Hydrol. Earth Syst. Sci. Discuss., 10, C506–C517, 2013
www.hydrol-earth-syst-sci-discuss.net/10/C506/2013/

Interactive comment on "Assessing the predictive capability of randomized tree-based ensembles in streamflow modelling" *by* S. Galelli and A. Castelletti

S. Galelli and A. Castelletti

castelle@elet.polimi.it

Received and published: 26 March 2013

NOTE: *he original comments by Prof. Renata Romanowicz (**R**) are in italics. Replies by the authors (**A**) are in regular text. Tables and Figires used in this review are reported in bold text.*

R: *General comment*

The authors present a comparison of the predictive capability of extremely randomized trees (Extra-Trees), including the accuracy, computational efficiency and explanation ability, with other tree-based methods and data-driven approaches in modelling rainfall runoff processes. The paper is of interest to the hydrological community. It is well

written and well structured. Its main novelty lies in the analysis of the applicability of the Extra-Trees approach to stream flow modelling. The authors follow a four-step assessment procedure, including random sampling of observational data, and a multi-criteria evaluation of model performance and they provide an uncertainty analysis of model performance. The method is applied to two catchments, a small urban Marina catchment in Singapore and the Cunning River, a large natural catchment in Australia. The authors compare the predictive capability, explanation ability and computational efficiency of the studied method to other data-driven methods (M5, CART, ANN and MLR). In particular, they compare the scatter plots of predicted and measured flows using all five methods and analyse the probability distributions of residuals.

A: We wish to thank Professor Romanowicz for her positive evaluation of the paper and useful comments. We agree that Extra-Trees performance should be better discussed and investigated (for example, performance for high and low flows), and we are willing to modify the paper following the reviewer's suggestion as explained point-by-point in the following.

R: *Specific comments*

1. The authors claim that the Extra-Trees method out-performs the other methods. Its largest advantage lies in its ability to rank the importance of the model input variables. Otherwise, its superiority is not so obvious. The M5 method has better results for all goodness-of-fit criteria. When it comes to computer efficiency, both Extra-Trees and M5 have much bigger computer time requirements than other methods. The authors are asked either to change their Conclusions or explain how they can justify their statement.

A: We thank Professor Romanowicz for her comment. Indeed, the conclusions are not in perfect agreement with our findings. From a predicting accuracy point of view we can claim that Extra-Trees have the same performance as M5 (the best performing model in the benchmarking exercise) on low and intermediate flows only (see reply no. 4), while their performance slightly decreases on base and high flow conditions. There is a structural reason for this behaviour, which is due to the Extra-Trees architecture (as commented in Section 5.2 and reply no. 4). Also, we also agree with the reviewer that Extra-Trees and M5 have larger computational requirements than the other models (although in the ANNs computational requirements it is necessary to account for 100 random initializations); however, this increase in the computational requirements is compensated by a better predicting accuracy. In synthesis, we can claim that Extra-Trees represent a good compromise between predicting accuracy and computational requirements, especially when adopted for large datasets (such as Marina catchment). In addition, Extra-Trees have the ability to rank the relative importance of the input variables. We will thus modify the conclusions in the new manuscript, where we will highlight the above-mentioned points.

R: *2. A detailed examination of the scatter plots (Fig. 5) shows that Extra-Trees under-predicts high flows for the Marina catchment. However, it is not visible in the error distribution shown in Fig. 7. The authors show the results of the fitted logistic distribution, which hides the model performance. It would be more informative if the empirical distributions were also shown.*

A: Yes, Extra-Trees tend to under-predict high flows in Marina catchment. As discussed in Section 5.2, the reason for this behaviour stands in the model architecture: Extra-Trees predictions correspond to the average of the output values associated to the inputs within a specific leaf. This can slightly limit their predictive capabilities, especially for high flows regimes, since the (generally few) flow peaks registered in the training dataset are averaged out in the model leaves. Both the fitted logistic and empirical probability distributions (**Figure 1** in this review file) are not very useful to address this modelling behaviour. This is because the majority of the inflow events (and thus model errors) are in the range 0 to about 20 m^3/s (see reply no. 4), while

the amount of flow peaks is too small to significantly change either the fitted or empirical pdfs. This point will be further commented in the revised version of the manuscript. However, we believe that the empirical distributions provide additional information on the models behaviour at low and intermediate flows, so we propose to still include them in the new manuscript. In Marina catchment (panel c), the empirical distributions show that Extra-Trees, M5 and CART have symmetrical distributions with low predictive uncertainty, while both ANNs and linear models are characterized by more prominent asymmetry and kurtosis. On the other hand, the empirical distributions for the Canning river (panel d) simply confirm the findings of the fitted logistic one.

R: *3. It is also not clear what time periods were used during the comparison of the scatter plots and the error distributions. The reader assumes that they were the same, but this should be clearly stated.*

A: For both scatter plots and error distribution analysis we used the models predictions (and errors) on the testing subsets, which were generated according to the sampling described in Section 3.2. We understand that this was not clearly explained within the text, so we will introduce this explanation in the revised version of the manuscript.

R: *4. In summary, the performance of the Extra-Trees method in predicting flow is the most interesting to the hydrological community, and that subject should be the main focus of the paper. At the moment, this particular point is not well explored, and the advantages of using the method are not convincing. In particular, the authors should comment on the ability of the methods they compare to reproduce the flow patterns, with a discussion of the model performance for high and low flows.*

A: We understand that the prediction accuracy is a key element of our three-fold evaluation (also consisting of explanation ability and computational efficiency) that is not exhaustively explored in the original manuscript, where it is limited to the analysis of different multi-assessment criteria (i.e. NS, RMSE, RRMSE and MAE). To better comment on the ability of Extra-Trees in reproducing different flow regimes we thus propose to i) identify four regimes (i.e. base flow, low flow, intermediate flow and high flow) and ii) analyze the models behaviour over these specific ranges.

The identification of flow regimes is based on the calculation of specific percentile values on the testing datasets. For both Marina catchment and Canning River, we categorized the flow regimes by using the 25^{th}, 75^{th}, 95^{th} and 99.5^{th} percentiles as in **Table 1**.

Both Marina catchment and Canning River are characterized by prolonged periods of low or null flow: in the former system this is due to the presence of large paved areas (reduced infiltration) and concrete lined canals, while in the latter this is due to the ephemeral nature of the river during the summer period. Because of this prolonged periods of no flow, 75% of the observations (in both cases) falls below what we categorized as base flow. The other extreme of the flow regimes (i.e. high flow) also corresponds to a high percentile value. This effect is due to the positive skewness characterizing both datasets, especially Marina catchment. The different flow regimes are represented in **Figure 2** and **Figure 3** in this review file, which we propose to introduce in the new manuscript in place of Figure 5.

Table 2 and **table 3** report the value of RMSE of the models predictions over the identified flow regimes. It shows that M5 and Extra-Trees are the most performing models, with M5 having better performance on base flow and low flow conditions. As explained in Section 5.2, the Extra-Trees tendency in underestimating high flows and over estimating low flows is due to the model architecture ("average of the output values associated to the inputs falling a specific leaf"). This is not the case with M5, which have a linear model in the final (pruned) leaves, and this allows them to extrapolate over unseen events. Unsurprisingly, other models that do not make use of a regression system (such as CART) have underestimation problems for high flows (and overestimation problems for low flows). On the other hand, Extra-Trees predictive

capability is either comparable or better than M5 one on low and intermediate flow conditions. We believe that this analysis could be used to expand Section 5.2 (lines 600 – 607), as it will allow to better analyzing the models behaviour without compromising the paper length and readability.

R: *5. The results obtained from the ranking of input variables look promising, but the outcome seems to be too obvious. For the Marina catchment, the influence of rainfall is the highest due to small catchment retention. For the Cunning River, the flow values in two previous days have the largest impact on flow predictions. The results suggest that Extra-Trees could be useful in exploring the dependence of input ranking on flow under varying meteorological conditions.*

A: This is indeed the result that we were expecting to obtain. The idea of adopting two case studies of 'known behaviour' (i.e. dependency on rainfall events in Marina catchment and on flow values for the Canning river) was aimed at showing and empirically validating the capability of Extra-Trees in exploring the relative importance of the input variables. Hopefully, Extra-Trees could then be adopted in more complex domains, for example under varying meteorological conditions. We will include this comment in the new version of the manuscript.

R: *6. It would also be advantageous if the authors showed the predictions with the confidence limits for the Extra-Trees method. The information contained in the error analysis does not give a proper perspective.*

A: We thank the Professor Romanowicz for the suggestion, and we propose to replace Figure 6 with **Figure 4** in this review file, where we compared measured and predicted streamflow for Extra-Trees and M5 only, and we then show the confidence limits for the Extra-Trees. The rational of this choice is to concentrate the graphical analysis of measured and predicted hydrographs on the two most performing models only.

Table 1. Categorized flow regimes.

Flow regime	Percentile	Marina catchment Flow limit [m^3/s]	Canning River Flow limit [m^3/s]
Base flow	$< 75^{th}$	3.17	0.16
Low flow	75^{th} – 95^{th}	3.17 – 17.10	0.16 – 1.44
Intermediate flow	95^{th} – 99.5^{th}	17.10 – 169.48	1.44 – 7.73
High flow	$>99.5^{th}$	169.48	7.73

Table 2. Testing results (RMSE [m^3/s]) of Extra-Trees and benchmarking models for Marina catchment dataset on four different flow regimes.

Model	Base flow	Low flow	Intermediate flow	High flow
Extra-Trees	2.143	4.562	31.636	141.273
M5	1.724	4.244	31.695	139.745
CART	2.869	6.567	35.357	152.091
ANNs	2.699	8.816	40.662	142.126
MLR	2.225	7.610	33.628	141.081

R: *7. An additional, important comment is on the quality of the figures. The figure labels are too small. I had to take a magnifying glass to be able to see the details.*

A: We will modify the labels as requested.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 10, 1617, 2013.

Table 3. Testing results (RMSE [m^3/s]) of Extra-Trees and benchmarking models for Canning River dataset on four different flow regimes.

Model	Base flow	Low flow	Intermediate flow	High flow
Extra-Trees	0.014	0.148	0.858	3.067
M5	0.012	0.170	0.885	2.389
CART	0.017	0.158	1.279	3.407
ANNs	0.156	0.211	0.980	3.211
MLR	0.104	0.148	0.830	3.168

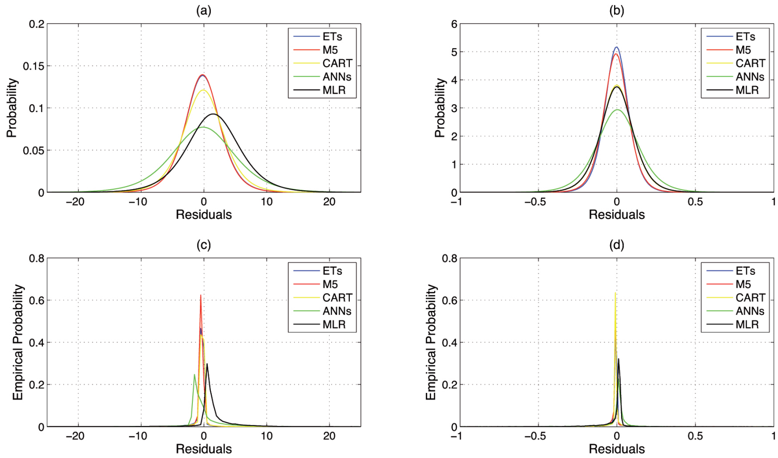

Fig. 1. Probability distribution (fitted logistic and empirical) of the models residuals in Marina catchment (a, c) and Canning river (b, d) on the testing datasets.

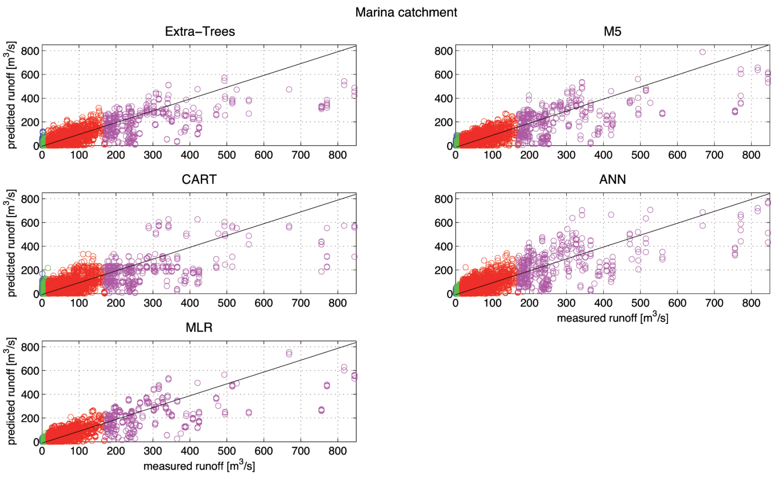

Fig. 2. Scatter plot of predicted and measured streamflow in Marina catchment for the different models (on the testing subset). Different colours are used for the different flow regimes: blue for the base flo

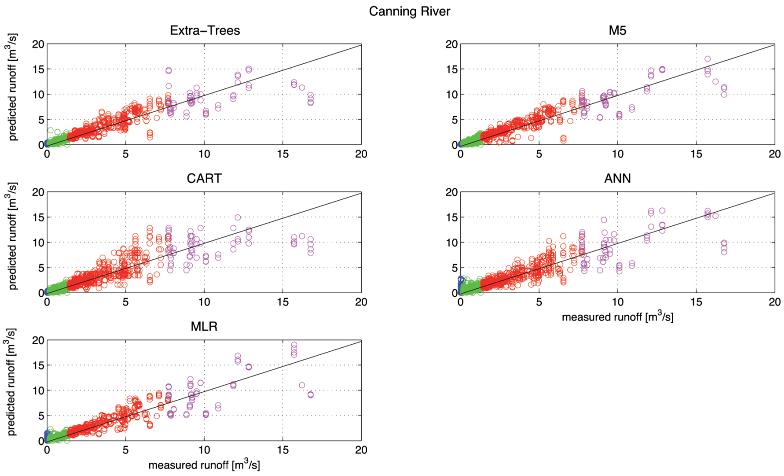

Fig. 3. Scatter plot of predicted and measured streamflow in Canning River for the different models (on the testing subset). Different colours are used for the different flow regimes: blue for the base flow,

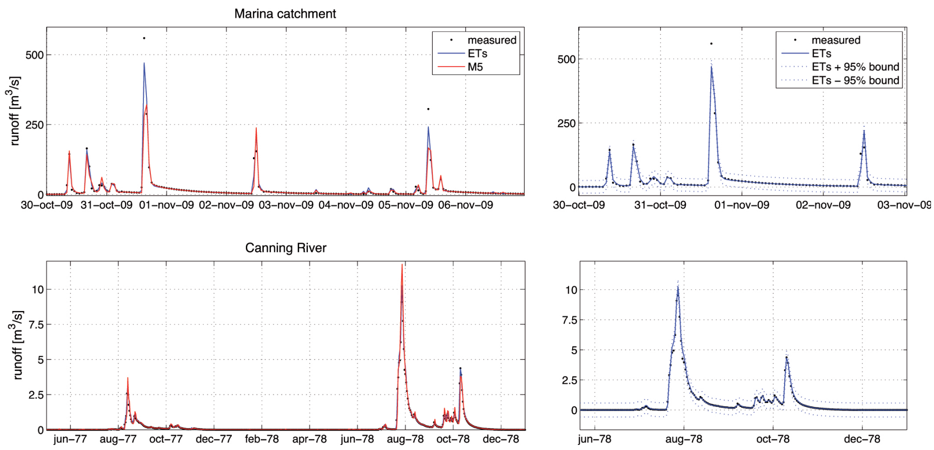

Fig. 4. Comparison between the measured and predicted streamflow for Marina catchment and Canning River (left panels), and comparison between measured and predicted streamflow with Extra-Trees and 95\% confid