Reply to John Quilty

Dear John Quilty,

We thank you very much for taking the time to review our manuscript, for your valuable comments/suggestions and the points of discussion you raised. We have wisely revised our manuscript according to your suggestions. Please find our detailed responses below:

Comment from John Quilty — ASSESSMENT:

This paper introduces a variational mode decomposition (VMD)-based support vector regression (SVR), i.e., VMD-SVR, model for multi-step ahead streamflow forecasting. The authors strive to address the 'boundary effects' problem common to many time series decomposition approaches that are typically coupled with data-driven models such as VMD, ensemble empirical mode decomposition (EEMD), singular spectrum analysis (SSA), wavelet transforms (WT), etc. outlined in earlier studies (Du et al., 2017; Maheswaran and Khosa, 2012; Quilty and Adamowski, 2018; Wang and Wu, 2016; Zhang et al., 2015). This is a worthwhile problem to address due to the growing interest in coupling these decomposition methods (VMD, EMD, SSA, WT, etc.) with data-driven models for hydrological forecasting and the vast majority of studies that overlook the impact of boundary effects on hydrological forecasting performance. Many of the just mentioned studies point out flaws in existing strategies for coupling decomposition methods with data-driven models and some go on to identify potential solutions.

In this paper, the authors put forth their own approach for addressing boundary effects. The authors claim that the main benefits of their proposed approach include that it "...can reduce the boundary effects, save the modelling time, and improve the prediction performance. This practical streamflow forecasting framework can be outlined as follows:

(1) Divide the entire streamflow data into training and validation sets and decompose each of these two sets separately into signal components. This procedure avoids using the validation information for training purposes.

(2) Combine the predictors of individual signal components into a final predictors, and select the original streamflow data as the prediction target in order to build only one optimized prediction model.

(3) Generate training and validation samples and divide the validation samples into development and testing samples. Mix and shuffle the training and the development samples to optimize the prediction model, and reduce the boundary effects."

Throughout the MAJOR COMMENTS section below, I raise several issues with how the authors' proposed approach actually satisfies these points. In my opinion, I think there is much clarification required on the authors' part to demonstrate that they adequately fulfill these points (in a way that is meaningful for operational forecasting problems, which the present study appears to be concerned with). In particular, the authors' methodology for how they decompose the time series using VMD (and other comparative approaches) and use it in training and validating their proposed VMD-SVR (and comparative) method(s) is not entirely clear. Out of all issues raised in this review, this point needs the most attention.

Nonetheless, I find that the paper is well-written, is properly structured, and is supported by appropriate figures. The references are sufficient. The analysis carried out in the paper is reasonable but the validity of the paper's results, in terms of how useful the results are to those concerned with developing operational forecasting models, largely depends on how the authors carried out decomposition of the time series (using VMD, DWT, etc.) and used it to develop the forecasting models.

If the authors can adequately address each of the comments/suggestions mentioned below, I would be happy to re-evaluate my current stance on the paper's suitability for being published in Hydrology and Earth System Sciences. In my opinion, the paper should not be published in its current form.

Reply: We thank you for the positive evaluation of our work and all the comments/suggestions on our manuscript. We have made improvements to the manuscript following your suggestions. Overall, the main corrections are:

1. We have analyzed the VMD shift-variance and sensitivity of the addition of new data using the underlying data of our study (the monthly runoff at Huaxian station) and discussed how to reduce the influence of boundary effects caused by the VMD shift-variance and sensitivity of addition of new data.

2. We have compared the proposed TSDP framework with the WDDFF framework proposed by Quilty and Adamowski (2018). Additionally, we have compared non-decomposition-based SVR with the ARIMA model in the revised manuscript.

3. We have moved the "Section 3.1 Study area and data observations" of the original manuscript to "Section 2 Monthly runoff data" to make it easier for the readers to start reading the methodology, knowing where the research was developed. Additionally, the "Experiment 2 Performance evaluation of the TSDP models"and "Experiment 5 Evaluation of runoff forecasting for long leading times" in Section 2.5 of the previous manuscript were combined to make the manuscript to be concise.

Please find more details about our changes below.

Comment from John Quilty — MAJOR ITEMS:

In the Introduction, the authors mention how their "... proposed scheme can reduce the boundary effects, save the modelling time, and improve the prediction performance. This practical streamflow forecasting framework can be outlined as follows:

(1) Divide the entire streamflow data into training and validation sets and decompose each of these two sets separately into signal components. This procedure avoids using the validation information for training purposes.

(2) Combine the predictors of individual signal components into a final predictors, and select the original streamflow data as the prediction target in order to build only one optimized prediction model.

(3) Generate training and validation samples and divide the validation samples into development and testing samples. Mix and shuffle the training and the development samples to optimize the prediction model, and reduce the boundary effects."

Some comments on each of these points are given below:

Point 1

To decompose the validation data, the authors imply that they append one validation record at a time onto the calibration dataset (and any previous validation data) then perform VMD. The authors do this for each validation record, keeping the VMD components for each previous validation record static (this I my interpretation, the latter assertion was not specifically mentioned by the authors, at least that I could find). This appears to avoid boundary effects in the validation data due to the 'future data' issue; however, there are two major issues with this approach:

• The first issue is that VMD is sensitive to the addition of new data. I.e., by adding an additional data point to a time series and performing VMD creates inconsistencies between the intrinsic mode functions (IMFs) with the appended data and the IMFs prior to the appended data. Sometimes, these inconsistencies can be very large and tend to be largest at the edges of a time series, the most important time series observations in real-world forecasting applications (see example R script attached at the end of this review for the sensitivity of VMD to the additional of new data points).

• Because of this last point, the parameters of a model calibrated on initial IMFs generated by VMD that are then fed with updated IMFs based on the newly appended data may need to be updated to account for these newly introduced errors not seen during model calibration. This begs the question of whether each time the IMFs are updated whether the model should be updated too. Which goes against the authors' desire to implement a computationally-efficient forecasting method. Perhaps the authors may wish to consider a Kalman Filter to update their model parameters if they follow such an approach (as opposed to completely re-training the model). The Kalman Filter could be used to update the model parameters at each time step or at larger intervals.

• The second issue is that VMD is shift-variant, meaning that performing VMD on lagged versions of the same time series leads to distortions in the IMFs derived by the VMD at the same times. This further exacerbates the issue raised above in terms of calibrating and validating a data-driven model based on using time-lagged inputs that are decomposed via VMD (see example R script attached at the end of this review that demonstrates the shift-variance problem in VMD).

Reply: Thank you for providing the R script to test the VMD shift-variance and sensitive to the addition of new data. We have tested the VMD shift-variance and sensitive to the addition of new data using a Matlab implementation. The secondary penalty parameter (α), the noise tolerance (τ), and the convergence tolerance (ϵ) remain static for decomposing the training set and appended sets. The decomposition level was tuned based on the training set and remain static for decomposing the appended sets. The last decomposition of appended sets for each signal component is a validation decomposition. We have clarified how to set and use decomposition parameters in our revised manuscript.

We agree with you that VMD is shift-variant and sensitive to the addition of new data. We think the boundary effects cause these two issues. The boundary effects lead to large decomposition errors at the edges of a time series, but the rest decomposition errors are very small by adopting appropriate decomposition parameters, which can be proved by Fig. 1. Fig. 1(b), (d), and (f) show that VMD has very small decomposition errors except for the boundary decompositions. Since the training set was concurrently decomposed and the validation set was sequentially appended to the training set and decomposed, the training decompositions (samples) are barely affected while the entire validation decompositions (samples) are affected by the introduced decomposition errors. In other words, the training and validation samples have different error distribution, which leads to the models calibrated on the training samples generalized poorly to validation samples.

Since the training and validation samples can come from different distribution (see Machine learning yearning by Andrew Ng), we think it is not necessary to update model parameters with Kalman Filter because we have already dealt with the boundary effects with two different approaches. However, we do believe that updating model parameters with Kalman Filter is a new great idea to deal with boundary effects and we will research it in our subsequent experiments. One approach is to make the driving pattern (the relationship between predictors and prediction targets) in the validation sample as close as possible to that of training sample and the other approach is to make the models assess validation error distribution during the training stage. In the first approach, the validation samples were generated from the decompositions of appended sets rather than validation decompositions because the driving pattern of boundary decomposition of appended sets is close to training decompositions (monotone increasing or decreasing, see minimap in Fig. 1a, c and e). In other words, the predictors selected from appended decomposition are more correlated to the predicted target (see Fig. 2). In the second approach, we mixed and shuffled the training and development (half validation) samples to build SVR models based on cross-validation because the model calibrated on training and validation error distribution simultaneously. We have proved that these two approaches worked by Experiment 1. The key ideas, potential reasons, and procedures of the TSDP framework have been clarified in the revised manuscript.

Comment from John Quilty — MAJOR ITEMS:

Point 2

Although there are numerous studies that have considered forecasting each IMF (in EEMD, VMD, etc. based models) separately (and summing their constituent forecasts to obtain the final forecast), the authors should note that in the literature other studies have also built forecasts using, for instance, all wavelet-decomposed time series in a single forecasting model (Maheswaran and Khosa, 2013; Quilty and Adamowski, 2018). Many other examples of this approach can be found in the literature. It is suggested that the authors 'downplay' this feature of their framework as being something new or different.

Reply: Yes, you are right and we have downplayed "building a single forecasting model" as the new feature of the proposed TSDP framework and we also have added references for building a single forecasting model.

Comment from John Quilty — MAJOR ITEMS:

Point 3

In terms of the mix and shuffle approach used to the training and validation data in the VMD-SVR models:

It is very difficult (for me) to see how taking all but the last 120 records of the red line in Figure 8 (b) (i.e., the development set) and randomly shuffling it with the red line from Figure 8 (a), (the training set) would lead to such a high performance on the last 120 records in Figure 8 (b) (i.e., the test set) as noted in Figure 9 and 11. Especially, when it appears that the training and combined development and test sets have completely different distributions (with the training set having a larger number of records).

I think it is necessary for the authors (at the very least) to provide pseudo-code for how they decomposed the time series using EEMD, SSA, VMD, and DWT as well as how it was partitioned into training, development, and testing sets including how the mixed sampling approach was carried out. It would be ideal if the authors could provide the code that they used for these steps and (if possible) the time series used to develop the models. This would allow for the substantial difference in results (between SVR and VMD-SVR) to be validated.

Reply:

Generating validation samples from the decompositions of appended sets (i.e., appended decompositions), which is performed before the mixing-and-shuffling step, helps a lot in improving the prediction performance. We have discussed why and how generating validation samples from appended decompositions and mixing and shuffling the training and development (half validation) samples improve the generalization ability of TSDP models.

We agree with you that providing code and data is helpful for the readers to validate our research results. The code and data for this study were uploaded to GitHub (see `https://github.com/zjy8006/`

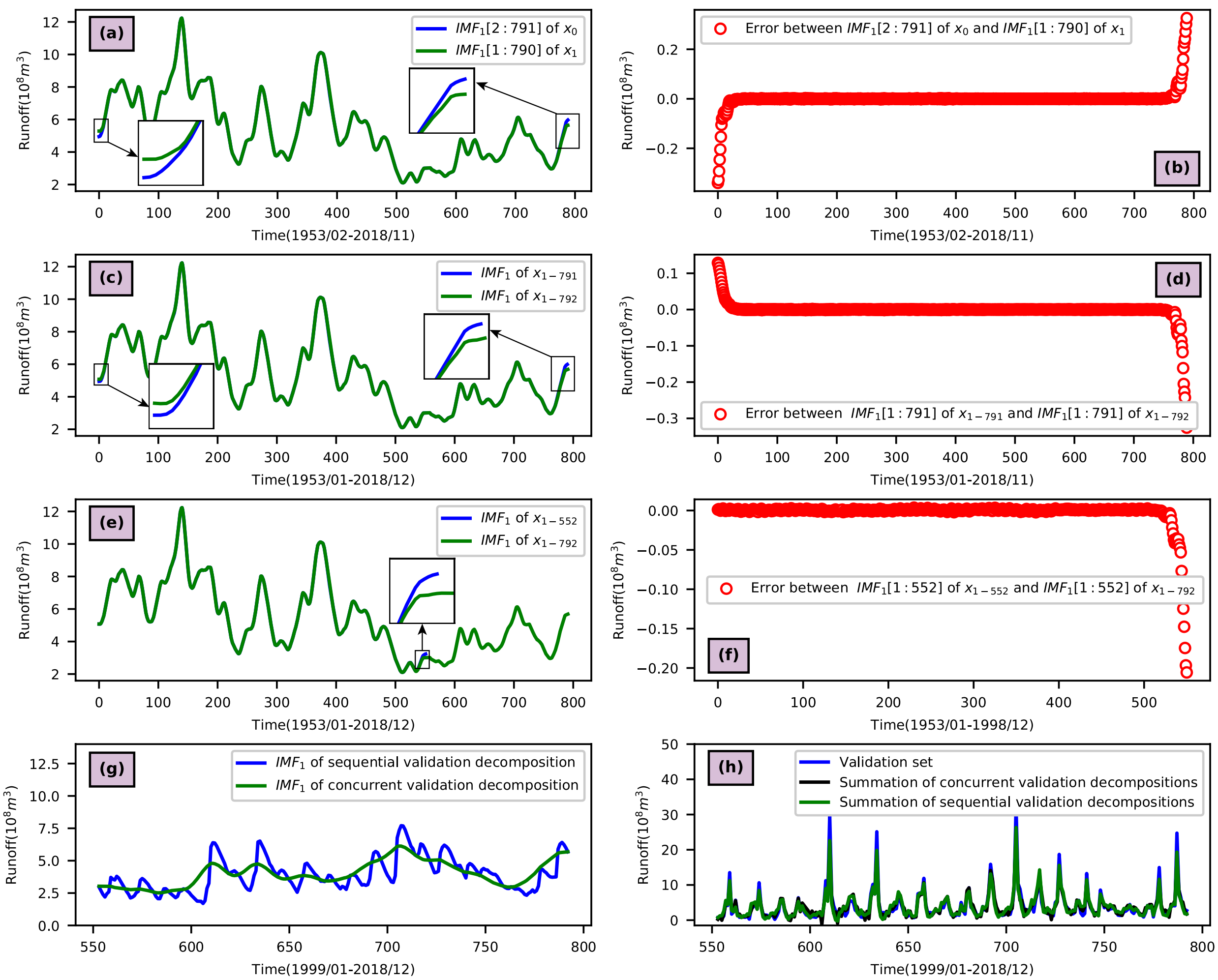

Figure 1: Diagram of boundary effect for illustrating instances of VMD (a and b) shift-variance, (c and d) sensitivity of appending one data point, (e and f) sensitivity of appending several data points, (g) difference between sequential and concurrent validation decompositions and (h) difference between the summation of sequential and concurrent validation decompositions at Huaxian station.

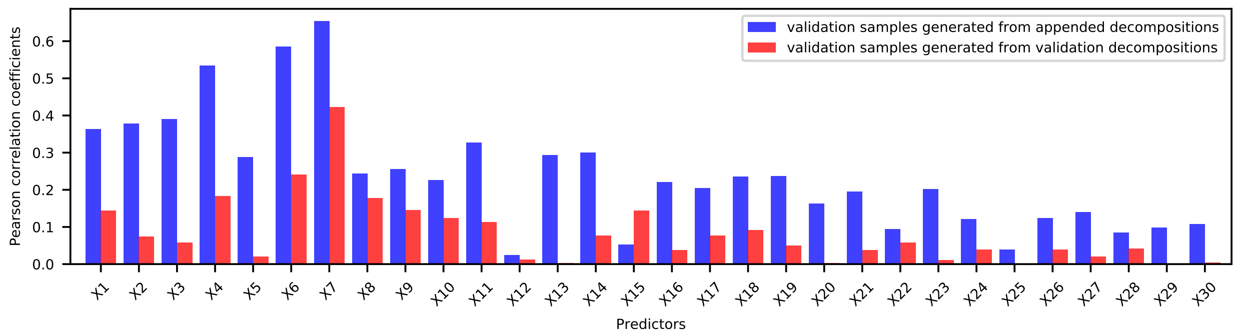

Figure 2: Absolute Pearson correlation coefficients (PCC) between predictors and predicted target of validation samples generated from VMD appended decompositions and validation decompositions at Huaxian station. .

MonthlyRunoffForecastByAutoReg.git) (see readme for how to reproduce our study) and we will later upload the code and data for this study to an open-source data repository (e.g., Mendeley Data).

Comment from John Quilty — Some other points include:

1. The authors seem to be familiar with the framework proposed by Quilty and Adamowski (2018) and note that these authors avoided boundary effects through their approach (Wavelet Data-Driven Forecasting Framework, WDDFF). Why did the authors not compare the WDDFF against their VMD approach? As it stands, each of the comparison methods (EEMD, SSA, and DWT) used in this study are all impacted by the boundary effect. At the end of this review, the reviewer has included a MATLAB code for how the authors could obtain boundary-corrected wavelet and scaling coefficients through the maximal overlap discrete wavelet transform. If the authors use this script to decompose their input time series and include it in their SVR model, they can easily replicate the WDDFF from Quilty and Adamowski (2018).

Reply: We thank you for providing the Matlab implementation of boundary-corrected maximal overlap discrete wavelet transform (BCMODWT) to build WDDFF. We have compared the WDDFF framework with the TSDP framework. However, we think the WDDFF is not feasible for our case study. Because our underlying data does not include meteorological observation and we have to choose the explanatory variables from the monthly runoff. Additionally, our underlying data only contain 792 monthly runoffs, and remove the boundary-affected decompositions will lead to the sample size less than 792. In the revised manuscript, we have selected twelve lagged monthly runoff as explanatory variables and tested several wavelet functions (haar, db1, fk4, coif1, sym4, db5, coif2, and db10) and decomposition levels (1, 2, 3 and 4). However, the WDDFF framework based on BCMODWT and SVR, namely BCMODWT-SVR failed to predict our underlying data (see Fig.3 and Fig. 4).

Comment from John Quilty — Some other points include:

2. Also, since the authors are following only a univariate time series forecasting problem without considering exogenous variables such as rainfall, evaporation, it would also seem plausible that they should compare their framework to simple time series methods such as ARIMA or even more appropriately, fractionally-differenced ARIMA (also known as Hurst-Kolmogorov processes, HKp) that are known to be suitable for forecasting time series with multiscale behaviour. The

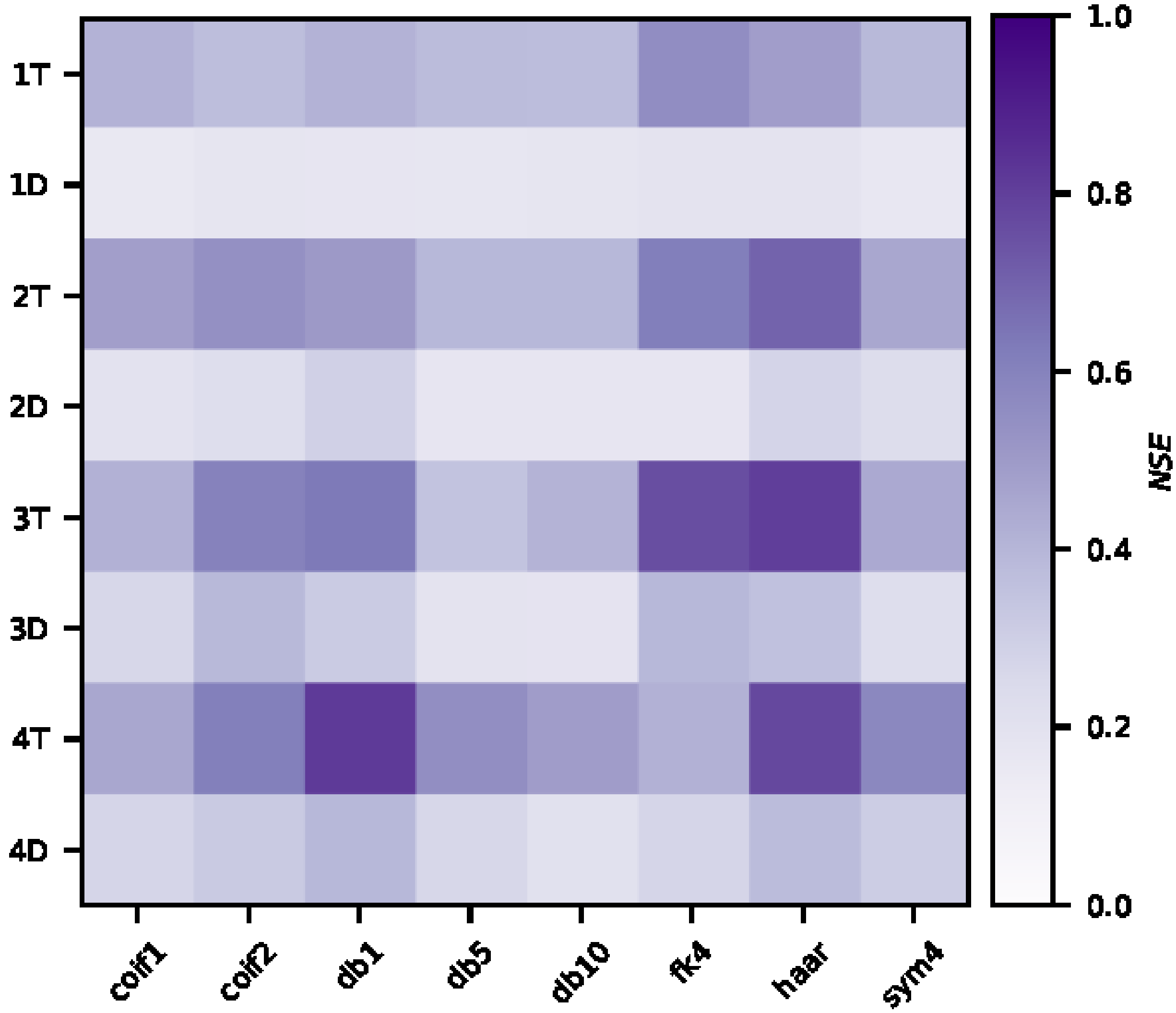

Figure 3: Nash–Sutcliffe efficiency of BCMODWT-SVR for different wavelets and decomposition levels. The horizontal axis represents the wavelets and the vertical axis represents the decomposition level together with the modeling stage (e.g., 1T and 1D represent the training and development stage with a decomposition level of 1).

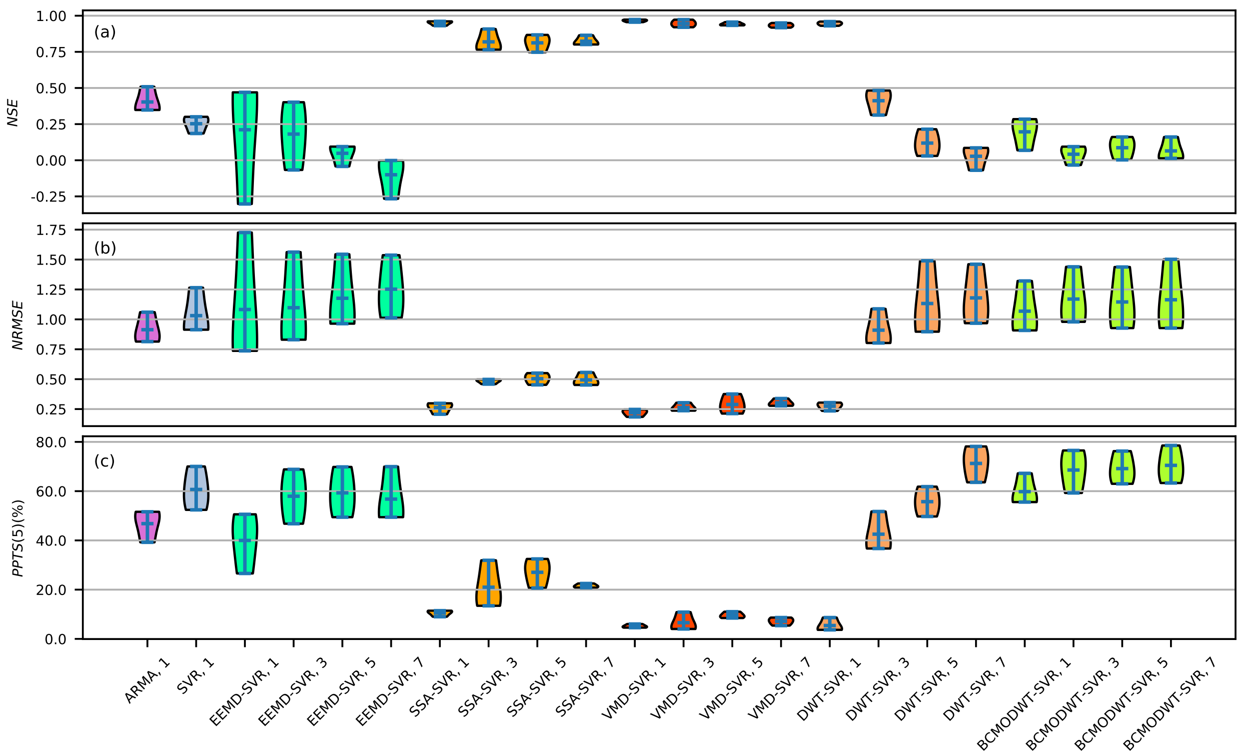

Figure 4: Violin plots of the evaluation criteria during the testing stage for the TSDP and benchmark models (the horizontal axes represent the model and leading time, e.g., "VMD-SVR, 1" represents VMD-SVR model for 1-month-ahead runoff forecasting).

HKp method has been shown to be a useful method for monthly streamflow forecasting, with the potential to outperform common machine learning methods (k nearest neighbours, neural networks) (Koutsoyiannis et al., 2010).

Reply: We have added the ARIMA models for predicting the underlying data of this study.

Comment from John Quilty — Some other points include:

3. Line 266-7: 'Therefore, only a few decomposition values of the training set are affected by the boundary effects.' Can the authors validate this claim (i.e., via a formula or through an experiment)? How can one determine which training records in the various VMD components include boundary effects?

Reply: That's a misstatement. We have confirmed that every decomposition of VMD will be affected by boundary effects through testing VMD shift-variance and sensitive to the addition of new data. The decomposition errors except for the boundary decomposition errors can be ignored because these decomposition errors are very close to zero (see Fig.1 b,d, and f). Since the training set was concurrently decomposed, this sentence means that most training decomposition errors except for the boundary decomposition errors can be ignored. We have rephrased this sentence in the revised manuscript.

Comment from John Quilty — Some other points include:

4. VMD has many tuning parameters that, by the discussion in section 2.1, seems to greatly impact VMD performance. How then is VMD more user-friendly than the MODWT, which only requires the selection of a decomposition level and wavelet filter (although not trivial), for which there are only a finite number? From what I can tell, the parameters in VMD (aside from the selection of the number of IMFs) can take on an infinite number of values. . .

Reply: We agree with you that the VMD has more parameters that should be pre-assigned than MODWT. However, we think the VMD is more controllable than MODWT through the VMD parameters. There are four parameters, i.e., the decomposition level (K), the secondary penalty parameter (α), the noise tolerance (τ), and the convergence tolerance (ϵ), mainly affect the decomposition performance of VMD. A value of K that is too small may lead to poor IMF extraction from the input signal, whereas a too-large value of K may cause information redundancy in the IMFs. A too-small value of α may lead to large bandwidth, information redundancy and additional noise to be included in the IMFs. A too-large value of α may lead to a very small bandwidth and loss of some signal information. The Lagrangian multiplier in VMD hinders the convergence when $\tau > 0$ is used and the signal has a large noise level. This can be avoided by setting τ to 0. Additionally, the value of ϵ affects the reconstruction error of the VMD decomposition. As suggested by Zuo et al. (2020), the values of α, τ, and ϵ were set to 2000, 0, and 1e-9, respectively. Setting τ to 0 can remove noise in the original time series as much as possible and setting ϵ to 1e-9 can obtain more accurate decomposition results. Setting the α to 2000 tends to get small bandwidth, hence, avoid information redundancy and additional noise to be included in the decomposed signal components. As suggested by Xu et al. (2019), the decomposition results are very sensitive to the K. Therefore, we only have to tune K by observing the center-frequency aliasing (see Fig. 5) as suggested by Zuo et al. (2020). This can avoid mode mixing and extract more uncorrelated signal components. The VMD is more controllable than DWT or MODWT because we do not know how to select wavelet functions and decomposition levels to obtain uncorrelated signal components with low noise level.

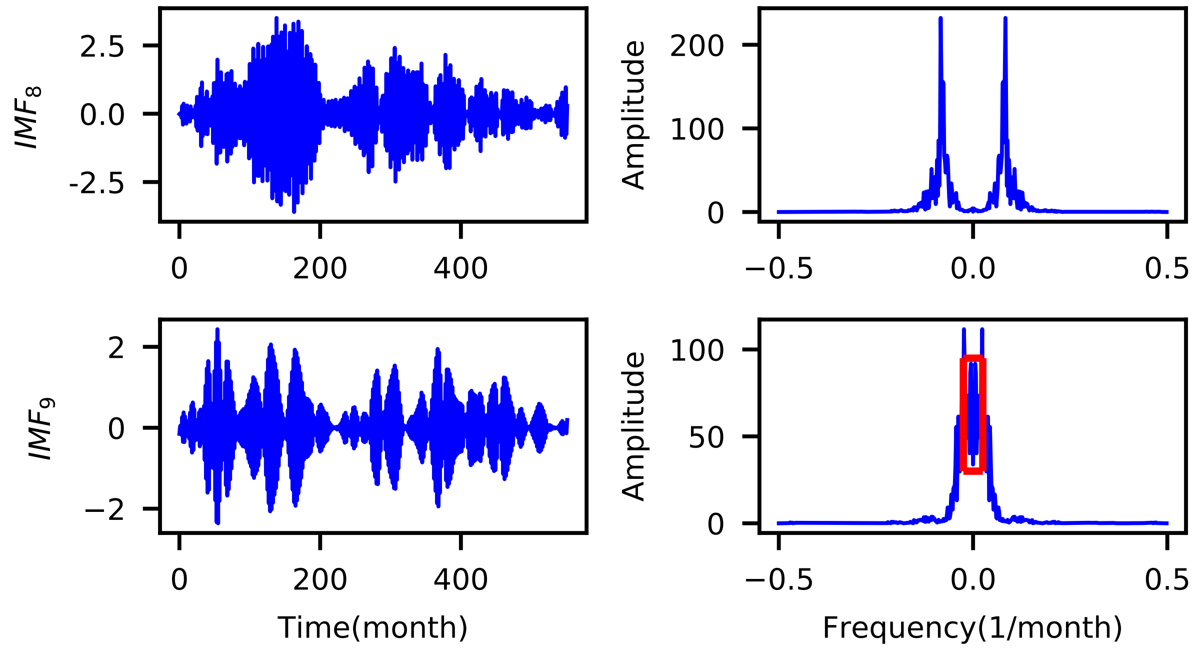

Figure 5: Center-frequency aliasing for the last signal component of the Huaxian station.

Comment from John Quilty — Some other points include:

5. I think Line 320 should be re-cast in light of the fact that selecting the right combination of VMD parameters is technically more computationally-intensive than for the DWT or MODWT.

Reply: We agree with you that tuning the all the four parameters (the decomposition level (K), the secondary penalty parameter (α), the noise tolerance (τ), and the convergence tolerance (ϵ)) is computationally-intensive than DWT or MODWT. However, we can control decomposition results in term of mode mixing and noise with the VMD parameters. We only need to tune the most sensitive parameter, i.e., the decomposition level, which has been demonstrated worked in our case study. Therefore, we think VMD is not more computationally-intensive than DWT or MODWT because the most VMD parameters do not need to be tuned.

Comment from John Quilty — Some other points include:

6. For Experiment 5, only odd numbered lead times were considered (3, 5, 7, and 9 months ahead). Why were even numbered lead times (2, 4, etc. months ahead) not considered?

Reply: We aim to evaluate the performance gap of the TSDP models for long lead times and the workload for evaluating both the odd- and even-numbered lead times is huge. Therefore, we think only evaluate the odd-numbered (or even-numbered) lead times is enough to tell the difference of models.

Comment from John Quilty — Some other points include:

7. Section 3.4: which open-source software was used for EEMD, SSA, VMD, and DWT?

Reply: We have clarified the open-source software of EEMD, SSA, VMD, and DWT.

Comment from John Quilty — Some other points include:

8. Given that a Bayesian approach (BOGP) was used for SVR hyper-parameter optimization, could it not also be used to select the VMD-related parameters? One would think that you could use the BOGP to optimize both VMD and SVR parameters at once. If possible, I think it would be interesting for the authors to consider this. If it is not feasible, a short discussion on why it is not feasible would be interesting.

Reply: Bayesian optimization is a sequential design strategy for global optimization of black-box functions. We did not search VMD parameters using BOGP because it is hard to define an objective function for VMD (for SVR the objective function is mean square error). Besides, only the decomposition level is needed to be tuned in the case study, and determining the decomposition level by observing the center-frequency aliasing can avoid mode mixing to obtain more uncorrelated signal components.

Comment from John Quilty — Some other points include:

9. Why was six-fold cross-validation selected for hyper-parameter optimization (why not 3, 5, or 10-fold cross-validation)?

Reply: The CV fold is a vital parameter that influences the forecasting performance of TSDP models. However, there is no theoretical method to determine the CV fold. The 10-fold CV and leave-one-out CV (LOOCV) are two frequently-used methods (Zhang and Yang, 2015; Jung, 2018). The research results of Zhang and Yang (2015) indicated the LOOCV has a better performance than a 10-fold or 5-fold CV. However, LOOCV is computationally expensive. Additionally, Hastie et al. (2009) empirically demonstrated that 5-fold CV sometimes has lower variance than LOOCV. Therefore, the selection of cross-validation folds needs to consider specific application scenarios. We used the 6-fold CV in the previous version of the manuscript because we referred to an SVR model example. We have changed the 6-fold CV to the frequently-used 10-fold CV scheme in the revised manuscript rather than LOOCV due to the limited computational resources. Additionally, the difference between the 6-fold CV and 10-fold CV is small in the case of this study.

Comment from John Quilty — Some other points include:

10. Normally one has to set a range for the different hyper-parameters in the BOGP approach. What range was set for the various SVR hyper-parameters? It would be good to include what guided your selection of these particular ranges.

Reply: We have clarified the search range for the parameters of SVR.

Comment from John Quilty — Some other points include:

11. Line 495: Figure 8 (a) – I find it hard to agree with the statement that the training data is 'barely affected by the boundaries'. Between record 550 and 555 there is a difference between the red and blue lines of 2 *108 m3! I think one can hardly dismiss this as being a small difference. . . I suggest acknowledging this rather large discrepancy as something significant.

In fact, this is one of the issues of VMD, EEMD, etc. They are not shift-invariant and are sensitive to the addition of new observations (see supporting R code at the end of this review). I suggest the authors discuss in detail these disadvantage of VMD, especially in relation to the MODWT, which does not suffer from these problems and which may also be used, in a mathematically sound manner, to decompose multiscale and/or non-stationary time series into sub-time series capturing their prominent features (which potentially includes trends, periodicities, transients, etc.). I think more effort needs to be devoted to clearly identifying the particular advantages of using the VMD-based approach in this study over the MODWT (which again, does not suffer from such issues).

Reply: That's a misstatement. We have rephrased this sentence in the revised manuscript. We agree with you that the training data is affected by boundary effects. However, this sentence means that the most decomposition errors except for the boundary decomposition errors caused by boundary effects can be ignored. Although the boundary decomposition errors are large, only a small number of training samples are affected by these boundary decomposition errors. We agree with you that VMD, EEMD, DWT, and SSA suffer form the shift-variance and sensitive to the addition of new data, which lead to decomposition errors. The boundary-corrected MODWT can avoid this problem. However, this study aims to propose a general solution to this problem by using different approaches. In other words, we think building practical forecasting models using VMD, EEMD, DWT and SSA without correcting and removing the boundary decompositions is worth to try. We have compared VMD, EEMD, DWT, SSA, and BCMODWT in the revised manuscript.

Comment from John Quilty — Some other points include:

12. Figure 8(b) drives the above-mentioned point home much further... Comparing Figure 8 (a) and Figure 8 (b) it also appears to be the case that the validation data and training data come from distributions, too. It would seem logical that the forecasting model should be updated to account for this change through time (e.g., perhaps through a Kalman Filter)?

Reply: As shown in Fig. 1(f), only a small number of decompositions at the edges of the training set suffers from boundary effect. As shown in Fig.1(g), all the validation decompositions are suffering from boundary effects. Therefore, we think the training decompositions and validation decomposition have a different (error) distribution. Since the training and validation set can have different distribution (see Machine learning yearning by Andrew Ng). We think it is not necessary to update the model parameters because generating validation samples from appended decompositions and mixing and shuffling the training samples and development samples (half validation samples) help a lot to improve the generalization ability of the trained models (see Fig. 6). Generating validation samples from appended decompositions obtain more correlated input predictors (see Fig. 2) and mixing and shuffling the training and development samples can assess the validation error distribution during the training stage.

Comment from John Quilty — Some other points include:

13. Figure 10: it would make it much easier to read if the authors reduced the marker size for the different methods in the scatter plots. For the hydrograph plots, it would be good to zoom in on a particular section, perhaps concentrating on the largest peak event?

Reply: We have redrawn the scatter plots to make it easier to read.

Comment from John Quilty — Some other points include:

14. Figure 18: Why was the standard SVR not included in this analysis? I think it should be included to show how much better the other approaches (DWT-SVR, VMD, SVR, etc.) are at longer lead times.

Reply: We have analyzed the standard SVR and ARIMA model in the revised manuscript.

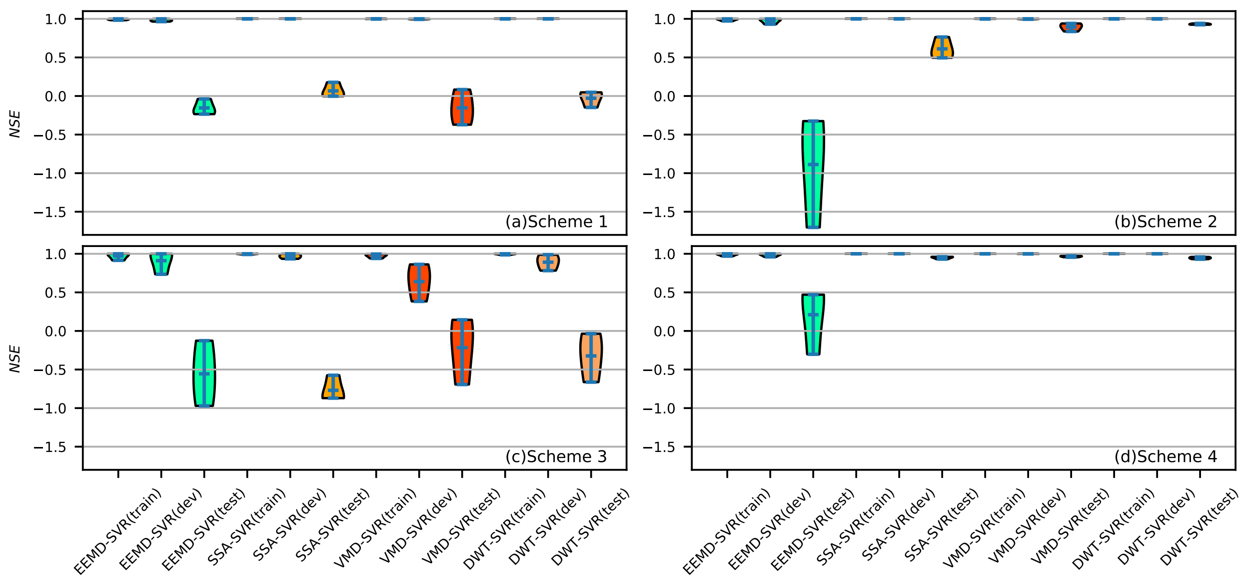

Figure 6: Violin plots of the NSE criterion of TSDP models for one-month-ahead runoff forecasting. (a) Generate samples from training-and-development and testing decompositions without mixing-and-shuffling step. (b) Generate samples from training-and-development and appended decompositions without mixing-and-shuffling step. (c) Generate samples from training and validation decomposition with mixing-and-shuffling step. (d) Generate samples from training and appended decompositions with mixing-and-shuffling step, which is exactly the proposed TSDP framework.

Comment from John Quilty — Some other points include:

15. Line 755-756: 'However, as far as we know, approaches of building a forecasting framework that is adapted to the boundary effect never be tried.'

Are you sure? Quilty and Adamowski (2018) explored the boundary effect existing in popular wavelet-based decomposition methods (DWT, MODWT, etc.), then introduced a set of best practices that addresses these boundary conditions (completely) and implemented these best practices in a new forecasting framework tailored for real-world forecasting. I think one could say that their framework 'adapted to the boundary effect'. In an earlier study by Maheswaran and Khosa (2012), the authors also discussed how to overcome some of the issues of the DWT by choosing a more appropriate wavelet decomposition method (à trous algorithm) that did not suffer from the same boundary conditions. I would also qualify their approach as 'adapting to the boundary effect'. In my opinion, the texted quoted from Line 755-6 is not entirely true and should be revised.

Reply: We have clarified this sentence. What we want to emphasize by this sentence is that the approaches, which are adapted to the boundary effect without correcting and removing the boundary-affected decompositions and providing users with a high confidence level on the unused data, never be tried.

Comment from John Quilty — MINOR ITEMS:

• There are numerous grammatical and spelling errors. I did not note all of these issues. I recommend that the authors carefully check the paper for grammatical and spelling issues (e.g., Line 673 '...increase and decrease patterns...' should read '...increasing and decreasing patterns...').

Reply: We have carefully revised the manuscript and check it for clarity and language. Additionally, our revised manuscript was edited for proper English language, grammar, punctuation, spelling, and overall style by one or more of the highly qualified native English speaking editors at Editideas. The editorial certificate is shown in Fig. 7.

Comment from John Quilty — MINOR ITEMS:

• Line 709: 'Orthometric'? I suggest trying to get your point across using different terms.

Reply: We have clarified this term. The 'Orthometric' term means uncorrelated signal components.

Comment from John Quilty — MINOR ITEMS:

• Line 761: simulating or forecasting? This applies to the whole paragraph. Simulation and forecasting are generally regarded as different procedures in hydrology.

Reply: We have rephrased this paragraph in the revised manuscript.

Comment from John Quilty — MINOR ITEMS:

• Line 764: It is not clear what is meant by '...predictor-runoff relationship and the decomposition-runoff relationship'.

Reply: We have clarified this sentence. We mean the relationship between input predictors and the output targets, and the relationship between the original signal and decomposed signal components.

Comment from John Quilty — MINOR ITEMS:

辑思编译
EDITIDEAS

EDITORIAL CERTIFICATE

This document certifies that the manuscript listed below was edited for proper English language, grammar, punctuation, spelling, and overall style by one or more of the highly qualified native English speaking editors at Editideas.

Manuscript title:

Two-stage Variational Mode Decomposition and Support Vector Regression for Streamflow Forecasting

Authors:

Ganggang Zuo, Jungang Luo, Ni Wang, Yani Lian, Xinxin He

Date Issued:

2020-4-29

Certificate Number:

E-202004120011

Editing Certified
Editideas
Edit Ideas

This document certifies that the manuscript listed above was edited for proper English language, grammar, punctuation, spelling, and overall style by one or more of the highly qualified native English speaking editors at Editideas. Neither the research content nor the authors' intentions were altered in any way during the editing process. Documents receiving this certification should be English-ready for publication; however, the author has the ability to accept or reject our suggestions and changes.

Editideas provides a range of editing, translation and manuscript services for researchers and publishers around the world. Our top-quality PhD editors are all native English speakers from famous institutions cross the U.S.. Our editors come from nearly every research field and possess the highest qualifications to edit research manuscripts written by non-native English speakers.

Figure 7: The editorial certificate.

• Line 765: I think you mean accuracy instead of reliability (the latter is generally measured using probabilistic performance metrics). The same comment also applies to the sentence two lines below.

Reply: You are right, we have clarified this in the revised manuscript.

Comment from John Quilty — MINOR ITEMS:
• Line 773: 'lead' not 'leading'.

Reply: We have revised this term.

Comment from John Quilty — MINOR ITEMS:
• Line 782: I would rephrase point 'c'. Perhaps mention something along lines 'Although some overfitting of the VMD-SVR occurs, the model still provides accurate out-of-sample forecasts'.

Reply: Thanks. We have rephrased point 'c'.

Comment from John Quilty — MINOR ITEMS:
• Line 789-90: Such as. . . ? It would be good to provide some ideas concerning how you think this can be realized.

Reply: Such as using interpretable models (e.g., decision trees) to analyze feature (predictor) importance, using partial dependence plots to observing the global or local convergence, visualizing the model structure and parameters to analyze how the model structure and parameters influence the prediction results, generating against data to test the model behavior.

References

Du, K., Zhao, Y., Lei, J., 2017. The incorrect usage of singular spectral analysis and discrete wavelet transform in hybrid models to predict hydrological time series. J. Hydrol. 552, 44–51. doi:10.1016/j.jhydrol.2017.06.019

Hastie, T., Friedman, J., and Tibshirani, R.: The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second, Springer Series in Statistics, Springer-Verlag New York, New York, NY, Online-Ressource, 2009.

Jung, Y.: Multiple predicting K -fold cross-validation for model selection, Journal of Nonparametric Statistics, 30, 197–215, doi:10.1080/10485252.2017.1404598, 2018.

Koutsoyiannis, D., Yao, H., Georgakakos, A., 2010. Medium-range flow prediction for the Nile: a comparison of stochastic and deterministic methods. Hydrol. Sci. J. 53, 142–164. doi:10.1623/hysj.53.1.142

Maheswaran, R., Khosa, R., 2013. Wavelets-based non-linear model for real-time daily flow forecasting in Krishna River. J. Hydroinformatics 15, 1022. doi:10.2166/hydro.2013.135

Maheswaran, R., Khosa, R., 2012. Comparative study of different wavelets for hydrologic forecasting. Comput. Geosci. 46, 284–295. doi:10.1016/j.cageo.2011.12.015

Quilty, J. and Adamowski, J.: Addressing the incorrect usage of wavelet-based hydrological and water resources forecasting models for real-world applications with best practices and a new forecasting framework, Journal of Hydrology, 563, 336–353, doi:10.1016/j.jhydrol.2018.05.003, 2018.

Wang, Y., Wu, L., 2016. On practical challenges of decomposition-based hybrid forecasting algorithms for wind speed and solar irradiation. Energy 112, 208–220. doi:10.1016/j.energy.2016.06.075

Xu, B., Zhou, F., Li, H., Yan, B., and Liu, Y.: Early fault feature extraction of bearings based on Teager energy operator and optimal VMD, ISA Transactions, 86, 249–265, doi:10.1016/j.isatra.2018.11.010, 2019.

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. doi:10.1016/j.jhydrol.2015.09.047

Zhang, Y. and Yang, Y.: Cross-validation for selecting a model selection procedure, Journal of Econometrics, 187, 95–112, doi:10.1016/j.jeconom.2015.02.006, 2015.

Zuo, G., Luo, J., Wang, N., Lian, Y., and He, X.: Decomposition ensemble model based on variational mode decomposition and long short-term memory for streamflow forecasting, Journal of Hydrology, 585, 124776, doi:10.1016/j.jhydrol.2020.124776, 2020.