in Fig. 2. The scatter plot shown in Fig. 2a shows good prediction around average

Table 2 Performance analysis of ARIMA model

ARIMA (p-d-q) models	Performance during model development period					Performance during model testing period				
	AIC	BIC	RMSE (m ³ /s)	E (%)	R	AIC	BIC	RMSE (m ³ /s)	E (%)	R
1-1-1	34,656.17	34,663.56	17.89	58	0.62	14,952.8	14,959.35	18.23	57	0.60
2-2-2	32,515.17	32,522.56	14.97	57	0.65	14,057.17	14,063.72	15.32	55	0.62
1-2-1	32,634.96	32,642.35	15.102	54	0.63	14,104.02	14,110.57	15.46	51	0.64
1-2-2	32,745.7	32,753.09	15.26	52	0.61	14,053.81	14,060.35	15.31	51	0.59
2-1-1	32,547.23	32,554.62	15.01	59	0.64	13,996.3	14,002.85	15.14	58	0.63
2-1-2*	31,718.85	31,726.25	14.01	58	0.64	13,945.02	13,951.57	14.99	56	0.66
3-2-2	32,682.54	32,689.94	15.18	47	0.45	14,312.92	14,319.47	16.10	45	0.43
4-1-4	34,980.83	34,988.22	18.38	42	0.45	14,910.25	14,916.8	18.08	40	0.43

Fig. 2 Comparison of **a** time series and **b** scatter plot of observed and predicted daily inflow by ARIMA during testing period



value of inflows, but for high inflows, the ARIMA model either under-predicted or over-predicted. Thus, it may be concluded that stochastic model (ARIMA) could not capture the highly nonlinear peak values in the daily reservoir inflow series. To improve the peak inflow prediction accuracy, the nonlinear soft computing techniques namely MT have to be applied for same data in hand.

The above-discussed MT model has been developed and applied to predict daily observed reservoir inflow using WEKA software version 3.5 (<http://www.cs.waikato.ac.nz/ml/weka/>). The time series data are not transformed, since

MT model technique does not need the condition assumption that data should follow normality. The daily time series is divided into 12,421 instances (34 years, from January 1, 1961, to December 31, 1994) for training and 5476 instances (15 years from January 1, 1995, to December 31, 2009) of data length for testing, and this split up length is same as that used in developing and testing of ARIMA model. Varieties of MT models were developed by varying the number of input from one antecedent inflow to eight antecedent inflows as input. The MT model with one lagged input is referred as MT1 and lagged by two

inputs as MT2, likewise so on. Thus, MT7 means seven antecedent inflows are given as input to predict one time step ahead output (the next time step inflow). Each MT model is trained and tested for four combinations of MT modeling technique such as pruned or un-pruned and smooth or un-smoothed. The MT models are evaluated with the same statistical performance indices used for evaluating ARIMA model. From the inter-comparison, it is found that UPUS MT model outperformed various other combinations in their respective input category. The reason for poor performance of pruned model is that leaves having smaller number of instances or lesser efficient are pruned, lesser number of instance leaves happens to be the peak value predicting leaves, pruning such leaves means cutting down the linear equations that predict such peak values, leading to poorer performance of peak value, because the number of instances are very less in peak value. Thus, the results of MT for the combination of UPUS are alone discussed.

The best performance of un-pruned and un-smoothed MT models during testing period is depicted in Table 3. At instance, the performance statistics of all the MT models depicts that the results of MT predicted inflows are much better than ARIMA models. From Table 3, it is found MT7 daily inflow model (marked * in Table 3) with seven input variables outperformed other input with a performance statistics of R (0.9812), MAE (1.9082) and RMSE (4.897). Table 3 shows that with the increase in number of inputs, there is considerable improvement in the efficiency of the MT model. The number of leaves or rules is coming out to be very large for every MT model as shown in Table 3, but with much advanced computing facility it is not a big problem while using them in real-life prediction at dam site. It is clear that the performance of MT model goes on increasing with the increase in number of inputs up to seven (MT7) and further increase in the input (MT8) reduced the performance of the model. In real life 7 days

represents previous week daily inflow as input for the model. Therefore, the best results found with seven lagged input, i.e., MT7, are presented here. The performance after 7 inputs was deteriorating as seen for MT8 model. The model has been stopped at eight lagged input because for further lagged inputs the performance is coming out to be less.

Figure 3 shows the time series (Fig. 3a) as well as scatter plot (Fig. 3b) of the observed and predicted inflow by MT7 UPUS model. From the scatter plot (Fig. 3b), it is very clear that MT model has outperformed ARIMA model. The entire range of variations in the observed data is well captured by the MT model. From the results, it may be concluded that for smaller time step reservoir inflows having highly nonlinear relationship and large zero values, UPUS-MT models are more suitable. MT technique has predicted the peak inflow as well as low inflow much better than ARIMA model. It is also found that UPUS-MT models edged the performance than the pruned and smoothed MT models. ARIMA has well predicted the moderate inflow, but peak and low inflows are under-predicted. MT models are found more advantageous in many aspects, such as less time-consuming, no parametric estimation, no prior knowledge and easy model setting, and the main advantage is understandable input and output in the form of linear equations.

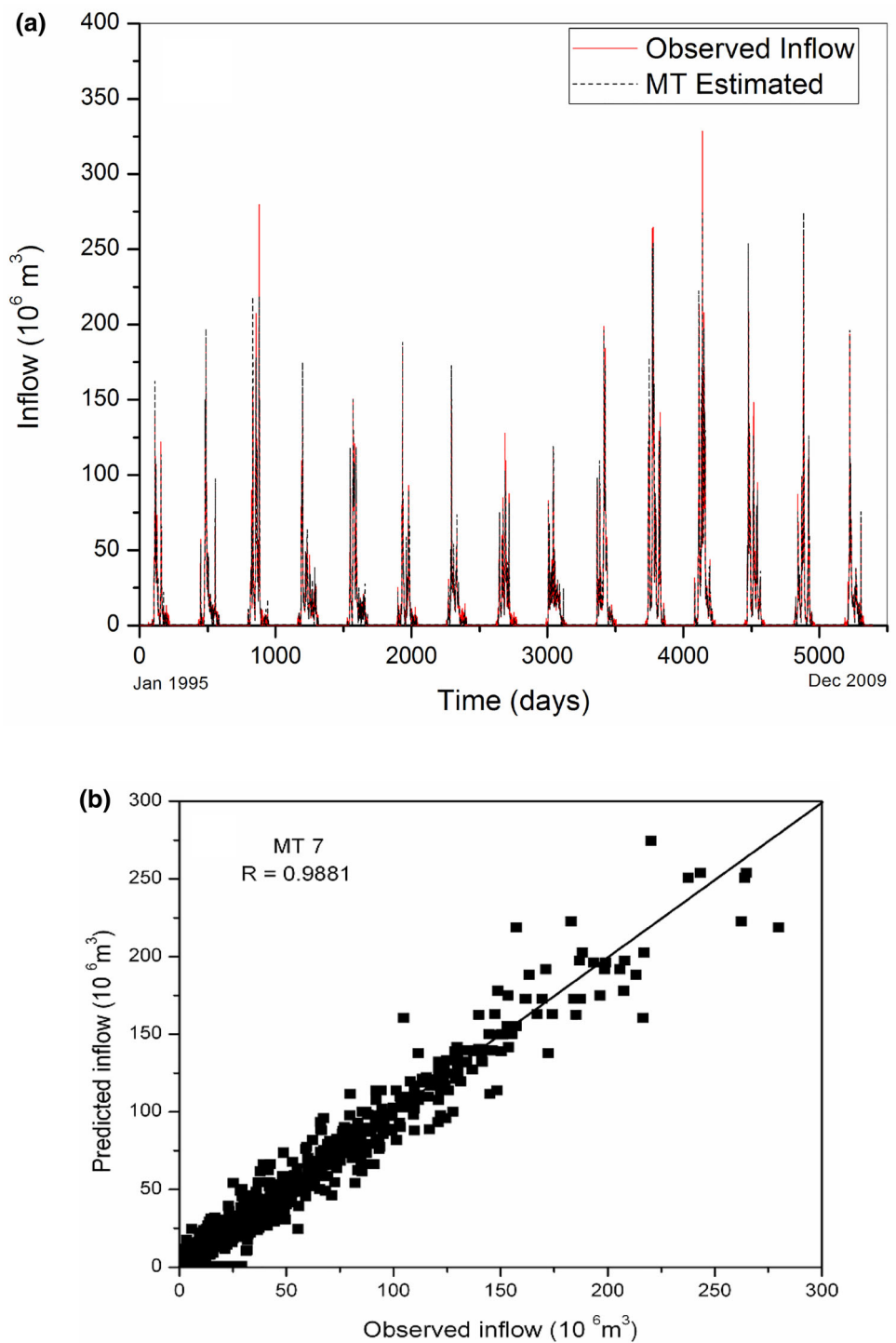
Conclusions

This paper presented two popular techniques of reservoir inflow prediction, conventional stochastic ARIMA models and data-driven MT technique based on AI, applied to the Koyna watershed, Maharashtra. Statistical performance measures along with scatter plot and time series plots were evaluated to find the performance of the developed model. The disadvantage of ARIMA models is that they are

Table 3 Performance analysis of MT

Model	Performance of MT (UPUS) model									
	Performance during model development period					Performance during model testing period				
	R	E	MAE (m ³ /s)	RMSE (m ³ /s)	No. of leaves	R	E	MAE (m ³ /s)	RMSE (m ³ /s)	No. of leaves
MT1	0.9516	0.9032	3.164	7.795	1383	0.9577	0.9172	2.932	7.816	671
MT2	0.9655	0.9327	2.529	6.604	1483	0.9706	0.9422	2.438	6.523	673
MT3	0.973	0.9325	2.25	5.859	1452	0.9765	0.9535	2.0925	5.859	643
MT 4	0.9775	0.9524	2.0929	5.348	1437	0.9816	0.9635	1.912	5.1938	640
MT 5	0.9783	0.8527	2.0054	5.256	1412	0.9816	0.883	1.8074	5.1902	651
MT 6	0.9804	0.9573	1.9368	4.992	1413	0.9839	0.9681	1.7897	4.9923	620
MT 7*	0.9812	0.9652	1.9082	4.898	1406	0.9881	0.9764	1.9082	4.8976	623
MT 8	0.967	0.9372	2.131	5.923	1400	0.9721	0.9459	3.012	6.0021	627

Fig. 3 Comparison of **a** time series and **b** scatter plot of observed and predicted daily inflow by UPUS MT7 during testing period



developed based on the assumptions of stationarity and linearity, which is not a requirement in case of MT model. ARIMA models have failed to capture nonlinear peak inflows, whereas MT model has predicted the peak flows better than ARIMA models. Also MT requires less input from modeler (thus human error is less) and has easy understanding and implementation in field. These advantages made the MT model a better choice than the ARIMA

model. MT results are found much better even with raw data, whereas the ARIMA model requires transformation of the data to model. Thus, MT can be applied to non-normal data set also and is much better than ARIMA for all the conditions of pruning and smoothing of the MT. UPUS-MT model showed 89% better performance than ARIMA models in terms of MSE. Based on the results of the study, it may be concluded that MT models predict reservoir

inflow much better than ARIMA model for smaller time step like daily.

Acknowledgements The authors convey their sincere thanks to Executive Engineer, Irrigation Department and office of Koyna Dam Division, Government of Maharashtra, India, for providing necessary data to carry out this research work and also for giving feedback on the developed MT equations to apply at dam site.

References

1. J. Adamowski, Development of a short-term river flood forecasting method for snowmelt driven floods based on wavelet and cross-wavelet analysis. *J. Hydrol.* **353**(3–4), 247–266 (2008)
2. H. Akaike, A new look at the statistical model identification. *IEEE Trans. Autom. Control* **19**, 716–723 (1974)
3. R. Arunkumar, V. Jothiprakash, Optimal reservoir operation for hydropower generation using non-linear programming model. *J. Inst. Eng. India Ser. A* **93**(2), 111–120 (2012)
4. B. Bhattacharya, D.P. Solomatine, Neural networks and M5 model trees in modelling water level-discharge relationship. *Neurocomputing* **63**, 381–396 (2005)
5. B. Bhattacharya, R.K. Price, D.P. Solomatine, A machine learning approach to modelling sediment transport. *ASCE J. Hydraul. Eng.* **133**(4), 440–450 (2007)
6. CDO, Final report on revised flood study for Koyna Dam, Central Design Office, Irrigation Department. Government of Maharashtra, India, 1992
7. V. Jothiprakash, A. Kote, Effect of pruning and smoothing while using M5 model tree technique for reservoir inflow prediction. *J. Hydrol. Eng.* **16**(7), 563–574 (2011)
8. V. Jothiprakash, S.A. Kote, Improving the performance of data driven techniques through data pre-processing while modeling the daily reservoir inflow. *Hydrol. Sci. J.* **56**(1), 168–186 (2011)
9. V. Jothiprakash, R.B. Magar, Multi-time-step ahead daily and hourly intermittent reservoir inflow prediction by artificial intelligent technique using lumped and distributed data. *J. Hydrol.* **450–451**, 293–307 (2012)
10. A.S. Kote, Single reservoir and multi-reservoir inflow prediction using artificial intelligent techniques. Ph.D. thesis report, Indian Institute of Technology Bombay, Mumbai, 2010
11. A.S. Kote, V. Jothiprakash, Stochastic and artificial neural network models for reservoir inflow prediction. *J. Inst. Eng. India* **90**(18), 25–33 (2009)
12. R.B. Magar, V. Jothiprakash, Intermittent reservoir daily-inflow prediction using lumped and distributed data multi-linear regression models. *J. Earth Syst. Sci.* **120**(6), 1067–1084 (2011). <https://doi.org/10.1007/s12040-011-0127-9>
13. R.B. Magar, Real time reservoir inflow using soft computing techniques. Ph.D. thesis report, Indian Institute of Technology Bombay, Mumbai, 2011
14. T. Mandal, V. Jothiprakash, Short term rainfall prediction using ANN and MT techniques. *ISH J. Hydraul. Eng.* **18**(1), 28–37 (2012)
15. M. Momani, P.E. Naill, Time series model for rainfall data in Jordan: case study for using time series analysis. *Am. J. Environ. Sci.* **5**(5), 599–604 (2009)
16. D.J. Nokes, I. McLeod, K.W. Hipel, Forecasting monthly river flow time series. *Int. J. Forecast.* **1**, 179–190 (1985)
17. E.K. Onyari, F.M. Ilunga, Application of MLP neural network and M5P model tree in predicting streamflow: a case study of Luvuvhu Catchment, South Africa. *Int. J. Innov. Manag. Technol.* **4**(1), 11–15 (2013). <https://doi.org/10.7763/IJIMT.2013.V4.347>
18. M. Pal, S. Deswal, M5 model trees based modeling of reference evapotranspiration. *Hydrol. Process.* **23**, 1437–1443 (2009)
19. J.R. Quinlan, Learning with continuous classes, in *Proceedings of the Australian Joint Conference on Artificial Intelligence* (World Scientific, Singapore, 1992), pp. 343–348
20. M.J. Reddy, B.N.S. Ghimire, Use of model tree and gene expression programming to predict the suspended sediment load in rivers. *J. Intell. Syst.* **18**(3), 211–227 (2009)
21. J. Rissanen, Modeling of short data description. *Automation* **14**(465), 471 (1978)
22. M.T. Sattari, M. Pal, K. Yurekli, A. Ünluğara, M5 model trees and neural networks based modelling of ETO in Ankara, Turkey. *Turk. J. Eng. Environ. Sci.* **37**(2), 211–219 (2013)
23. B. Sivakumar, R. Berndtsson, *Advances in Data-Based Approaches for Hydrologic Modeling and Forecasting* (World Scientific, Singapore, 2010)
24. D.P. Solomatine, K.N. Dulal, Model trees as an alternative to neural networks in rainfall-runoff modeling. *Hydrol. Sci. J.* **48**(3), 399–411 (2003)
25. D.P. Solomatine, M.B.L.A. Siek, Flexible and optimal M5 model trees with application to flow predictions, in *Proceedings of 6th International Conference on Hydroinformatics* (2004). ISBN 981-238-787-0
26. D.P. Solomatine, Y. Xue, M5 model tree and neural networks: application to flood forecasting in the upper reach of the Huai River in China. *J. Hydrol. Eng. ASCE* **9**(6), 491–501 (2004)
27. L. Štravs, M. Brilly, Development of a low-flow forecasting model using the M5 machine learning method. *Hydrol. Sci. J.* **52**(3), 466–477 (2007)
28. H. Witten, E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementation* (Morgan Kaufmann, San Mateo, 2000)
29. K. Yurekli, A. Kurunc, H. Simsek, Prediction of daily maximum streamflow based on stochastic approaches. *J. Spat. Hydrol.* **4**(2), 1–12 (2004)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.