Journal of Hydrology 600 (2021) 126459

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

Discussion

Closure to the discussion of Ebtehaj et al. on "Comparative assessment of time series and artificial intelligence models to estimate monthly streamflow: A local and external data analysis approach"

Saeid Mehdizadeh [a,*], Farshad Fathian [b], Mir Jafar Sadegh Safari [c], Jan Adamowski [d]

[a] Department of Water Engineering, Urmia University, Urmia, Iran
[b] Department of Water Science and Engineering, Faculty of Agriculture, Vali-e-Asr University of Rafsanjan, P.O.Box 77188-97111, Rafsanjan, Iran
[c] Department of Civil Engineering, Yaşar University, Izmir, Turkey
[d] Department of Bioresource Engineering, Faculty of Agricultural and Environmental Sciences, McGill University, Canada

ARTICLE INFO

This manuscript was handled by Geoff Syme, Editor-in-Chief

ABSTRACT

In this closure, we respond to the comments of Ebtehaj et al. (2020), and also provide additional details regarding several features of our study.

1. Introduction

The discussion paper of Ebtehaj et al. (2020), hereafter referred to as EZB, provided several comments related to our Mehdizadeh et al. (2019a) paper. In this closure, we respond to the comments of EZB, and further explain the aims and reasoning behind our approaches and models, and show why our approaches and models are suitable based on the aims and scope of our study. We also provide additional details regarding several features of our study.

EZB had five main points in their discussion paper that we address in this closure: 1. Individual linear time series models and hybrid linear-ARCH models (ARCH = autoregressive conditional heteroscedasticity): We explain in more detail our aims, scope, reasoning and approach used to develop our individual linear time series and hybrid linear-ARCH models, and explain why our models were suitable given our aims. 2. Artificial intelligence models: We explain why data preprocessing can be a useful step in artificial intelligence modeling and why it was used in our study. 3. Hybrid models: We explain that comparing the performance of ARCH models 'alone' with our other models, as suggested by EZB, did not make sense since our ARCH models were fitted to the residuals of the linear time series models (to remove the ARCH effects in the residuals and build the hybrid linear-ARCH models). We also point out that developing hybrid time series – artificial intelligence models (and other possible types of hybrid models), as suggested by EZB, was outside the scope of our study, which aimed to develop hybrid linear-ARCH models. 4. External analysis: We explain why our external analysis approach was valid and useful. 5. RMSE, MAE and R: We explain why our performance indices (i.e., RMSE, MAE and R) were suitable to compare our models.

We want to point out that the data analysis and model results/comparisons, and comments, provided by EZB in their discussion paper (regarding their approaches and results and our approaches and results in our original study) often used different data sets (e.g., EZB used raw/original data for their artificial intelligence models while we used preprocessed/transformed data in our original study), or different model development approaches (e.g., EZB used different, and more than twice the number of inputs for their artificial intelligence models compared to our artificial intelligence model inputs in our original study), or different methods (e.g., EZB used the Kwiatkowski-Phillips-Schmidt-Shin, Mann-Kendall and Mann-Whitney tests for their data analysis of the Port Elgin study site data although we believe these tests are not appropriate given the nature of our data), or included inaccurate assumptions (e.g., EZB assumed that we could compare the performance of our ARCH models 'alone' with our other models while this did not make sense given that the ARCH models in our study were fitted to the residuals of the linear time series models to remove the ARCH effects in the residuals and develop hybrid linear-ARCH models), all of which make simple/effective comparisons between the suggestions/approaches/results of EZB,

* Corresponding author.
E-mail address: saied.mehdizadeh@gmail.com (S. Mehdizadeh).

https://doi.org/10.1016/j.jhydrol.2021.126459
Received 12 May 2021; Received in revised form 12 May 2021; Accepted 12 May 2021
Available online 19 May 2021

and our approaches and results, not straightforward in some cases.

This closure is organized point by point to address the comments of EZB. It should be noted that because our aims and overall approach were the same in all four study sites in the original paper, and because EZB focused on one of our study sites (Port Elgin) for points 1 and 2, we do the same and focus on the Port Elgin study site in points 1 and 2 of our closure (to also avoid making an already long closure longer). Also, in this closure, when we refer to individual linear time series models we are referring to models such as AR (autoregressive) and MA (moving-average), when we refer to hybrid linear-ARCH models (ARCH = autoregressive conditional heteroscedasticity) we are referring to models such as AR-ARCH and MA-ARCH, and when we refer to artificial intelligence models, we are referring to a general class of models that encompass non-parametric regression methods such as MARS (multivariate adaptive regression splines) and evolutionary methods such as GEP (gene expression programming), developed in our study.

2. Point 1: Individual linear time series models and hybrid linear-ARCH models

We first want to provide some additional details regarding our aims and overall approach to develop our individual linear time series models and hybrid linear-ARCH models (the latter of which were one of the main topics of our original study). Following this, we respond to the specific points raised by EZB regarding our individual linear time series models (which EZB focus on in point 1).

In our original study, for each study site, we transformed the original data via our normalization and standardization process (which included de-seasonalization of the data), with the aim of providing us with the ability to explore the use of the simplest possible model classes in linear time series analysis for our individual linear time series models and to form our hybrid linear-ARCH models. The use of the original (i.e., not preprocessed/transformed) data in the study sites would have necessitated the use of ‘complex’ model classes in time series analysis, for example the seasonal ARIMA model class (given the nature of the original monthly streamflow data) and, given our aim to also incorporate conditional heteroscedasticity (given the ARCH-type effects in the residuals, i.e., significant autocorrelations in the *squares* of the residuals), this would have resulted in even more complex models. On the other hand, following our data transformation in each study site, we were able to focus on the simplest model classes in linear time series analysis (i.e., the AR and MA classes of time series models) - for our individual linear time series models and to form our hybrid linear-ARCH models - given the nature of the transformed data, and seeing that our selected models in each study site (which adhered to our aim of focusing on the simplest possible model classes in linear time series analysis for our individual linear time series models and to form our hybrid linear-ARCH models), offered a good fit to the data in the sense of the residuals having a white noise structure and practically zero autocorrelation. We developed hybrid linear-ARCH models (i.e., AR-ARCH and MA-ARCH in our case) since we observed ARCH-type effects in the residuals of the AR and MA models (i.e., significant autocorrelations in the *squares* of the residuals). The simplest model classes in linear time series analysis, for example the AR model class, have a simple structure, are easy to use and interpret, are widely used in hydrology (e.g., by practitioners), and have a long and well-developed literature (Salas et al., 1988; Rajagopalan et al., 2010). ARCH models, which are a class of non-linear time series models, are an approach that can remove ARCH effects in the residuals of linear time series models. We explored simple hybrid linear-ARCH models (i.e., AR-ARCH and MA-ARCH) for the reasons we have explained, and also because hybrid linear-ARCH models have not been investigated extensively in the hydrological literature and we were interested to see how these simple hybrid linear-ARCH models performed, and compared with simple individual linear time series models (i.e., AR and MA). We want to mention that some of the authors of the original study previously explored various more ‘complex’ model classes in time series analysis (e.g., the SARIMA model class; or the Autoregressive Fractionally Integrated Moving Average (ARFIMA) model class; etc.), and also previously explored coupling different time series analysis models (e.g., Vector AR; Self-Exciting Threshold AR; SARIMA; etc.) with ARCH and GARCH (Generalized ARCH) models in, for example, Fathian et al. (2016), Fathian et al. (2019a), Fathian et al. (2019b), Fathian et al. (2019c), and Mehdizadeh et al. (2020), so we were aware of other possible model options. However, for the reasons we have discussed, we developed simple individual linear time series (i.e., AR and MA) and simple hybrid linear-ARCH (i.e., AR-ARCH and MA-ARCH) models in our original study.

We now respond to the specific points raised by EZB, who focus on our individual linear AR and MA models in point 1. EZB first comment that the transformed data for the Port Elgin study site (which is the study site EZB focus on in point 1) is not stationary (based on their use of the Kwiatkowski-Phillips-Schmidt-Shin test), and has trend (based on their use of the Mann-Kendall test) and jump (based on their use of the Mann-Whitney test). EZB then comment that because of these points, our use of individual linear AR and MA time series models was not appropriate for the Port Elgin study site. We respectfully disagree; we show below that the transformed data (recall that our transformation included de-seasonalization) is stationary, and has no trend or jump, and our use of AR and MA time series models was suitable.

Fig. 1 shows the transformed Port Elgin data (after our preprocessing process described in the original paper). It is clear that the Port Elgin transformed data can suitably be modelled as a stationary process: there do not appear to be any clear non-stationarities therein according to Fig. 1. We used an advanced change point detection methodology - Wild Energy Maximization and gappy Schwarz Criterion (WEM.gSC) (Cho and Fryzlewicz, 2021) - for detecting (possibly multiple) change points in the mean of an autocorrelated time series (both our original and transformed time series exhibit clear serial correlation). The WEM.gSC change point detection methodology is comprised of two procedures that are combined: i) a solution path generation procedure that is based on the principle of ‘wild energy maximisation’ (i.e., the WEM component, which is useful to separate shifts in the mean from fluctuations stemming from serial correlations); and ii) an information criterion-based model selection strategy labelled ‘gappy Schwarz criterion’ (i.e.,

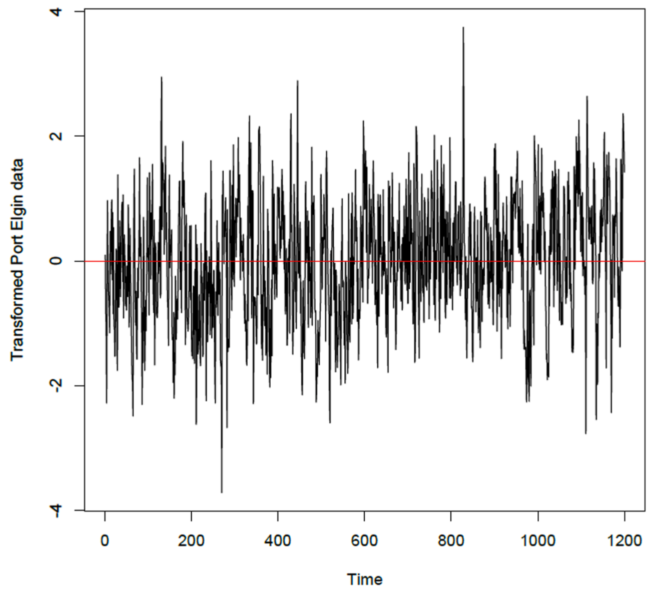

Fig. 1. Transformed Port Elgin study site data. Red: change-point fit using the technique of Cho and Fryzlewicz (2021). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

the gSC component, which estimates the dependence structure as well as the number of change points simultaneously) (Cho and Fryzlewicz, 2021). We ran the WEM.gSC technique from Cho and Fryzlewicz (2021), appropriate for detecting changes in the mean in the presence of serial correlations in time series data, which did not detect any change-points in the transformed data (Fig. 1). In other words, this means that our transformed data is stationary, and has no trend or jump.

We want to highlight that, in their discussion paper, EZB used the KPSS (Kwiatkowski-Phillips-Schmidt-Shin) test (to test for stationarity), but this does not appear appropriate in this setting as this tests against the presence of a deterministic trend in the direction of a unit-root alternative, while our Port Elgin data do not display either feature. In addition, EZB used the Mann-Kendall test (to test for trend) and the Mann-Whitney test (to test for jump), but this also does not appear appropriate in this instance, as both tests are designed for uncorrelated/independent time series, while our data (both original and transformed) exhibit clear serial correlation. Below, we show a brief example, in the R language for statistical computing, of how the Mann-Kendall test can be 'fooled' by the presence of serial correlation. The Kendall R package (McLeod, 2011) is adopted.

```
> install.packages("Kendall", repos='http://cran.us.r-project.org')
> library(Kendall)
> set.seed(1)
> MannKendall(arima.sim(list(ar = 0.9), n = 1000))
tau = -0.0647, 2-sided pvalue = 0.0021775
```

This performs the Mann-Kendall test on a realization of a stationary AR(1) process and (incorrectly) strongly rejects stationarity. In the presence of serial correlation, other approaches such as WEM.gSC (Cho and Fryzlewicz, 2021) should be used.

We also present the Autocorrelation Function (ACF) structure of our original and transformed data for the Port Elgin study site in Fig. 2. The ACF behavior of the transformed data shows that using our data preprocessing approach (which included de-seasonalization as we subtracted the means of each month separately), the seasonality is removed from the original data, and no statistically significant seasonal autocorrelation exists in the transformed data.

Based on all of the above points, to re-iterate what we mentioned earlier and in contrast to what EZB state, our transformed data for the Port Elgin study site is stationary, and has no trend or jump, and our use of AR and MA time series models was suitable for our individual linear time series models (as well as to form our hybrid linear-ARCH models).

On a different note, regarding EZB's comment on the Port Elgin test phase results for the individual linear time series models (i.e., AR and MA), upon re-examining the model results in our original study, we discovered that we had made an error in reporting the performance index values (i.e., RMSE, MAE and R values) for a sub-set of individual linear time series and hybrid linear-ARCH models for the Port Elgin (as well as Walkerton) study sites. For this sub-set of models, the fitted models themselves and their orders are correct; however, we had made a small error when re-scaling the data to the original scale prior to calculating the performance indices. Hence the error in reporting the performance index values for this sub-set of models. The corrected results for this sub-set of models is as follows: i) Port Elgin RMSE (m^3/s), MAE (m^3/s), R [AR(1) test phase = 37.65, 23.86, 0.709; AR(1)-ARCH(1) train phase = 10.74, 4.95, 0.99; MA(4) test phase = 37.62, 23.87, 0.715; MA(4)-ARCH(1) train phase = 10.15, 4.76, 0.991]; ii) Walkerton RMSE (m^3/s), MAE (m^3/s), R [AR(1) test phase = 19.24, 12.35, 0.704; AR(1)-ARCH(1) train phase = 5.89, 2.73, 0.99]. All other results - for all other models and all other study sites - in our original paper are correct and are not affected, and our overall conclusions in the original study are not affected and remain the same.

On a final note, EZB comment on our individual linear AR and MA models for the Beinerahe Roodbar study site, and state that "it is very rare to find practical usages of AR(n) and MA(n) where n is larger than two". We respectfully disagree. In contrast to what EZB state, we want to point out that numerous studies in the streamflow modeling literature have detailed the development and use of AR and MA models, including for monthly data, where $n > 2$ (e.g., Salas et al., 1988; Wang and Salas, 1991; Kişi, 2004; Wang et al., 2005; Myronidis et al., 2018; etc.). Regarding EZB's suggestion on the use of seasonal ARMA (SARMA) models for the Beinerahe Roodbar study site instead of our individual AR and MA models, we want to point out that seasonal AR models (SAR), a sub-class of SARMA/SARIMA, have an AR representation where the AR lag has to be at least as long as the seasonal period, so, for example, for modeling monthly data with a yearly season, it would be reasonable to use a seasonal AR model for which the AR representation is AR(12). Having said this, regarding our individual AR and MA models for the Beinerahe Roodbar study site, we emphasize again that our data preprocessing/transformation in the original study included de-seasonalization (as we subtract the means of each month separately), which is why our transformed data for all study sites in the original study no longer displayed seasonality in the ACF, and so we respectfully disagree with EZB's point regarding the possible use of SARMA models when applied to our transformed data for the Beinerahe Roodbar study site. We were satisfied with our AR and MA models for the Beinerahe Roodbar study site given the goodness-of-fit of our selected models (i.e., the residuals had a white noise structure and practically zero autocorrelation) in conjunction with the fact that our use of AR and MA models adhered to our aim of focusing on the simplest possible model classes in linear time series analysis for our individual linear time series models (and to form our hybrid linear-ARCH models). Take our AR model for the Beinerahe Roodbar study site as an example; in our original research, as with each of our models, we had assessed the goodness-of-fit of this model (fitted to the transformed Beinerahe Roodbar data) by testing the residuals of the AR model.

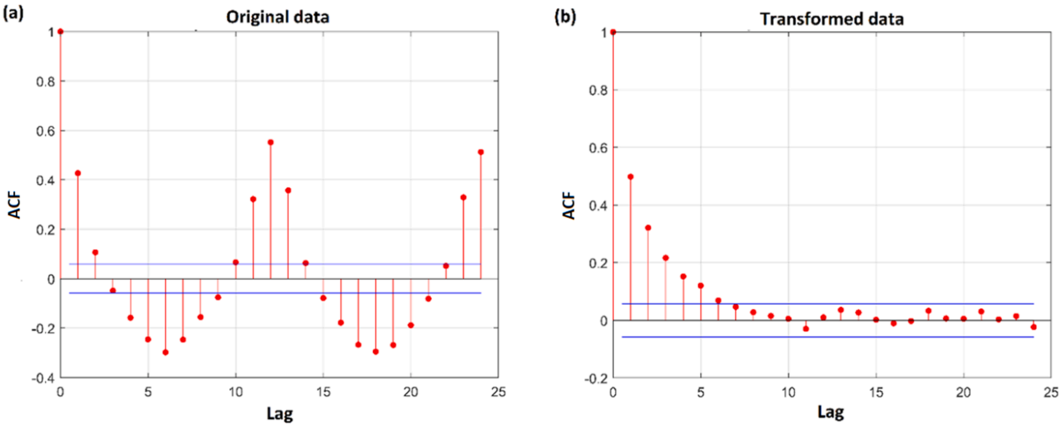

Fig. 2. ACF structure of (a) original data and (b) transformed data (Port Elgin study site).

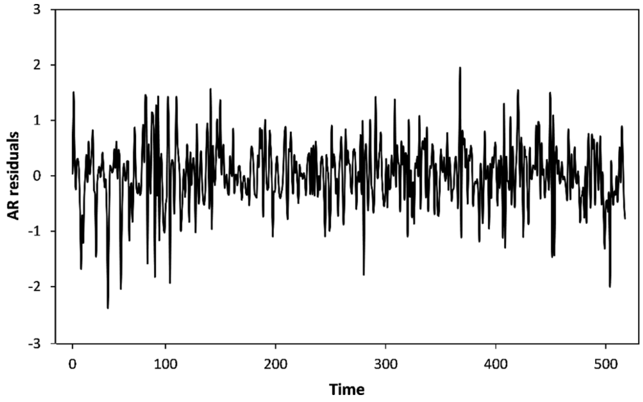

Fig. 3. Residuals from the AR model fitted to the transformed Beinerahe Roodbar time series.

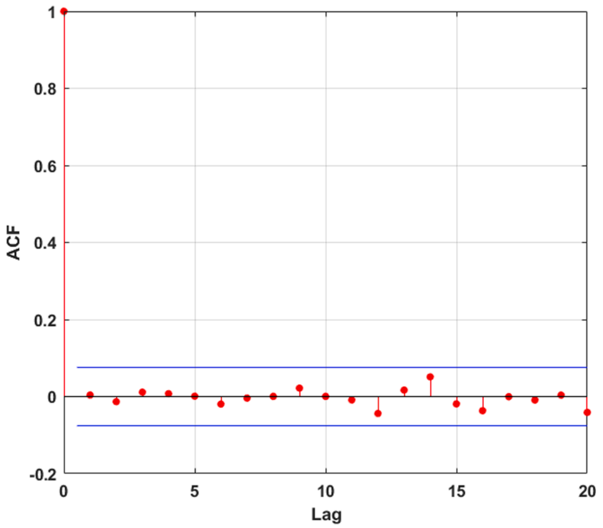

Fig. 4. ACF of the residuals of the AR model fitted to the transformed Beinerahe Roodbar time series.

The residuals from this model are shown in Fig. 3. Visually, they have a white noise structure, which is further confirmed in their ACF plot, shown in Fig. 4. They do also display ARCH effects in the squares (not shown), which, as already mentioned earlier, explains our ARCH modeling of the residuals (in the hybrid linear-ARCH, i.e., AR-ARCH, model). The most important point is that both figures provide strong visual evidence for the goodness-of-fit of the AR model: the residuals from this fit have a stationary white noise structure - which is what should be expected from residuals from an adequate time series fit.

3. Point 2: Artificial intelligence models

Before we respond to the comments of EZB in point 2, we want to mention that, for ease of reference in our original paper, we referred to and grouped the MARS and GEP methods as artificial intelligence models (in the sense that we are referring to a general class of models that encompass non-parametric regression methods (such as MARS) and evolutionary methods (such as GEP)).

In point 2, EZB first comment that our data preprocessing procedures are not necessary to develop our artificial intelligence (AI) models. We respectfully disagree. First, we want to mention that to conduct a fair performance comparison of the individual linear time series and hybrid linear-ARCH models with the artificial intelligence models (i.e., GEP and MARS) in our original study, we believe it was important to use the same preprocessed/transformed data to develop the individual linear time series, hybrid linear-ARCH, and artificial intelligence models to allow us to maintain the same conditions when conducting our performance comparison of models. Since preprocessed/transformed data were used to develop the individual linear time series and hybrid linear-ARCH models in our original study (for the reasons we discussed in point 1 of this closure), we felt that the same preprocessed/transformed data should be (and was) used to develop our GEP and MARS models to maintain the same conditions. Second, and on a more general note, we believe that data preprocessing procedures (of which there are many different types ranging from simple to complex approaches that go well beyond our approach in the original study), can often be useful to use before developing an artificial intelligence model, and are common (e.g., Zhang and Qi, 2005; Wu et al., 2009; Zhang et al., 2015; Kalteh, 2016; Quilty et al., 2019). We believe that the many different data preprocessing approaches that are available can be useful to consider before developing artificial intelligence models for many reasons, including: 1. Artificial intelligence models can sometimes have challenges to simultaneously deal with several components of the data appropriately, and therefore data preprocessing can be useful (Quilty et al., 2019; Zuo et al., 2020; Hammad et al., 2021). 2. Preprocessed data can help facilitate not only optimal model structure building processes, but also alleviate potential overfitting problems of artificial intelligence models, and can help improve the performance and stability of the model, as well as speed up the training process and convergence during training (Nelson et al., 1999; Zhang and Qi, 2005; Fijani et al., 2019; Zuo et al., 2020). 3. Preprocessing of data can sometimes have a significant effect on the performance of artificial intelligence models, and can help to deal with the impacts of, for example, noisy and unreliable data, errors, and outliers (Kotsiantis et al., b, 2006a; Garcia et al., 2015; Chitsaz et al., 2016; Fijani et al., 2019; Hammad et al., 2021). The above are just some examples of reasons why data preprocessing, of which there are many different types, is often considered to be useful before developing artificial intelligence models.

EZB then comment that for the local analysis approach in the Port Elgin study site (which is the only study site and type of analysis EZB provide GEP and MARS model results for in point 2), using the original/raw data, rather than the preprocessed/transformed data, results in more accurate MARS and GEP models in the Port Elgin study site. We want to first mention an important point: In their discussion paper, EZB used *nine inputs* to develop their one MARS model, and one GEP model, which they developed using the original/raw data for the Port Elgin study site. In contrast, our MARS and GEP models in the original study, which were developed using our preprocessed/transformed data, used *less than half the number of inputs* (i.e., *one* to *four* inputs, with our best MARS model using *four* inputs, and our best GEP model using *three* inputs). We now want to mention a second important point: if we compare our best MARS model (developed using our preprocessed data) for the Port Elgin study site in our original study (i.e., MARS4 which has *four* inputs and has, for the test phase, R = 0.731, RMSE = 36.69 m^3/s, MAE = 23.33 m^3/s) with the one MARS model developed by EZB in their discussion paper (developed using the original/raw data) for the Port Elgin study site (which has *nine* inputs and has, for the test phase, R = 0.714, RMSE = 36.38 m^3/s, and no MAE result reported), our R is actually 2.3% *better*, while our RMSE is 0.85% worse, compared to the model of EZB. So, comparing our best MARS model with the one MARS model developed by EZB, shows that our R was actually *better*, and our RMSE *marginally worse* (less than 1%) compared to the model developed by EZB. These are very marginal differences and, importantly, our model used less than half the number of inputs. Following the important principle of parsimony in modeling, we believe that this shows that our approach in our original study was suitable in that our model provided good results with few inputs (e.g., basically the same results as the model of EZB but with less than half the model inputs), and, importantly, our approach for the MARS (and GEP) model used the same (preprocessed) data that was used to develop our other models in the original study, which we believe was important to allow us to maintain the same conditions when conducting our performance comparison of models. For the one GEP model that EZB developed that had a better performance compared to our best GEP model, we believe this was likely due, in part, to EZB using three times the number of inputs for their one GEP model (nine inputs for their GEP model, versus three inputs for our best GEP model). This implication is also clear in the results of our original study, where our models with more inputs generally had better results (e.g., our best GEP model (GEP3 with three inputs) had better results than our GEP1 (with one input), or our best MARS model (MARS4 with four inputs) had better results than our MARS1 (with one input)), which is to be expected. The decision of selecting what modeling approach to pursue depends on, among other things, the aims of the modeller. And, we want to re-iterate that, in our original study, our overall aim and approach for our models (i.e., individual linear time series, hybrid linear-ARCH, and artificial intelligence) was to develop simple (and adequate) models, and also use the same data set (i.e., preprocessed/transformed in our case) to develop all the models, since we believe using the same data set was important in order to conduct a fair performance comparison of models. As such, our GEP and MARS models with few inputs but good results, which were developed using the same preprocessed data as the other models in our original study, adhered to our aims. Based on what we have discussed, and keeping in mind our aims and the important principle of parsimony in modeling, the above described approach that we selected in our original study to develop our MARS and GEP models was suitable.

On a final note, EZB comment on the GEP parameter values (e.g., number of chromosomes, mutation rate, etc.) that we used for our GEP models, and then provide what they state are the optimum GEP parameter values, which they found through a trial and error process. We want to point out that the use of our GEP parameter values in the original study was suitable; many other hydrological and environmental modeling studies have used the same, or a very similar, approach to ours with respect to the GEP parameter values (e.g., Kisi and Shiri, 2012; Hashmi and Shamseldin, 2014; Zorn and Shamseldin, 2015; Kisi et al., 2015; Al-Juboori and Guven, 2016; Samadianfard et al., 2018; Bateni et al., 2019; Wang et al., 2019). On a related note, we want to point out that the GEP parameter values provided by EZB were found through a trial and error process; a trial and error approach contains variability, especially one involving optimizing eleven different parameters as is the case for the GEP model, and different researchers undertaking a trial and error approach may each find a different set of ‘optimal’ model parameter values. We also want to point out that EZB did not provide any results of a GEP model that used their optimal GEP parameter values, to assess if these GEP model results differ very significantly from our GEP model results, using the same preprocessed data and the same set of model inputs as we used in the original paper. However, we believe it is quite likely that the different GEP model parameter values that EZB suggest would not result in very significant/important differences in GEP model performance (e.g., as was the case with the earlier described marginal differences in model results between using raw/original data and processed data for the MARS model). Having said all of this, we want to emphasize again the most important point, which is that the use of our GEP parameter values in the original study was suitable; as mentioned earlier, many other hydrological and environmental modeling studies have used the same, or a very similar, approach to ours with respect to the GEP parameter values.

4. Point 3: Hybrid models

EZB first comment that “no results are given for ARCH”, and that “using ARCH may provide results close to the hybrid models (AR-ARCH and MA-ARCH)”. EZB are stating that we should have compared the ARCH models ‘alone’ with our other models (e.g., AR-ARCH and MA-ARCH). We respectfully disagree. We want to point out that it did not make sense to compare the performance of ARCH models ‘alone’ with our other models (e.g., AR-ARCH, MA-ARCH) since the ARCH models in our study were fitted to the residuals of the linear AR and MA models (to remove the ARCH effects in the residuals of the linear AR and MA models, and build the hybrid AR-ARCH and MA-ARCH models).

EZB also comment that “combining stochastic methods with GEP can also yield good results”. We were aware that such hybrid models can be developed and may also provide good results. Coupling stochastic or time series models with artificial intelligence models for streamflow modeling has already been investigated in other studies (e.g., Moeeni et al., 2017), including by the authors of the original paper (e.g., Mehdizadeh et al., 2019b; Fathian et al., 2019c). However, developing these types of hybrid time series - artificial intelligence models (and other possible types of hybrid models) was outside the scope of our original study; our aim was to develop hybrid linear-ARCH models.

5. Point 4: External analysis

EZB state that in our external analysis approach "no value for Q(t) at station 2 is recorded that can serve as a model input to predict Q(t) at station 1", and comment that our external analysis approach is not valid. We respectfully disagree; EZB misinterpreted what we did in our external analysis approach. In our valid external analysis approach, we developed a procedure to estimate missing streamflow data at a target station (station 1) using streamflow data from a neighboring station (station 2), where we assumed that Q(t) (as well as Q(t-1), etc.) is observed/available at the neighboring station, but where Q(t) is missing at the target station for any reason. More specifically, in our external analysis approach in the original study, we had developed artificial intelligence models that could be used in our study sites to estimate missing monthly streamflow data at a target station (Q(t)), using streamflow data that is observed/available for the same time (i.e., Q(t)) as well as for the past (i.e., Q(t-1), etc.) from a neighboring station in the same river (geographically near and with temporally similar characteristics), when the target station streamflow data (Q(t)) is not available (which can be a common issue in practice, e.g., if the target station has missing data due to a malfunction in the streamflow gauge). In the original paper, we state that “under the external analysis approach, the data of a neighboring station is used to estimate streamflow at each target station” and, for example, that “for the MARS1 and GEP1 models under the external analysis approach, the streamflow data of the same month (Q(t)) at a nearby station were used” to estimate the streamflow (Q(t)) at the target station (in the same river). We want to point out that a similar external analysis approach was successfully used by Sanikhani and Kisi (2012) in the context of estimating missing monthly streamflow data at a target station. We obtained accurate results using our above described external analysis approach in the original study, and this valid and suitable approach was useful to have explored.

6. Point 5: RMSE, MAE and R

Regarding EZB’s comments on our use of the AIC, we had used the AIC, as a suitable criterion (e.g., Salas et al., 1988; Hipel and McLeod, 1994), to help select the optimum model (i.e., best fitted model with optimum order among satisfactory candidates) for each type/class of individual linear time series model (i.e., AR, MA) and each type/class of hybrid linear-ARCH model (i.e., AR-ARCH, MA-ARCH) that we used (and then compared based on their performance as explained below). However, we did not show the details of the selection process based on AIC results since the original paper was already overly long. After using the AIC to select the optimum model for each type/class of time series model (i.e., AR, MA, AR-ARCH, MA-ARCH), we compared these selected models based on their performance (via the RMSE, MAE and R performance indices). We respectfully disagree with EZB's comment that our approach (described above) to compare our selected individual linear time series models and hybrid linear-ARCH models based on their performance (via, in our case, the RMSE, MAE and R performance indices) is not appropriate, and that we must instead use the AIC to compare these selected time series models. Our approach, which is valid and suitable, was to compare our selected individual linear time series models and hybrid linear-ARCH models based on their performance, and we used the RMSE, MAE and R performance indices, as a suitable approach, to do this.

7. Conclusion

In this closure, we responded to the comments of EZB, and further explained the aims and reasoning behind our approaches and models, and showed why our approaches and models are suitable based on the aims and scope of our study. We also included additional details regarding several features of our study (that we did not provide in the original paper since it was already overly lengthy). To summarize the main points we discussed in this closure:

1. Individual linear time series and hybrid linear-ARCH models: We explained in more detail our aims, scope, reasoning and approach used to develop our individual linear time series and hybrid linear-ARCH models, and explained why our models were suitable given our aims.
2. Artificial intelligence models: We explained why data preprocessing can be useful in artificial intelligence modeling and why it was used in our study.
3. Hybrid models: We explained that comparing the performance of ARCH models 'alone' with our other models, as suggested by EZB, did not make sense since our ARCH models were fitted to the residuals of the linear time series models (to remove the ARCH effects in the residuals and build the hybrid linear-ARCH models). We also pointed out that developing hybrid time series - artificial intelligence models (and other possible types of hybrid models), as suggested by EZB, was outside the scope of our study, which aimed to develop hybrid linear-ARCH models.
4. External analysis: We explained why our external analysis approach was valid and useful.
5. RMSE, MAE and R: We explained why our performance indices (i.e., RMSE, MAE and R) were suitable to compare our models.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We want to thank Prof. Geoff Syme (Editor, Journal of Hydrology). The COVID-19 pandemic had significant and very negative impacts on several authors of this closure, and we appreciate the help and guidance of Prof. Syme during this very challenging year. We also want to thank Prof. Piotr Fryzlewicz (London School of Economics), Mr. Mohammad Reza Alizadeh (McGill University), Prof. Reza Modarres (Isfahan University of Technology), Prof. John Quilty (University of Waterloo), and Prof. Babak Vaheddoost (Bursa Technical University) for very useful discussions.

References

Al-Juboori, A. Mahmood, Guven, A., 2016. A stepwise model to predict monthly streamflow. J. Hydrol. 543, 283–292.

Bateni, S., Vosoughifar, H., Truce, B., Jeng, D., 2019. Estimation of clear-water local scour at pile groups using genetic expression programming and multivariate adaptive regression splines. J. Waterway Port Coastal Ocean Eng. 145 (1), 04018029.

Chitsaz, N., Azarnivand, A., Araghinejad, S., 2016. Pre-processing of data-driven river flow forecasting models by singular value decomposition (SVD) technique. Hydrol. Sciences J. 61 (12), 2164–2178.

Cho, H., Fryzlewicz, P., 2021. Multiple change point detection under serial dependence: Wild energy maximisation and gappy Schwarz criterion. eprint arXiv:2011.13884 [stat.ME].

Ebtehaj, I., Zeynoddin, M., Bonakdari, H., 2020. Discussion of "Comparative assessment of time series and artificial intelligence models to estimate monthly streamflow: A local and external data analysis approach" by Saeid Mehdizadeh, Farshad Fathian, Mir Jafar Sadegh Safari and Jan F Adamowski. J. Hydrol. 583, 124614. https://doi.org/10.1016/j.jhydrol.2020.124614.

Fathian, F., Modarres, R., Dehghan, Z., 2016. Urmia Lake water-level change detection and modeling. Mod. Earth Syst. Env. 2 (4), 1–16.

Fathian, F., Fakheri-Fard, A., Ouarda, T.B.M.J., Dinpashoh, Y., Mousavi Nadoushani, S. S., 2019a. Multiple streamflow time series modeling using VAR–MGARCH approach. Stoch. Environ. Res. Risk Assess. 33 (2), 407–425.

Fathian, F., Fakheri Fard, A., Ouarda, T.B.M.J., Dinpashoh, Y., Mousavi Nadoushani, S.S., 2019b. Modeling streamflow time series using nonlinear SETAR-GARCH models. J. Hydrol. 573, 82–97.

Fathian, F., Mehdizadeh, S., Kozekalani Sales, A., Safari, M.J.S., 2019c. Hybrid models to improve the monthly river flow prediction: Integrating artificial intelligence and non-linear time series models. J. Hydrol. 575, 1200–1213.

Fijani, E., Barzegar, R., Deo, R., Tziritis, E., Skordas, K., 2019. Design and implementation of a hybrid model based on two-layer decomposition method coupled with extreme learning machines to support real-time environmental monitoring of water quality parameters. Sci. Total Env. 648, 839–853.

Garcia, L.P.F., de Carvalho, A.C.P.L.F., Lorena, A.C., 2015. Effect of label noise in the complexity of classification problems. Neurocomputing 160, 108–119.

Hammad, M., Shoaib, M., Salahudin, H., Baig, M., Khan, M., Ullah, M., 2021. Rainfall forecasting in upper Indus basin using various artificial intelligence techniques. Stoch. Environ. Res. Risk Assess. doi.org/10.1007/s00477-021-02013-0.

Hashmi, M.Z., Shamseldin, A.Y., 2014. Use of gene expression programming in regionalization of flow duration curve. Adv. Water Resourc. 68, 1–12.

Hipel, K.W., McLeod, A.I., 1994. Time series modelling of water resources and environmental systems. Developments in Water Science, 1st edition. Elsevier, Amsterdam, Netherlands.

Kalteh, A.M., 2016. Improving forecasting accuracy of streamflow time series using least squares support vector machine coupled with data-preprocessing techniques. Water Resour. Manag. 30 (2), 747–766.

Kişi, Ö., 2004. River flow modeling using artificial neural networks. J. Hydrol. Eng. 9 (1), 60–63.

Kisi, O., Sanikhani, H., Zounemat-Kermani, M., Niazi, F., 2015. Long-term monthly evapotranspiration modeling by several data-driven methods without climatic data. Comput. Electron. Agric. 115, 66–77.

Kisi, O., Shiri, J., 2012. River suspended sediment estimation by climatic variables implication: comparative study among soft computing techniques. Comp. Geosc. 43, 73–82.

Kotsiantis, S.B., Kanellopoulos, D., Pintelas, P.E., 2006a. Data preprocessing for supervised learning. Internat. J. Comput. Inform. Eng. 1 (2), 111–117.

Kotsiantis, S.B., Zaharakis, I.D., Pintelas, P.E., 2006b. Machine learning: a review of classification and combining techniques. Artif. Intell. Rev. 26 (3), 159–190.

McLeod, A.I., 2011. Kendall: Kendall rank correlation and Mann-Kendall trend test. R package version 2, 2. https://CRAN.R-project.org/package=Kendall.

Mehdizadeh, S., Fathian, F., Safari, M.J.S., Adamowski, J.F., 2019a. Comparative assessment of time series and artificial intelligence models to estimate monthly streamflow: a local and external data analysis approach. J. Hydrol. 579, 124225. https://doi.org/10.1016/j.jhydrol.2019.124225.

Mehdizadeh, S., Fathian, F., Adamowski, J.F., 2019b. Hybrid artificial intelligence-time series models for monthly streamflow modeling. App. Soft Compt. 80, 873–887.

Mehdizadeh, S., Fathian, F., Safari, M.J.S., Khosravi, A., 2020. Developing novel hybrid models for estimation of daily soil temperature at various depths. Soil Till. Res. 197, 104513. https://doi.org/10.1016/j.still.2019.104513.

Moeeni, H., Bonakdari, H., Ebtehaj, I., 2017. Integrated SARIMA with Neuro-fuzzy systems and neural networks for monthly inflow prediction. Water Resour. Manag. 31 (7), 2141–2156.

Myronidis, D., Ioannou, K., Fotakis, D., Dörflinger, G., 2018. Streamflow and hydrological drought trend analysis and forecasting in Cyprus. Water Resour. Manage. 32 (5), 1759–1776.

Nelson, M., Hill, T., Remus, W., O'Connor, M., 1999. Time series forecasting using neural networks: should the data be deseasonalized first? J. Forecast. 18 (5), 359–367.

Quilty, J., Adamowski, J., Boucher, M.-A., 2019. A stochastic data-driven ensemble forecasting framework for water resources: a case study using ensemble members derived from a database of deterministic wavelet-based models. Water Resour. Res. 55 (1), 175–202.

Rajagopalan, B., Salas, J. Lall, U., 2010. Stochastic methods for modeling precipitation and streamflow. In: Advances in data-based approaches for hydrologic modelling and forecasting. Sivakumar, B., Berndtsson, R. (Ed.), World Scientific Publishing. Hackensack, NJ.

Salas, J.D., Delleur, J.W., Yevjevich, V., Lane, W.L., 1988. Applied modeling of hydrologic time series, 3rd edition. Water Resources Publications, Littleton CO.

Samadianfard, S., Asadi, E., Jarhan, S., Kazemi, H., Kheshtgar, S., Kisi, O., Sajjadi, S., Manaf, A.A., 2018. Wavelet neural networks and gene expression programming models to predict short-term soil temperature at different depths. Soil Till. Res. 175, 37–50.

Sanikhani, H., Kisi, O., 2012. River flow estimation and forecasting by using two different adaptive neuro-fuzzy approaches. Water Resour. Manag. 26 (6), 1715–1729.

Wang, D.C., Salas, J.D., 1991. Forecasting streamflow for Colorado River systems. Colorado Water Resources Research Institute Report No. 164. Colorado State University. Fort Collins, CO.

Wang, S., Lian, J., Peng, Y., Hu, B., Chen, H., 2019. Generalized reference evapotranspiration models with limited climatic data based on random forest and gene expression programming in Guangxi China. Agric. Water Manage. 221, 220–230.

Wang, W., Van Gelder, P.H.A.J.M., Vrijling, J.K., Ma, J., 2005. Testing and modelling autoregressive conditional heteroscedasticity of streamflow processes. Nonlin. Process. Geophys. 12, 55–66.

Wu, C.L., Chau, K.W., Li, Y.S., 2009. Predicting monthly streamflow using data-driven models coupled with data-preprocessing techniques. Water Resour. Res. 45 (8), W08432.

Zhang, G.P., Qi, M., 2005. Neural network forecasting for seasonal and trend time series. Eur. J. Oper. Res. 160 (2), 501–514.

Zhang, X., Peng, Y., Zhang, C., Wang, B., 2015. Are hybrid models integrated with data preprocessing techniques suitable for monthly streamflow forecasting? Some experiment evidences. J. Hydrol. 530, 137–152.

Zorn, C.R., Shamseldin, A.Y., 2015. Peak flood estimation using gene expression programming. J. Hydrol. 531 (3), 1122–1128.

Zuo, G., Luo, J., Wang, N.i., Lian, Y., He, X., 2020. Two-stage variational mode decomposition and support vector regression for streamflow forecasting. Hydrol. Earth Syst. Sci. 24 (11), 5491–5518.