Research papers

Variable transformations in the spectral domain – Implications for hydrologic forecasting

Ze Jiang , Ashish Sharma [*], Fiona Johnson

School of Civil and Environmental Engineering, University of New South Wales, Sydney, New South Wales, Australia

ARTICLE INFO

ABSTRACT

Forecasting of hydrologic extremes across a range of timescales is critical for minimizing the socio-economic costs of these events. Regression-based prediction is commonly adopted even in operational forecasting systems, often necessitating the use of distributional transformations to improve model specifications. One of the issues in such predictions, however, is the marked differences that distinguish the frequency spectrum of the hydrologic response from the predictor variables used. This raises the question of whether there exists an optimal predictor variable transformation that can mimic the frequency spectrum embedded in the observed response variable series. The present paper discusses the need to transform predictor variables to improve hydrologic forecasts, and specifically focuses on the frequency domain of the variables involved. A number of alternatives using wavelet-based approaches are presented as a means of transforming the variance associated with different frequency bands in each predictor variable. The limitations and advantages of these transformations are summarized and demonstrated using synthetic examples. A stepwise variance transformation framework is further proposed that facilitates transformations of the residual error from a given predictor variable conditioned on existing (or pre-identified) predictor variables. Results of the stepwise framework demonstrate that the response is better characterized using both synthetic case studies and when applied to forecasting the El Niño–Southern Oscillation (ENSO) over long lead times.

1. Introduction

Hydrologic time series are often positively skewed with a larger proportion of low values and very few high to extreme values. Using a predictive model that adopts as predictors variables that are more evenly distributed, creates difficulties in modelling especially when linear predictive alternatives are used. It is for this reason that we adopt distributional transformations, as is the case when relating total dissolved solids (TDS) concentrations and streamflow, where the discharge will often be logarithmically transformed (Mosteller and Tukey, 1977). Such a transformation is often sufficient to ensure a distributional correspondence between the response and associated predictors, but is inadequate if these variables exist as a time series and there is a mismatch in their respective frequency spectrums. Consider, for instance, the teleconnection between the El Niño–Southern Oscillation (ENSO) and rainfall, the association of which has been studied extensively in the literature (D'Arrigo et al., 2005; Pui et al., 2012; Torrence and Compo, 1998; Westra and Sharma, 2010). Here, a simple linear model is not sufficient to relate the two because the response (rainfall)

exhibits a markedly different spectrum to the predictors. As such, for many complex natural systems, formulating a simple predictive model becomes challenging especially when there exists no physical rationale to adopt an explicit or implicit transformation of the variables that define the system.

There are three key reasons we use variable transformations as a means of improving model specification:

1) to simplify the patterns in the variable (e.g., more symmetric or constant) by removing known modes of variability including random noise,
2) to allow relationships between variables to be expressed using simpler (often linear) models, and
3) to force the spectrum of the predictor variables to exhibit a greater consistency with that of the response being modelled (Helsel and Hirsch, 2002; Hyndman and Athanasopoulos, 2014; Jiang, Sharma et al., 2020; McInerney et al., 2017; Wu et al., 2019).

Table 1 lists a variety of mathematical transformations that have

* Corresponding author.
 *E-mail address:* A.Sharma@unsw.edu.au (A. Sharma).

Table 1

Summary of mathematical transformations used in hydro-climatology.

Data transformation	Equation	Notation
Log transformation	$y' = \log(y + \lambda_2)$	log
Reciprocal (inverse) transformation	$y' = \dfrac{1}{y + \lambda_2}$	*recip*
Box-Cox transformation (Box and Cox, 1964)	$y' = \begin{cases} \dfrac{(y + \lambda_2)^{\lambda_1} - 1}{\lambda_1}, & if \lambda_1 \neq 0; \\ log(y + \lambda_2), & if \lambda_1 = 0. \end{cases}$	*bc*
Log-sinh transformation (Wang et al., 2012)	$y' = \dfrac{log(sinh(\lambda_1 y + \lambda_2))}{\lambda_1}$	log-*sinh*
Variance transformation (Jiang, Sharma, et al., 2020)	$X' = \widehat{R}_X \alpha$	VT
Phase randomization (Chavez and Cazelles, 2019)	$x' = W_{x'}(t,f)^{-1} \; W_{x'}(t,f) = |(W_x(t,f))| exp(i\varphi_{noise}(t,f))$	PR

Note: y is the response and y' is the transformed response; X is the predictor and X' is the transformed predictor; x is the original variable and x' is the simulated variable. λ_1 and λ_2 are power parameters; $\widehat{R}_X$ is the standardized wavelet reconstruction matrix of X and α is the transformed standard deviation matrix; $W_x(t,f)$ is the wavelet transform of the target variable and $W_{x'}(t,f)^{-1}$ is the inverse wavelet transform of $W_{x'}(t,f)$, which is the transformed time–frequency representation of x with the randomized phase from Gaussian white noise $\varphi_{noise}(t,f)$.

been adopted for hydrological applications in the past. These include transformations that modify the probability distribution, such as logarithmic, reciprocal, and Box-Cox (Box and Cox, 1964), mainly applied to the response (y). Note that log and reciprocal transformations are variants of the Box-Cox transformation when $\lambda_1 = 0$ and $\lambda_1 = -1$, respectively. The log-sinh transformation has also been proposed to address heteroskedasticity often resulting from a Box-Cox Transformation (Wang et al., 2012). Table 1 also includes a variance
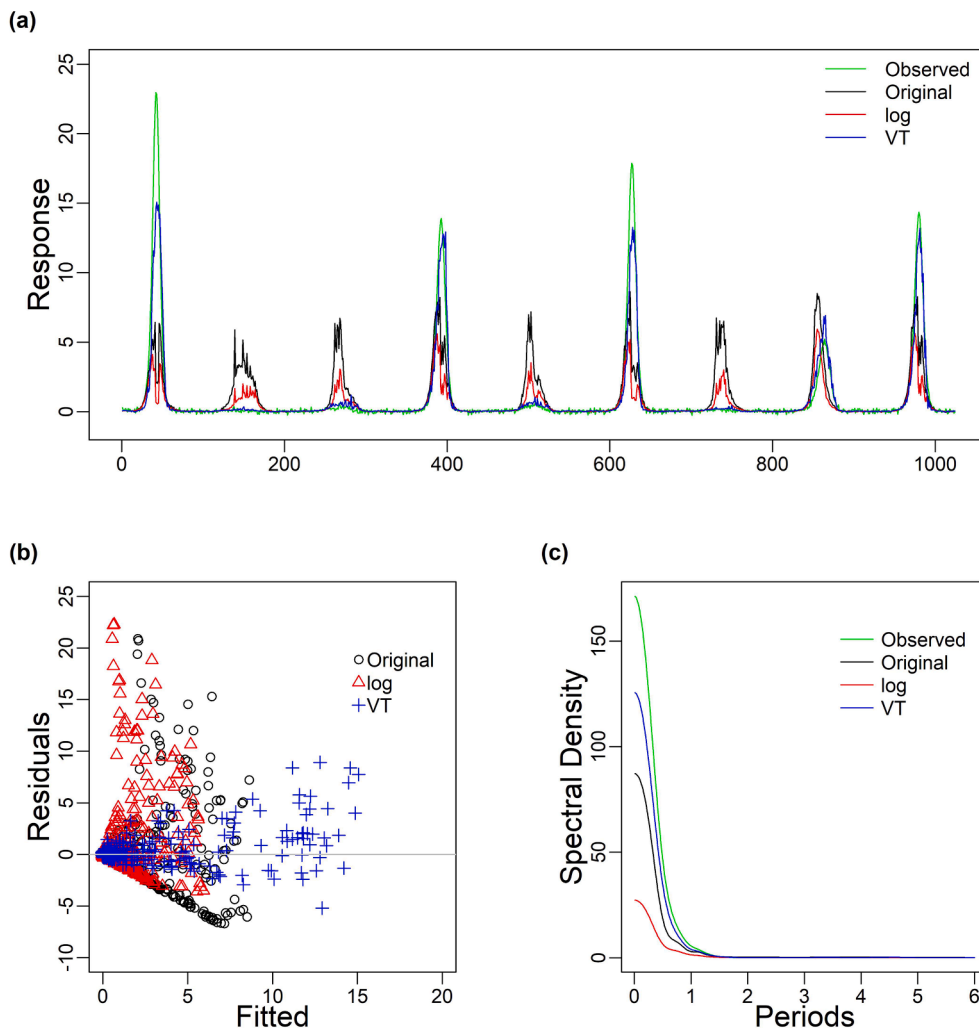
Fig. 1. Example of log transformation and variance transformation using residual plot and spectrum analysis: (a) Observed and fitted response using original data, log-transformed response, and variance-transformed predictors in the time domain (b) Scatter plot of fitted response against the associated residuals using three different approaches; (c) Spectral density plot of observed and fitted response using three different approaches.

transformation (Jiang, Sharma, et al., 2020) which was developed to refine the spectral representation of predictor variables, and phase randomization as a related means of surrogate data generation (Brunner et al., 2019; Chavez and Cazelles, 2019; Schreiber and Schmitz, 2000), both defined in the frequency domain which forms the focus of our study. In all cases, transformations aim to assist in the specification of a more accurate and stable predictive model.

Fig. 1 illustrates two different transformations for a nonlinear system where predictors and the response have markedly different spectral properties. The nonlinear dataset is characterized by the Rössler system with a *k*-nearest neighbor (knn) regression model (more details are provided in Section 3.4). The log transformation in the temporal domain and a recently developed variance transformation in the frequency domain are demonstrated. It is observed that the transformation in the spectrum allows a balanced representation of both regression diagnostics and spectral attributes. Not only residuals are close to normality (although heteroscedasticity still exists), but fitted values are more similar to observations in both the spectral and temporal domains. It is worth noting that the residuals of log transformation are in their original space and the synthetic experiment is constructed to illustrate the merit of the spectral transformation when there is a significant mismatch in the spectrum of the predictors and the response.

In this study, we focus on the variance transformation approach in the frequency domain to further explore its applicability for hydrologic forecasting. As discussed in Jiang, Sharma, et al. (2020), one of the most common ways to characterize the spectrum of a time series is the discrete wavelet transform (DWT). DWT can decompose the original signal into sub-time series representing information at a range of frequencies, which captures periodicity, short- and long-term dependence, and non-stationarity that exist in the original signal. If these decomposed sub-time series are used in a data-driven model, there can be significant improvements in prediction accuracy (Nourani et al., 2014; Sang, 2013). However, the DWT decomposed components at any point in time are calculated using future information as well as past information (Du et al., 2017; Quilty and Adamowski, 2018). This means that DWT cannot be used in forecasting applications, where future information on the predictors is not available. In contrast to DWT, the maximal overlap DWT (MODWT) decomposes the time series independent of future information and a secondary benefit is that it does not require time series to be of dyadic length (Percival and Walden, 2000). Another alternative to these approaches, Algorithme à trous (AT) also has no dependence on future information, and one of its characteristics is that the decomposed sub-time series is redundant but captures the main feature and variability of the original time series (Dutilleux, 1990; Fowler, 2005). More details on the difference and advantages of each wavelet transform mentioned above are provided in Section 2.

There are two common means of using decomposed sub-time series from any wavelet transform in system modeling and forecasting. The direct approach (Nguyen and Nabney, 2010), uses the decomposed sub-times series to predict the target response (e.g., Kişi, 2011; Rashid et al., 2018). The multi-component approach first uses decomposed sub-time series to forecast decomposed components of the target response at the equivalent scales and then aggregates these sub-time series predictions (e.g., Rathinasamy et al., 2014; Shafaei and Kisi, 2016). In addition to these approaches, a new technique, namely variance transformation, was developed to improve the predictability of a response by transforming the weights assigned to each of the decomposed frequency components in the predictor and thus forming a new predictor variable that improves the prediction of the response (Jiang, Sharma, et al., 2020). The key assumption of this method is that if the predictor and the corresponding response have similar spectral properties, the predictive model using the transformed predictor will exhibit better accuracy. This method was initially developed based on the DWT and was then extended to include the MODWT so that it could be used in forecasting without requiring future information on the predictor (Jiang, Rashid et al., 2020). In this study, three further contributions to the method are

proposed. Firstly, the theoretical formulation of variance transformation using AT is introduced and tested on a synthetic dynamical system model. Secondly, the original method transformed each predictor individually based on its relationship with the response. Here, a stepwise variance transformation approach is proposed which transforms the predictor variables according to the residual of the response conditional on pre-existing predictors. It is hypothesized that the stepwise transformation further improve the prediction accuracy even when the predictor variables and the associated response have matching spectral attributes. Finally, we comprehensively investigate whether the newly developed stepwise variance transformation approach can improve long lead forecasting of ENSO.

The remainder of the paper is organized as follows. Section 2 reviews wavelet transforms of DWT, MODWT, and AT to provide a better understanding of their implications on the variance transformation technique. In Section 3, the variance transformation technique derived from DWT and MODWT is reviewed and extended to include AT. Section 4 introduces the stepwise variance transformation and compares it against the direct variance transformation method. The capability of the stepwise variance transformation framework in a real-world forecasting example is demonstrated in Section 5. Summary and conclusions are presented in Section 6.

2. Wavelet transforms

This section provides some basic concepts and introduces notations for readers with limited knowledge about wavelet transforms. Wavelet transforms are used to characterize the spectrum of time series. The variance transformation method of Jiang, Sharma, et al. (2020) uses wavelet transforms, and modifies predictor variable time series to better match a response variable in the frequency domain. This is unlike the Box-Cox and related transformations in Table 1 which attempt to modify the distributional representation of the variable, with smaller changes to the spectrum.

In this section, we review and summarize the characteristics of DWT, MODWT, and AT, respectively to motivate extending the variance transformation method to AT. DWT can be implemented in two ways – either through additive decomposition (also known as multiresolution analysis) or via variance (or energy) decomposition. The variance transformation method is based on the multiresolution analysis (MRA) and thus when we use DWT hereafter in the paper, it refers to DWT-MRA. Given a discrete data sample $X = [x_0, x_1, ..., x_{n-1}]^T$, the DWT of the time series by decomposing the signal into wavelet and scaling coefficients is given by Percival and Walden (2000):

$$W = wX \tag{1}$$

where W is a vector of DWT coefficients ($n \times 1$), and w is the $n \times n$ orthonormal transform matrix. The reconstruction of the original data set can be achieved by the equation (Aussem et al., 1998; Percival and Walden, 2000):

$$X = w^T W = \sum_{j=1}^{J} d_j + a_J \tag{2}$$

where W can be partitioned as $W = [D_1, ..., D_J, A_J]$, including detail coefficients (D_j) and approximation coefficients (A_J); n is the sample size while J is the maximum decomposition level; d_j is the reconstructed details while a_J is the reconstructed approximations. Thus, the reconstruction matrix is given by $R = [d_1, ..., d_J, a_J]$ with a dimension of $n \times (J+1)$, and the associated standard deviation matrix is given by $I = [\sigma_{d_1}, ..., \sigma_{d_J}, \sigma_{a_J}]^T$. This leads to additive decomposition which is known as multiresolution analysis. The MODWT is a variant of the DWT, and MODWT decomposes the original time series X into a $n \times (J+1)$ matrix of wavelet and scaling coefficients $\widetilde{W} = [\widetilde{D}_1, ..., \widetilde{D}_J, \widetilde{A}_J]$, with the associated standard deviation matrix being $\widetilde{I} = [\sigma_{\widetilde{D}_1}, ..., \sigma_{\widetilde{D}_J}, \sigma_{\widetilde{A}_J}]^T$. Similarly,

Table 2
Summary of properties for the three types of wavelet transforms.

Wavelet Method	Additive decomposition	Variance decomposition	No dependence on future data	Dyadic sample size
DWT-MRA	✓	✓		✓
MODWT		✓	✓	
AT	✓		✓	

Note: When the Haar wavelet filter is used, MODWT and AT are equivalent, and both preserve additive and variance decomposition.

AT can decompose the original time series into a $n \times (J+1)$ matrix of $\widetilde{W}^a = [\widetilde{D}_1^a, ..., \widetilde{D}_J^a, \widetilde{A}_J^a]$, and the associated standard deviation matrix is given by $\widetilde{I}^a = [\sigma_{\widetilde{D}_1^a}, ..., \sigma_{\widetilde{D}_J^a}, \sigma_{\widetilde{A}_J^a}]^T$. It is noted that dyadic down-sampling is avoided in the MODWT and AT resulting in wavelet and scaling coefficients that have the same length as the original time series so that their coefficients matrix ($\widetilde{W}$) have a different dimension from W, the coefficients matrix of DWT (Walden, 2001).

The mathematical formulations of all three types of wavelet transforms are presented by Quilty and Adamowski (2018) and Percival and Walden (2000), and readers are referred to their work for more detailed information. Here, we provide a general discussion on all three methods including their strengths and drawbacks which are summarized in Table 2. The main benefit of using DWT-MRA is that it ensures both additive decomposition and variance decomposition. These two features are required to find the theoretically optimal variance transformation (Jiang, Sharma, et al., 2020) discussed in the next section. However, the major challenge in using DWT-MRA for predictions is that it depends on future data to decompose the original time series into the frequency domain. MODWT and AT decompositions are both independent of future information and are therefore appropriate for forecasting applications. The second drawback of DWT-MRA is that it requires a dyadic sample size, whilst MODWT and AT have no restrictions in terms of sample size. Although MODWT and AT have advantages over DWT-MRA on these two points, they only partially meet the requirements for an optimal variance transformation. MODWT only preserves variance in the decomposition while AT only permits additive decomposition. However, the logic of variance transformation can still be applied using either method, but the resulting transformed variable could be dramatically different from the original time series (e.g., the variance of the time series). Theoretically, MODWT is preferable to AT since it ensures variance decomposition so that it aligns with the key idea of variance transformation (i.e., redistributing variance across the frequency domain) while AT gives more redundant information of the original time series. However, the redundancy of information in the decomposition has many advantages in real applications, and an example demonstrating these has been included in Section 3.4.

As discussed by Quilty and Adamowski (2018), there are a number of wavelet filters that have associated benefits and drawbacks in representing the variable time series in the frequency domain. When a Haar wavelet filter is used, MODWT and AT are equivalent, resulting in the same wavelet and scaling coefficients. More importantly, with the Haar wavelet, MODWT (and thus also AT) preserve variance and additive decomposition. As a result, the proposed variance transformation technique can be applied to MODWT and AT. However, there is a risk that the Haar filter may not properly characterize the spectrum of variables of interest and thus we also consider in Section 3 the implications of different choices of the filter. However, the advantage of MODWT or AT
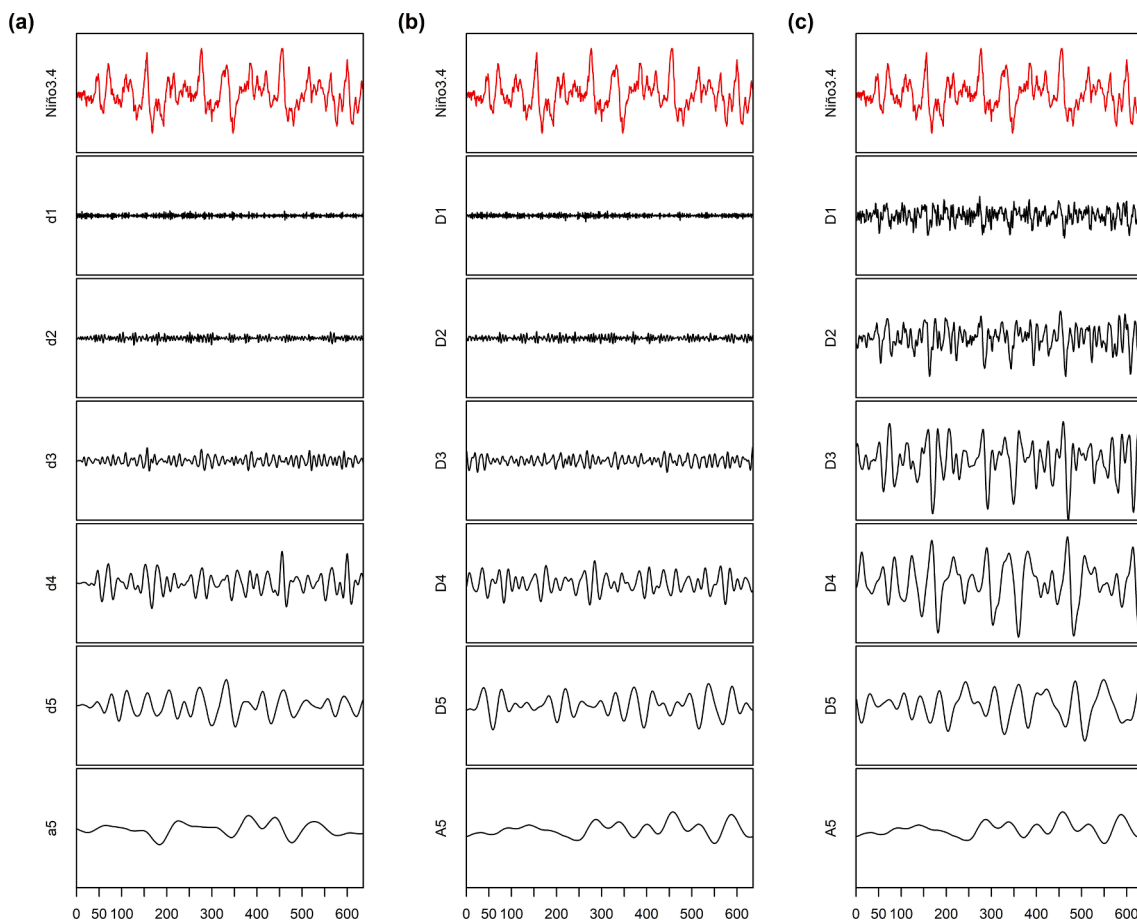

Fig. 2. Illustration of three types of wavelet transforms using db8: (a) DWT-MRA; (b) MODWT; (c) AT.

in being independent of future information makes their use very attractive. Consequently, when we adopt the Haar wavelet filter for MODWT and AT, hereafter it will be referred to as MODWT/AT due to their equivalence in the remainder of our presentation.

Fig. 2 illustrates how the decomposition varies for the three wavelet transforms (i.e., DWT, MODWT, and AT) using the target response from the forecasting example presented in Section 5 (i.e., time series of ENSO index, Niño3.4). The variance in the original time series (red line) is concentrated at low frequencies (high to low frequency shown from upper to lower panels). A long wavelet filter (Daubechies 8, for short db8 or d16) is used to demonstrate the difference between the decompositions in the frequency domain. For reference, the equivalence of MODWT and AT when using the Haar wavelet filter is given in Figure S 1 of the Supporting Information. Note that the time series is padded with zeros to bring the sample size to the next higher power of two (i.e., dyadic sample size) which is necessary for DWT in this study.

3. Variance transformation technique

The DWT-based variance transformation technique was introduced by Jiang, Sharma, et al. (2020), and is briefly reviewed in Section 3.1. The solution to DWT requiring future information was addressed by Jiang, Rashid, et al. (2020) using MODWT, and here we introduce the mathematical basis of the AT-based approach. A synthetic dataset generated from the Rössler system is used to compare the three alternatives of variance transformation. Finally, a short summary of all three wavelet-based variance transformation methods is given.

3.1. Variance transformation using DWT

As described previously, DWT decomposes the original time series X into a vector of wavelet and scaling coefficients $W = [D_1, ..., D_J, A_J]$ with a dimension of $n \times 1$. Reconstructed details and approximations using wavelet and scaling coefficients are given by $R = [d_1, ..., d_J, a_J]$ with a dimension of $n \times (J + 1)$, and the associated standard deviation matrix is given by $I = [\sigma_{d_1}, ..., \sigma_{d_J}, \sigma_{a_J}]^T$. The resultant R is also called DWT-MRA. In the form of matrix multiplication, the original time series X can be expressed as $X = \widehat{R}I$, where $\widehat{R}$ is the standardized reconstruction matrix. Our objective is to find a transformed predictor variable X' using the rotated variance structure α following the direction of covariance vector C between the variable set $(Y, \widehat{R})$. Y is the associated response of X. This can be written as (Jiang, Sharma, et al., 2020):

$$X' = \widehat{R}\alpha$$
$$\alpha = \sigma_X \widehat{C} \tag{3}$$

$$C = \frac{1}{n-1} Y^T \widehat{R} = \left[S_{\widehat{Y d_1}}, ..., S_{\widehat{Y d_J}}, S_{\widehat{Y a_J}} \right] \tag{4}$$

where $\widehat{C}$ is the normalized covariance vector of C. Basically, the transformed predictor X' is obtained by redistributing the variance in its spectrum whilst ensuring the resulting variance of transformed time series X' is the same as the original predictor X. In the end, a theoretical optimal prediction accuracy, as measured by root mean square error (RMSE), can be derived as (Jiang, Sharma, et al., 2020):

$$RMSE_{min} = \sqrt{\frac{n-1}{n}(\sigma_Y^2 - \|C\|^2)}, \tag{5}$$

where σ_Y denotes the standard deviation of the response Y, and $\|\cdot\|$ denotes the norm of a vector in Euclidean space. As can be seen in Equation (5), both the positive and negative signs of covariance can result in the same optimal prediction accuracy. However, the sign of covariance affects how the transformed predictor variables behave, and we found that the transformed predictor variables sometimes show trends physically inconsistent with observations if only the positive sign

of covariance is used. The correct sign for the covariance is that it should match the sign of its associated correlation coefficient. An auto-selection process to achieve this is therefore recommended and has been applied to all variance transformation alternatives discussed in this study. The implementation of the auto-selection process ensures the trend of the transformed predictor is consistent with observations.

3.2. Variance transformation using MODWT

As discussed in Section 2, MODWT, a modified version of DWT, has no limitation on the sample size and has no future data dependence (Nason and Von Sachs, 1999). MODWT decomposes the original time series X into a $n \times (J+1)$ matrix of wavelet and scaling coefficients $\widetilde{W} = [\widetilde{D}_1, ..., \widetilde{D}_J, \widetilde{A}_J]$, and the associated standard deviation matrix is given by $\widetilde{I} = [\sigma_{\widetilde{D}_1}, ..., \sigma_{\widetilde{D}_J}, \sigma_{\widetilde{A}_J}]^T$. MODWT always preserves the variance decomposition, which provides a way to investigate and transform the variance structure of the coefficient matrix, $\widetilde{W}$. It is mentioned that the coefficients matrix W decomposed from DWT has a dimension of $n \times 1$ while the coefficients matrix $\widetilde{W}$ from MODWT has a dimension of $n \times (J + 1)$. This is another reason it can be used for the variance transformation directly. Consequently, using the covariance C between the variable set $(Y, \widehat{\widetilde{W}})$ the variance transformed X' can be obtained by (Jiang, Rashid, et al., 2020):

$$X' = \widehat{\widetilde{W}}\alpha$$
$$\alpha = \sigma_X \widehat{C} \tag{6}$$

where $\widehat{\widetilde{W}}$ is the standardized coefficients matrix $\widetilde{W}$. Note that although the variance decomposition is ensured with the MODWT, it cannot lead to the same variance as present in the original time series since the decomposed components are not additive after the transformation even when the Haar wavelet filter is used.

3.3. Variance transformation using AT

Algorithme à trous, introduced by Holschneider et al. (1990) and Shensa (1992), is a wavelet transform algorithm designed to overcome the problems that occur in DWT associated with data reduction due to down-sampling and future data dependence. AT is a redundant transform because the decomposed time series is computed by repeated filtering with a maximal sampling rate at all dyadic scales, and the inherent redundancy of this transform has many advantages in applications. So, taking a J level decompositions of an original time series of size n for example, AT can decompose the original time series of X into a $n \times (J+1)$ matrix of $\widetilde{W}^a = [\widetilde{D}_1^a, ..., \widetilde{D}_J^a, \widetilde{A}_J^a]$, and the associated standard deviation matrix is given by $\widetilde{I}^a = [\sigma_{\widetilde{D}_1^a}, ..., \sigma_{\widetilde{D}_J^a}, \sigma_{\widetilde{A}_J^a}]^T$. Therefore, the proposed variance transformation using AT can be expressed as:

$$X' = \widehat{\widetilde{W}}^a \alpha$$
$$\alpha = \sigma_X \widehat{C} \tag{7}$$

where $\widehat{C}$ is the normalized covariance vector of C between the variable set $(Y, \widehat{\widetilde{W}}^a)$. As mentioned in Section 2, AT ensures additive decomposition but violates the variance decomposition. This means that the variance transformation in Equation (7) could lead to a transformed variable with the substantially increased variance due to the redundant information included in the decomposition. The implications for this increased variance are considered in the case study presented next. Alternatively, the transformed variable could be rescaled to match the variance of the original time series where appropriate for any particular predictive models.

3.4. Application to synthetic examples

Synthetic datasets generated from statistical models with known attributes give opportunities to evaluate the effectiveness of a method or algorithm. In this study, the Rössler system, a dynamical system that exhibits chaotic dynamics proposed by Rössler (1976), was used to investigate the proposed methods. The Rössler system is defined by the following three differential equations:

$$\frac{dx}{dt} = -y - z,$$
$$\frac{dy}{dt} = x + ay, \qquad (8)$$
$$\frac{dz}{dt} = b + z(x - c).$$

The chaotic system includes three constant parameters, namely a, b and c. The values of $a = 0.2$, $b = 0.2$, and $c = 5.7$ are commonly used (Harrington and Van Gorder, 2017; Strogatz, 2000), and it is initialized from a given condition of $(-2, -10, 0.2)$. The Rössler system generates a dependent variable set (x_t, y_t, z_t), where x_t and y_t are used as predictor variables while z_t is considered as the corresponding response. An illustration of the Rössler system in phase space and time domain is presented in Fig. 3.

In the experiment, the dynamical system is further complicated by adding Gaussian white noise ε to each generated time series, where $\varepsilon \sim N(0, 0.1^2)$. All synthetic time series in this section have the same sample size of $n = 100,000$, and the first half of the data is used as the calibration dataset while the second half is used as the validation dataset. Variance transformation is carried out after the data partition, and the derived covariance from the calibration set is then transferred to the validation set along with the fitted predictive model. The knn regression from Beygelzimer et al. (2006) is used here as the predictive model.

Using the generated data seen above, illustrations of how three different types of wavelet-based transformations perform using two different wavelet filters (Haar and db8) are given in Figs. 4 and 5, respectively. The increased variance of transformed predictors is observed for the AT-based approach using both Haar and db8 wavelet filter, but it will not affect the final prediction performance using regression-based models as is the case presented here (see the following results). Although the transformed predictors from the MODWT-based

approach with db8 are able to match the variance of the original predictor, it fails to characterize the predictor variables properly. In terms of the DWT-based approach, the transformed predictor of y using either Haar or db8 wavelet presents an inconsistent pattern with the original variable. Last, it is clear that wavelet transforms investigated here have minor issues of characterizing predictors at the beginning of the time series. This is the curse of wavelet transforms resulting from boundary related issues, investigated and addressed in Jiang, Rashid, et al. (2020).

A radar chart showing three metrics (including RMSE, standard deviation, and correlation) for all three types of wavelet-based models using original and transformed predictor variables is presented in Fig. 6. Note that since the Haar wavelet filter is adopted, MODWT and AT are equivalent to each other, and the observation is included here mainly for indicating what the optimal value for the standard deviation. In terms of calibration, models using transformed predictor variables perform similarly for all three alternatives, and they are better than models using original predictor variables. For validation, all wavelet-based models outperform the reference model in both correlation and RMSE. It is noted that there is an exceptionally good correlation in the validation results for the MODWT/AT-based approach, which is most likely due to the Rössler system characteristics (i.e., the modeled response (z_t) has a large number of low values) as well as the trade-off among the three metrics. This trade-off has been well described and discussed through MSE decomposition (Gupta et al., 2009). Different wavelet transforms characterize the spectrum of variables differently resulting in different transformed predict variables (seen in Fig. 4), and the MODWT/AT-based approach works better than DWT in this dynamic system.

What we show above is a case under ideal conditions with sufficiently long data samples. However, this rarely happens in reality. Furthermore, there are many other practical issues related to wavelet transforms, such as the selection of wavelet filters and the decomposition level. In Fig. 7, we have assessed the sensitivity of the approach to the choice of wavelet family by using the db8 wavelet filter. As shown in Fig. 5, when a longer wavelet filter is used, DWT and MODWT are not able to characterize the spectrum of the predictor well and show comparable performance in all three metrics. However, the AT-based model has higher correlation and lower RMSE than the other two methods since its redundant decomposition allows the preservation of high-frequency information even when a longer wavelet filter is adopted. Therefore, the introduction of AT-based variance transformation is valuable when wider wavelet filters are required and are more likely to
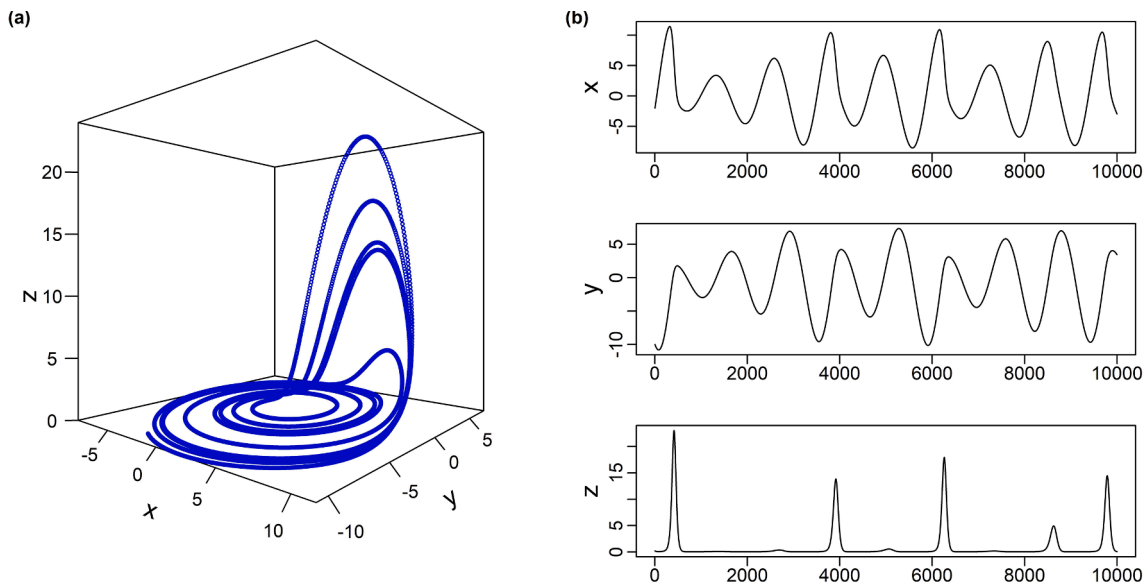

Fig. 3. Example of the Rössler system. **(a)** Phases space plot of the Rössler system corresponding to $a = 0.2$, $b = 0.2$, and $c = 5.7$; **(b)** Time domain plot of the dependent variable set (x_t, y_t, z_t).
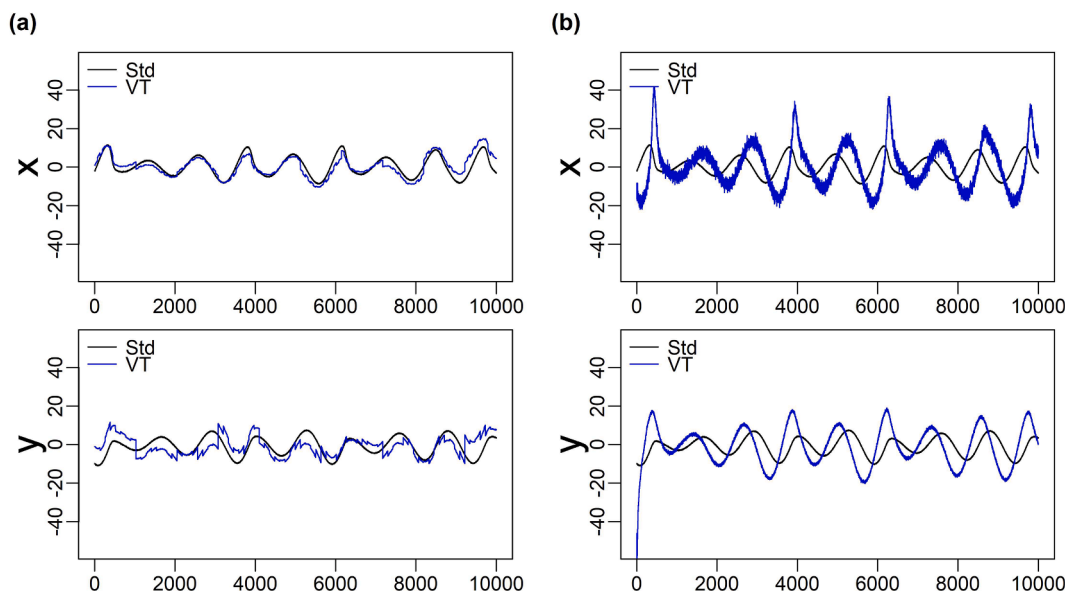
Fig. 4. Example of original (black line) and transformed (blue line) predictors using Haar: (a) DWT-MRA (b) MODWT/AT. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
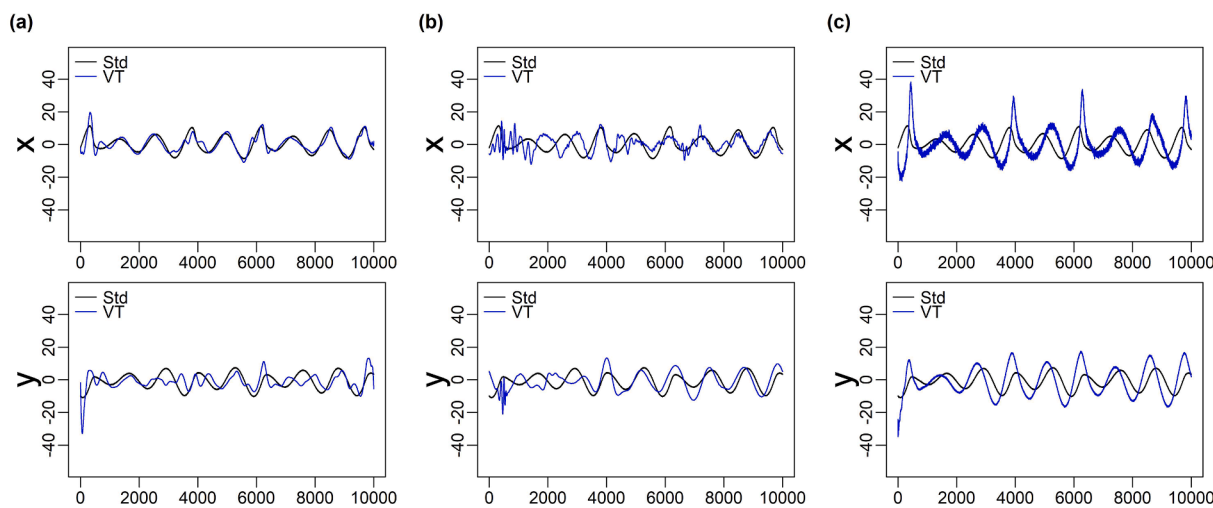

Fig. 5. Example of original (black line) and transformed (blue line) predictors using db8: (a) DWT-MRA (b) MODWT (c) AT. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

characterize the process of interest well, for example, an instantaneous process with polynomials of high order coefficients, such as a constant, linear, quadratic components, etc (Maheswaran and Khosa, 2012).

In closing, the above results investigate the model performance across different wavelet-based variance transformation approaches. Their limitations and advantages over each other have been discussed in detail. A comparison against the traditional wavelet forecasting approaches (i.e., direct and multi-component approaches, which use decomposed frequency components as predictors) is included in Text S 1 of the Supporting Information.

4. Variance transformation framework

In this section, we first present the variance transformation technique as a generalized model specification framework that uses a stepwise approach to implement the transformation. Secondly, we compare and highlight the difference between the stepwise transformation and direct transformation using a synthetic example generated from an autoregressive model with multiple predictor variables.

4.1. Stepwise variance transformation

The variance transformation technique is designed for regression problems where the aim is to better match the spectrum of the predictors with the response variable. Thus, the transformation of predictor variables can be based on the original response with each predictor variable individually. An alternative approach in the case of multiple predictors is to consider the transformation of a particular predictor based on the residual information in the response given existing predictor variables that have already been selected for use in the model. A stepwise selection logic calculating partial informational correlation (PIC) based on mutual information in information theory was introduced by Sharma (2000) and Sharma and Mehrotra (2014), and it is now combined with the variance transformation, resulting in stepwise variance transformation. An overall flowchart of the proposed framework is provided in Fig. 8. This framework has been implemented in an R library named WAvelet System Prediction (WASP), and more details can be found in Jiang, Rashid, et al. (2020).

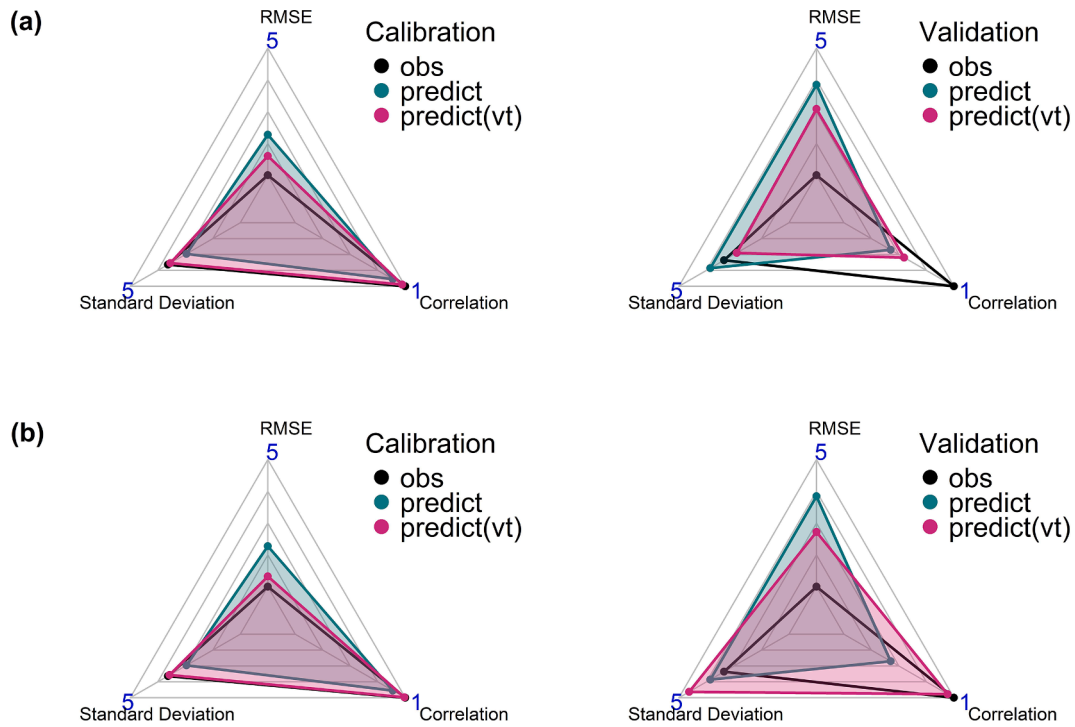The first step is to transform each predictor variable corresponding to

Fig. 6. Comparison of model performance using original and transformed predictors with the Haar wavelet filter: (a) DWT-based model; (b) MODWT/AT-based model.

the response individually. Second, the first significant predictor (i.e., the first-order predictor) is identified, which is the predictor with the strongest dependence between the transformed predictor and response using PIC. Next, residual information in the response and predictors is calculated, conditioned on the pre-identified predictors, and then higher-order predictor variables are obtained by transforming their residuals in predictors with respect to the residual in the response. This process is repeated until a pre-specified number of predictors have been selected or until the remaining residuals of predictors have no further dependence with the response residual. Although there are many options to calculate the residual in the response of interest, residuals here are computed by using knn leave-one-out cross-validation. To help illustrate this process, the stepwise variance transformation process can be formulated as Equation (9) as follows:

$$X^{'} = g(X|Z, Y|Z) \tag{9}$$

where $g(\cdot)$ denotes the variance transformation operation represented in the previous section and Z represents the pre-existing predictor(s). It is clear that when Z is empty, the transformed predictor variable is the first-order predictor, while when Z is not empty, it is a higher-order predictor variable. The above processes summarize the stepwise variance transformation.

Table 3 shows an illustration of both variance transformation (VT) and stepwise variance transformation (SVT) approaches. We use three predictor variables here, but there is no mathematical limit to the number of predictor variables. Clearly, the proposed two approaches share the same first-order predictor but otherwise, they are different. Although the same order of X_1, X_2, and X_3 is shown in the table, the actual order that the predictors are selected is likely to be different for the higher-order predictors. This is because the SVT method transforms the residual information conditioned on the previously identified predictors such that the transformed predictors are different from the VT method. An example of the first- and second-order predictors using the real-world example can be found in Section 5.

4.2. Application to synthetic examples

The autoregressive model of order nine (AR9) is adopted here to assess the performance of VT and SVT, respectively. The equation of the AR9 model is given by Sharma (2000):

$$x_t = 0.3x_{t-1} - 0.6x_{t-4} - 0.5x_{t-9} + \varepsilon_t \tag{10}$$

where ε is random Gaussian noise with zero mean and unit standard deviation. For each dataset, x_t was arbitrarily initialized and a total of $N + 500$ data points were generated, and the first 500 points were discarded to reduce the effects of arbitrary initializations. Nine candidate inputs, $x_{t-1}, x_{t-2}, \ldots,$ and x_{t-9}, were generated with only three of these, $x_{t-1}, x_{t-4},$ and x_{t-9}, needed in the AR9 model. A total number N of 1024 samples are generated, and the first half of the data is used for calibration while the second half is used for validation. Similarly, transformation is applied after the data partition and the knn model is used as the predictive model. The dyadic sample size is used to remove any boundary effects. DWT was used as the basis of wavelet transform for both VT and SVT, with the Haar wavelet filter adopted. Under this controlled environment, we can isolate the impacts of the stepwise transformation logic.

Three models were run, one with original predictors (referred to as standard approach, Std), one with VT predictors, and finally a model with the SVT predictors. Table 4 presents the results of predictor selection in terms of the number of times a predictor is identified out of 100 realizations and the total number of predictors identified using original and transformed predictors. The true predictors of AR9, x_{t-1}, x_{t-4}, and x_{t-9}, are identified in 100% of the realizations using all models. However, many additional predictor variables are selected using the VT approach and a handful of additional predictor variables are identified by the Std approach, while SVT performs between the two. Because both VT and SVT force predictors to be more similar to the associated response, this leads to higher dependence between the predictors and response, and thus more significant predictors are identified. However, the SVT approach tends to select fewer predictors than VT as predictor variables are transformed based on the residual information in the
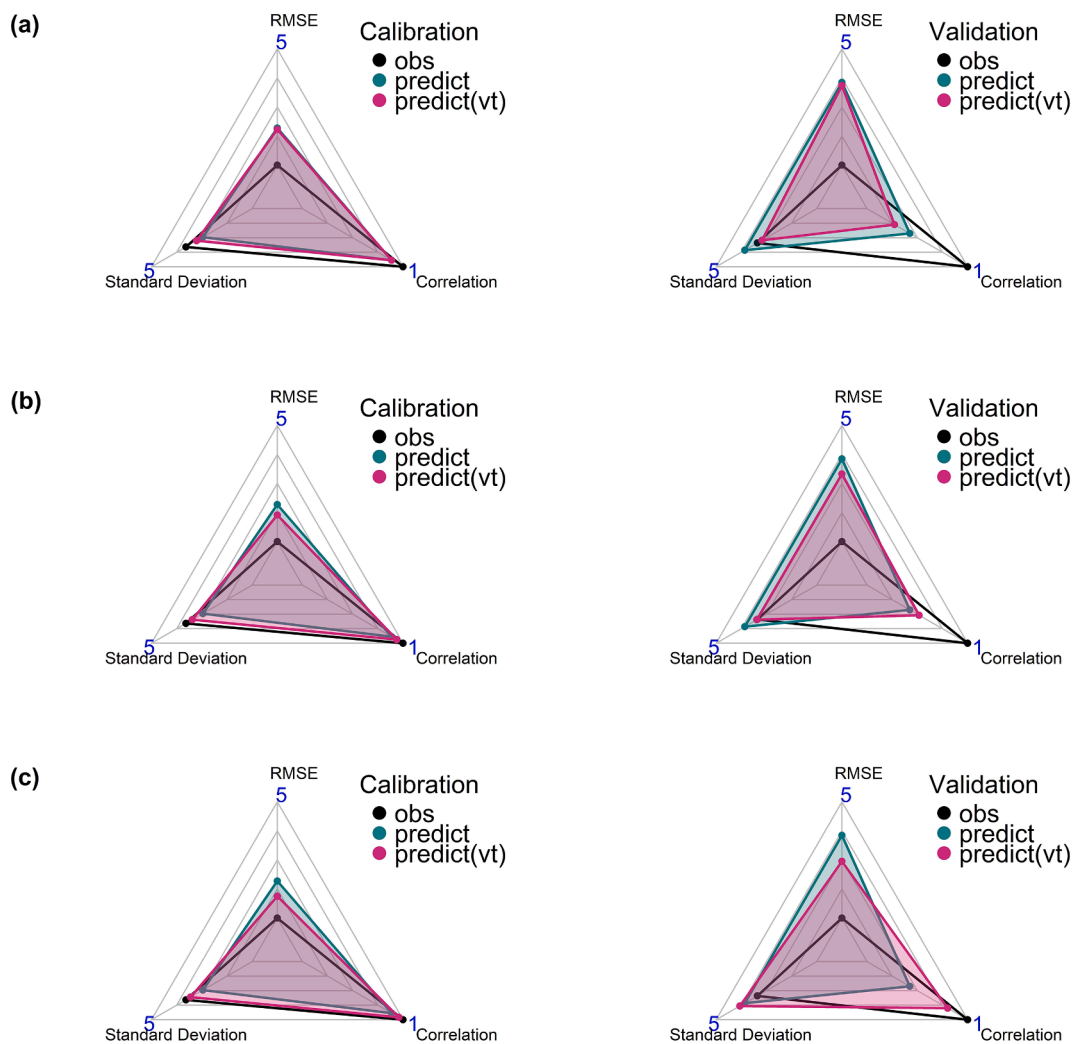
Fig. 7. Comparison of model performance using original and transformed predictors with the db8 wavelet filter: (a) DWT-based model; (b) MODWT-based model; (c) AT-based model.

response, meaning that fewer predictors are needed to characterize the response. Table S 1 presents the general agreement that exists between the PIC scores between calibration and validation (averaged across predictors and synthetic realizations), indicating the appropriate goodness of fit achieved in the process of defining the prediction model.

In addition, SVT has the best performance as measured by RMSE among the three models in both calibration and validation (Fig. 9). Based on the one-sided, two-sample Kolmogorov–Smirnov test, the RMSE values from the SVT model are significantly smaller than that of Std and VT, with *p*-values of 0.0004 and 0.001, respectively. It is worth noting that in terms of prediction accuracy there are only incremental improvements over Std using the proposed VT technique. This is mainly due to the nature of the AR9 system, where the spectral properties of the predictor and response are quite similar since the predictors in the system are lagged values of the response (Jiang, Sharma, et al., 2020). The SVT approach not only shows further improved prediction accuracy but also has advantages over the VT approach in terms of true predictor identification. These advantages of SVT will be further demonstrated using a real-world example in the next section.

5. Application to forecasting ENSO

Long-lead ENSO forecasting over the Niño3.4 region is used as a real-world example in this study. We adopted wind stress and temperature variables as regression predictors as they have been used in a range of

ENSO forecasting works (Dijkstra et al., 2019; Petrova et al., 2019). We have adopted the same regression predictors over three different regions in the equatorial area as given by Petrova et al. (2019), and details of these variables are given in Table S 2. The monthly time series of these variables are the averaged values over the three regions. The sea surface temperature datasets used for the predictors are NOAA ERSST-V4. The zonal wind stress is derived from ICOADS data provided by the NOAA/OAR/ESRL PSL, Boulder, Colorado, USA. The subsurface temperature dataset used for the subsurface ocean predictors is the Subsurface Temperature and Salinity Analyses dataset by Ishii et al. (2005), and subsurface temperature variables are extracted at depths of 50, 100, 150, 200, 250, 300, 400, and 500 m. The monthly anomalies of Niño3.4 are derived from monthly sea surface temperature values of Hadley Centre Global Ice and Sea Surface Temperature (HadISST) datasets (Rayner et al., 2003). In total, the entire dataset includes the target response (i.e., Niño3.4) and 30 predictor variables (i.e., wind stress and temperature variables at a range of depths for each of the three regions). The data is partitioned into two subsets: training period 1960–93 and testing period 1994–2012, This temporal division is due to the improved data quality and coverage of ocean subsurface variables by Tropical Ocean-Global Atmosphere Program (TOGA) in 1994 (Petrova et al., 2019). This is a retrospective experiment and the forecast skills have been assessed using both correlation and RMSE for a range of lead times up to 24 months.

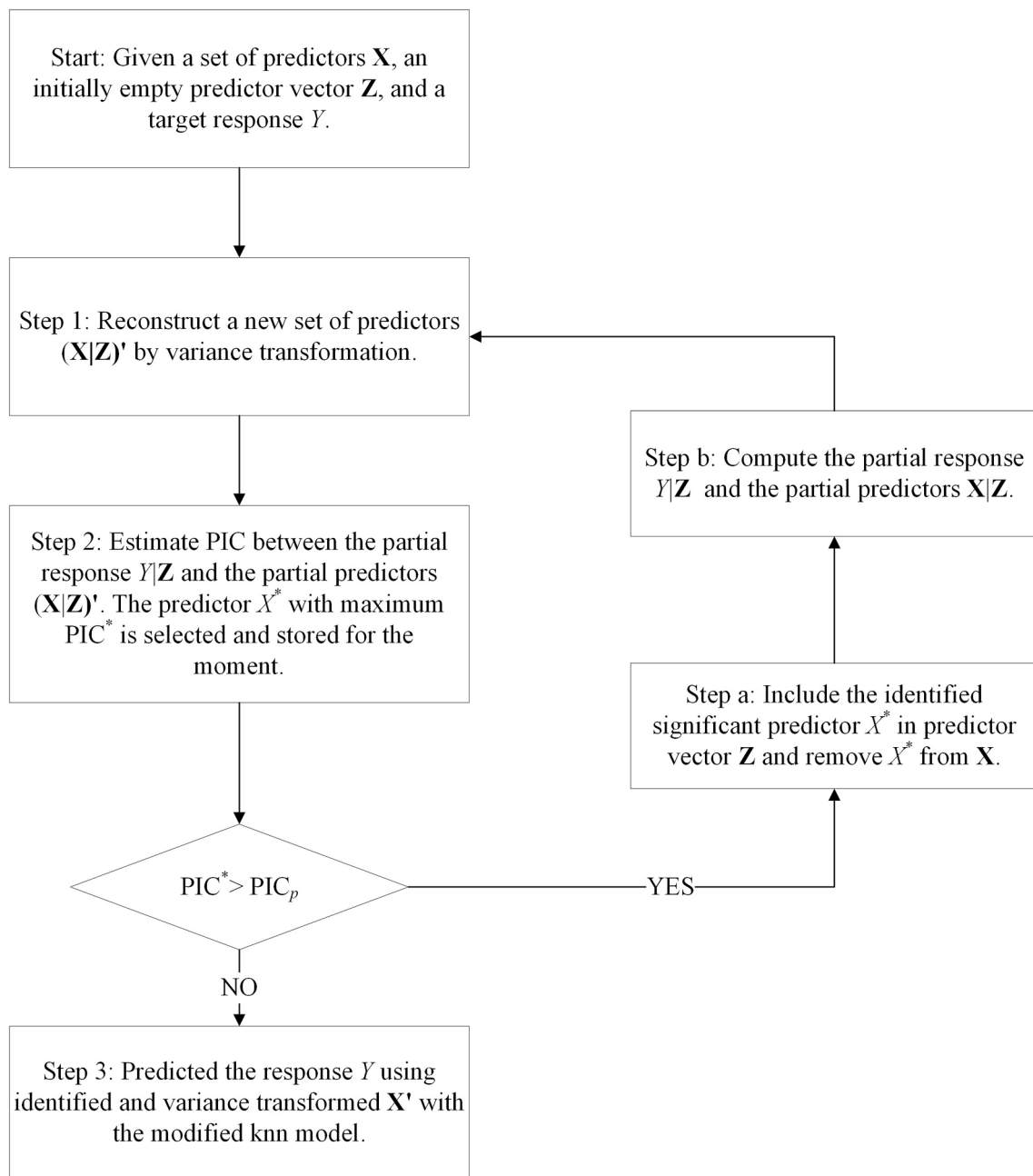The dynamical linear model (DLM) is used for forecasting in this

Fig. 8. Flowchart of the stepwise variance transformation method.

Table 3

Illustration of variance transformation and stepwise variance transformation.

	Predictor	Response	Transformed Predictor
Variance Transformation (VT)	X_1	Y	$X'_1 = g(X_1, Y)$
	X_2	Y	$X'_2 = g(X_2, Y)$
	X_3	Y	$X'_3 = g(X_3, Y)$
Stepwise Variance Transformation (SVT)	X_1	Y	$X'_1 = g(X_1, Y)$
	$X_2\|X'_1$	$Y\|X'_1$	$X'_2 = g(X_2\|X'_1, Y\|X'_1)$
	$X_3\|(X'_1, X'_2)$	$Y\|(X'_1, X'_2)$	$X'_3 = g(X_3\|(X'_1,X'_2),Y\|(X'_1,X'_2))$

Table 4

Frequency of predictor selection and the number of predictors selected using Std, VT, and SVT approaches.

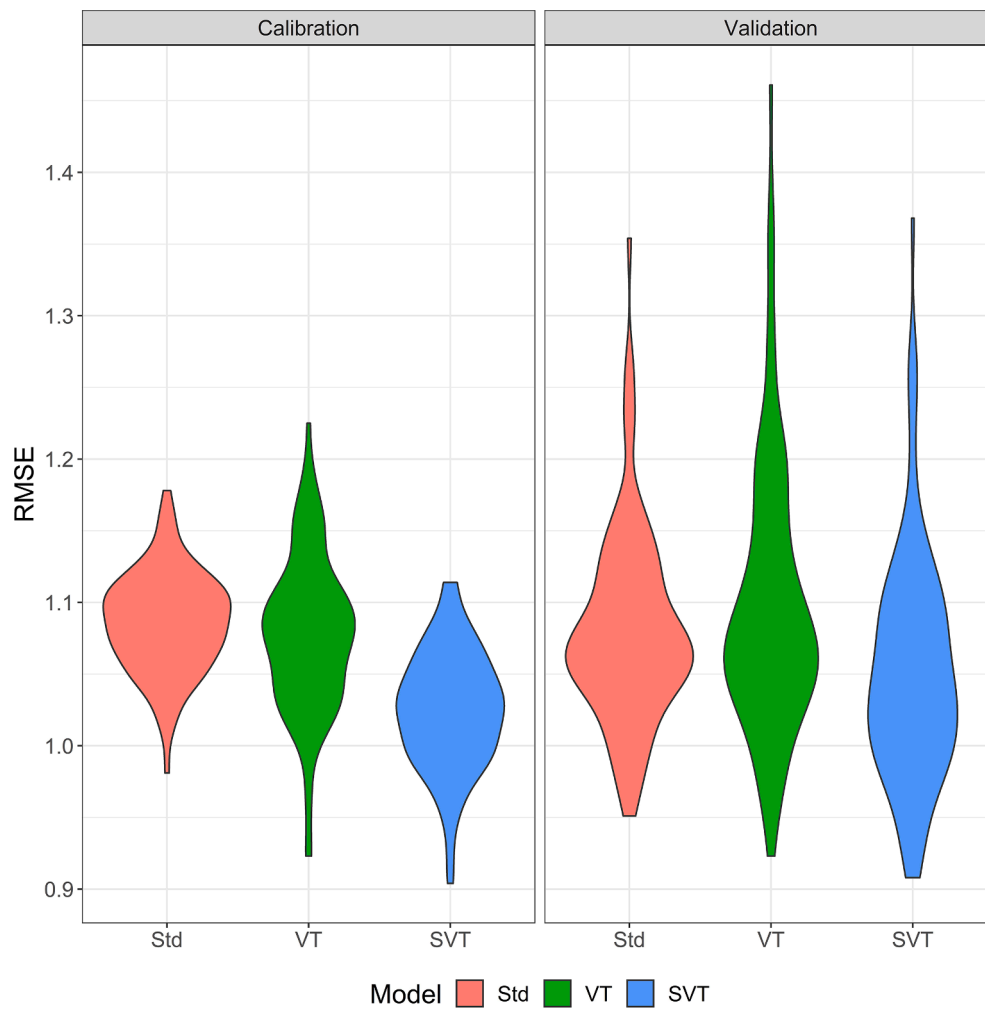| Percentage of Samples where Predictor x_{t-i} is selected				Percentage of Samples Total Number of Predictors Identified			
Predictor	Std	VT	SVT	Number of Predictors	Std	VT	SVT
1	100	100	100	1	0	0	0
2	1	51	62	2	0	0	0
3	2	31	6	3	92	1	32
4	100	100	100	4	8	15	68
5	2	84	0	5	0	45	0
6	1	60	0	6	0	27	0
7	2	10	0	7	0	10	0
8	0	0	0	8	0	2	0
9	100	100	100	9	0	0	0

Fig. 9. Comparison of prediction accuracy between Std, VT, and SVT approaches.

study, and concurrent predictor variables are adopted as covariates in the model, and it can be written as (Petris, 2010):

$$
\begin{aligned}
y_t &= F_t\theta_t + v_t, \quad v_t \sim N(0, V_t), \\
\theta_t &= G_t\theta_{t-1} + w_t, \quad w_t \sim N(0, W_t).
\end{aligned}
\tag{11}
$$

where θ_t is the unobserved state vector while y_t is the observed data that is the target response here. F_t, G_t, V_t, and W_t are real matrices of the appropriate dimension according to y_t and θ_t. Equation (11) can also be cast into a component form, including trend, seasonal, and covariate (i. e., predictor variable) components. The DLM model can be easily defined within the framework of the dlm R package (Petris, 2010), and the time-invariant DLM with three components (i.e., trend, seasonal, and covariate components) was used as the predictive model here. The forecasting model used here is a general forecasting model, which is different from the model proposed by Petrova et al. (2019) specifically designed for Niño prediction based on the ENSO dynamics. However,

the merit of models using transformed predictor variables will be demonstrated as long as the same baseline model is adopted. In addition, the transformed predictor variables are obtained using MODWT/AT-based variance transformation, and therefore no future information is required for forecasting. The Haar wavelet filter is used here so MODWT and AT are equivalent to each other.

Significant predictor variables are identified using the original and transformed predictor variables, respectively. With these identified significant drivers, the target response Niño3.4 is predicted using the DLM model. Table 5 summarizes the predictor selection using Std, VT, and SVT approaches. Two common predictor variables have been identified as the significant variables by three approaches, and transformed predictors associated with their original time series are given in Fig. 10. First, the two identified predictors, zonal wind stress and subsurface temperature, are transformed by filtering out irrelevant information with both VT and SVT approaches. It is apparent that transformed predictor variables are smoother particularly for the predictor variable of zonal wind stress. Second, VT and SVT share the same first-order predictor variable, and the second predictor from the VT model was also selected as the second-order predictor in the SVT method. However, the second-order predictor by SVT is obtained from transformed residual information conditioned on the pre-existing predictor variable (i.e., the first-order predictor) thus it has a completely different range and variability compared to the predictor used in the VT and Std models. Transformed variables have same trend but with larger variance than untransformed variable, and VT transformed temperature variable has larger variance than SVT transformed variable.
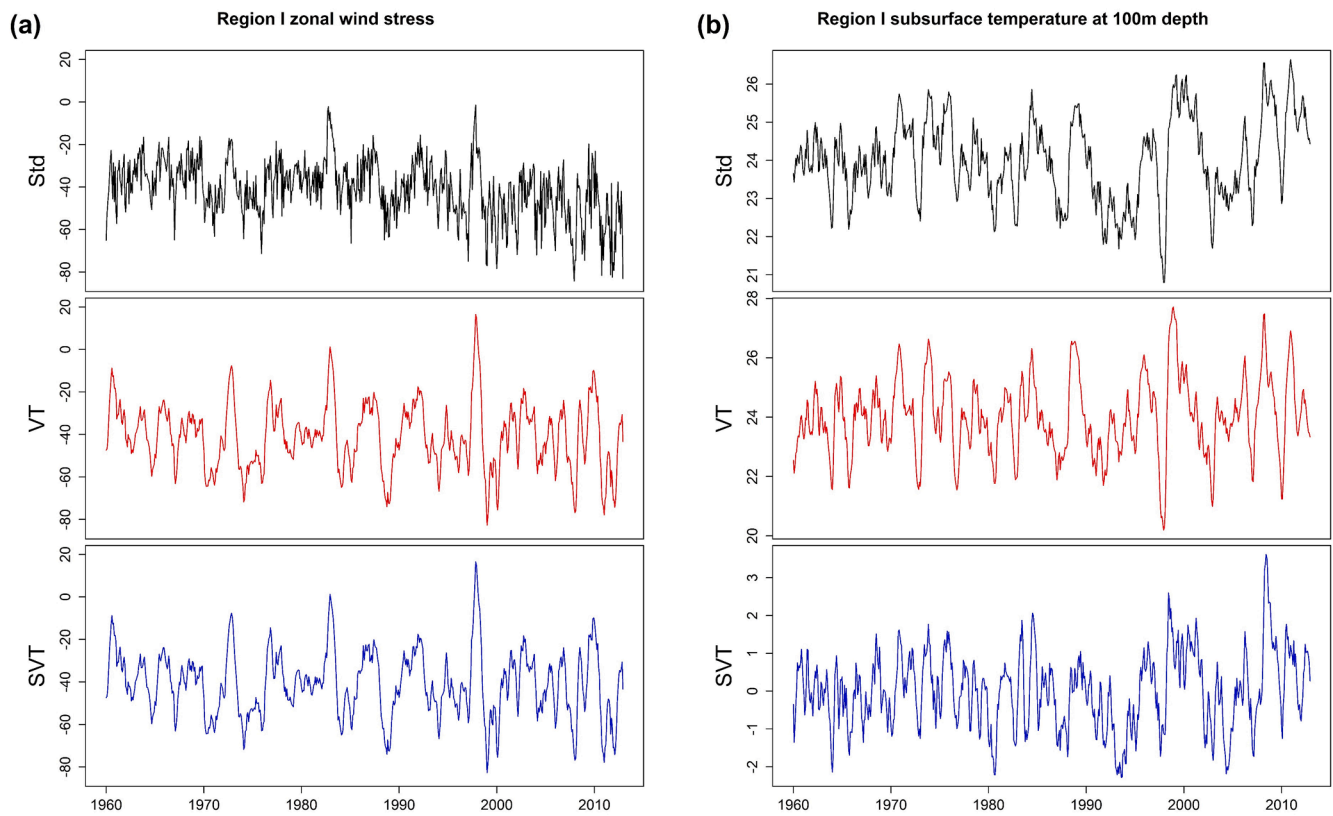
Table 5
Identified significant predictor variables.

No.	Std	VT	SVT
1	Region I subsurface temperature at 100 m depth	Region I Zonal wind stress	Region I Zonal wind stress
2	Region I Zonal wind stress	Region I subsurface temperature at 100 m depth	Region I subsurface temperature at 100 m depth

Fig. 10. Transformed predictor variables and associated original time series: (a) the first-order predictor (b) the second-order predictor.
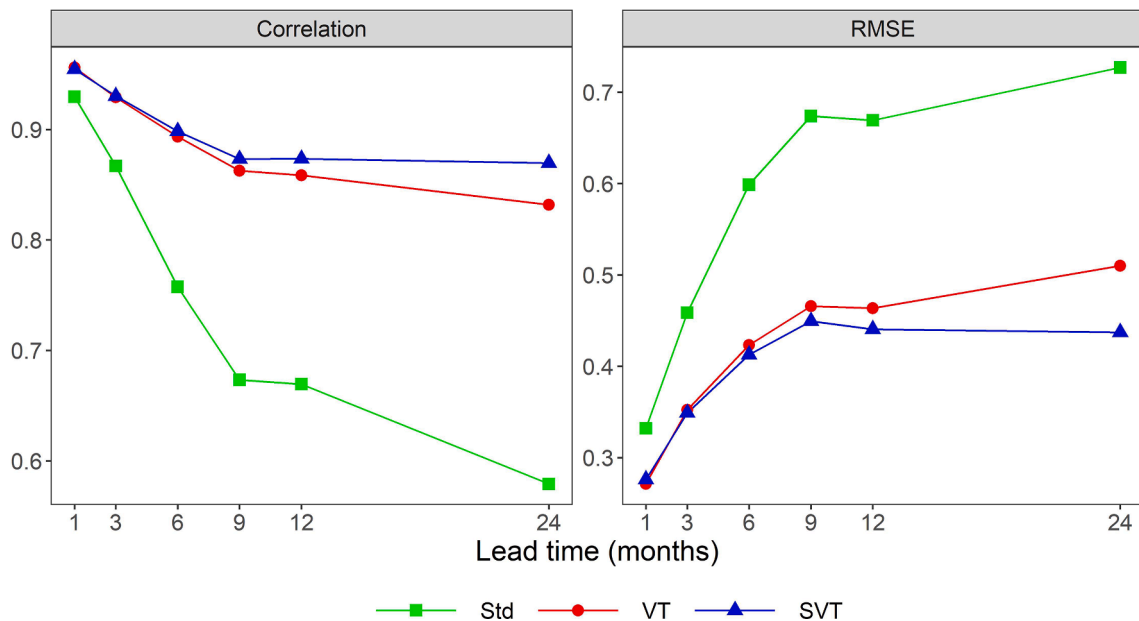

Fig. 11. Forecast skills of three models by correlation and RMSE against varying lead time, respectively.

Fig. 11 presents the predictive accuracy including both RMSE and correlation against lead time (*l*). Given the same predictor variables, the model using transformed predictor variables has better performance for both metrics. More importantly, the comparison between VT and SVT approach demonstrates the capability of the stepwise variance transformation by transforming the predictor variable according to the residual in response conditioned on identified or pre-existing predictor variables. As a result, the SVT predictor variables can characterize the

response better even with the same identified predictor variables. Better forecast skills than what Petrova et al. (2019) have shown are observed which is likely due to the concurrent predictor variables up to the current time of *t* are used in this study instead of lagged predictor variables at times of *t*-1, *t*-2, …, and *t*-*l* adopted in their work. The objective of this study is to demonstrate the improvement in the predictability of Niño3.4 by using transformed predictor variables given the same forecasting model, and this has been clearly shown in the example. Another

important feature of the proposed method is that the forecast skills deteriorate more slowly with increasing lead time than the reference model. This again shows the value of the variance transformation technique as it can preserve important features of the predictors across longer timescales. The stepwise approach shows consistent higher skills and slower deteriorations across lead times than the Std and VT approaches.

6. Summary and conclusions

In many natural systems particularly in the field of hydro-climatology, it is well acknowledged that long-term oscillations or low-frequency variabilities exist, and this forms the motivation for improving predictive capability using the proposed variance transformation technique. Our proposed approach investigates the frequency domain wherein the original time series is transformed with a new variance structure across its spectrum with respect to a chosen response. The introduction of AT as the wavelet transformation alternative provides an additional choice of characterizing the spectral domain and provides advantages over other wavelet alternatives when a longer wavelet filter is required. This, especially, has an impact on the prediction of instantaneous processes (e.g., floods, storms, and bushfires) in hydro-climatology, an important issue with wide implications for the discipline and the communities that are affected. Another contribution of the work is the proposed stepwise variance transformation framework, which has advantages over the direct variance transformation approach in terms of both selection and prediction accuracy especially for complex natural systems where multiple predictor variables are needed. Applications to both synthetic and real examples demonstrate the merits of the wavelet-based transformation technique and the capability of stepwise variance transformation.

There are unique aspects of the proposed variance transformation technique no matter which wavelet transform is used as the underlying basis. The first feature that must be noted is that the method identifies a unique variance transformation for each decomposed sub-time series, the transformation being shown to result in optimal predictive accuracy with respect to a chosen response. The second feature is that the proposed method can identify additional meaningful drivers using the transformed predictor variables given the need for predictor selection (as is the case presented in our real example). The third feature is that the proposed method is a generic alternative potentially applicable in modelling any natural system. Theoretically, the DWT-based method leads to the best performance of proposed alternatives since it always fulfills the requirements of both additive and variance decomposition as long as orthogonal wavelets are used. However, the MODWT- and AT-based alternatives provide values in certain systems as shown in this study. Furthermore, these extensions have allowed applications in a forecasting setting, and the potential usage in long-lead time forecasts, such as the ENSO example given here, is very promising. Climate model simulations from decadal prediction experiments (Choudhury et al., 2019) could be another scenario where these extensions can be applied for forecasting applications. More conventional downscaling (Mehrotra and Sharma, 2006) or conditional generation alternatives that use exogenous covariates (Roderick et al. 2019; Wasko and Sharma, 2017) generated using climate models, are also likely to be benefitted. While their performance largely depends on how well they can characterize the spectrum of variables of interest, it is preferable to use the DWT-based method when predictor variables can be simulated into the future. The emerging problem such as the assessment of climate change to the hydro-climatological system is a perfect example where climate model simulations are used to study impacts across affected various sectors (e.g., renewable resources, water resources, agriculture, etc.).

In closing, it is worth noting that we have implemented the DWT-, MODWT-, and AT-based variance transformation and stepwise variance transformation in the R-package WASP (WAvelet System Prediction). WASP is an open-source tool with sufficient help-files and can be downloaded from the Web site at http://www.hydrology.unsw.edu.au/software/WASP. A practical application of the software is given by Jiang, Rashid, et al. (2020).

CRediT authorship contribution statement

Ze Jiang: Conceptualization, Methodology, Software, Validation, Formal analysis, Writing – original draft, Writing - review & editing, Visualization. **Ashish Sharma:** Conceptualization, Methodology, Validation, Resources, Writing - review & editing, Visualization, Supervision, Project administration. **Fiona Johnson:** Conceptualization, Methodology, Validation, Resources, Writing - review & editing, Visualization, Supervision, Project administration.

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.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jhydrol.2021.126816.

References

Aussem, A., Campbell, J., Murtagh, F., 1998. Wavelet-based feature extraction and decomposition strategies for financial forecasting. J. Comput. Intell. Finan. 6 (2), 5–12.

Beygelzimer, A., Kakade, S., Langford, J., 2006. Cover trees for nearest neighbor. Proceedings of the 23rd international conference on Machine learning, pp. 97–104.

Box, G.E.P., Cox, D.R., 1964. An analysis of transformations. J. Roy. Stat. Soc.: Ser. B (Methodol.) 26 (2), 211–243. https://doi.org/10.1111/j.2517-6161.1964.tb00553.x.

Brunner, M.I., Bárdossy, A., Furrer, R., 2019. Technical note: stochastic simulation of streamflow time series using phase randomization. Hydrol. Earth Syst. Sci. 23 (8), 3175–3187. https://doi.org/10.5194/hess-23-3175-2019.

Chavez, M., Cazelles, B., 2019. Detecting dynamic spatial correlation patterns with generalized wavelet coherence and non-stationary surrogate data. Sci. Rep. 9 (1), 7389. https://doi.org/10.1038/s41598-019-43571-2.

Choudhury, D., Mehrotra, R., Sharma, A., Sen Gupta, A., Sivakumar, B., 2019. Effectiveness of CMIP5 decadal experiments for interannual rainfall prediction over Australia. Water Resour. Res. 55 (8), 7400–7418. https://doi.org/10.1029/2018WR024462.

D'Arrigo, R., Cook, E.R., Wilson, R.J., Allan, R., Mann, M.E., 2005. On the variability of ENSO over the past six centuries. Geophys. Res. Lett. 32 (3) https://doi.org/10.1029/2004GL022055.

Dijkstra, H.A., Petersik, P., Hernández-García, E., López, C., 2019. The application of machine learning techniques to improve El Niño prediction Skill. Front. Phys. 7 (153) https://doi.org/10.3389/fphy.2019.00153.

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. https://doi.org/10.1016/j.jhydrol.2017.06.019.

Dutilleux, P., 1990. An implementation of the "algorithme à trous" to compute the wavelet transform. In: Wavelets. Springer, pp. 298–304.

Fowler, J.E., 2005. The redundant discrete wavelet transform and additive noise. IEEE Signal Process Lett. 12 (9), 629–632. https://doi.org/10.1109/Lsp.2005.853048.

Gupta, H.V., Kling, H., Yilmaz, K.K., Martinez, G.F., 2009. Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. J. Hydrol. 377 (1), 80–91. https://doi.org/10.1016/j.jhydrol.2009.08.003.

Harrington, H.A., Van Gorder, R.A., 2017. Reduction of dimension for nonlinear dynamical systems. Nonlinear Dyn 88 (1), 715–734. https://doi.org/10.1007/s11071-016-3272-5.

Helsel, D.R., Hirsch, R.M., 2002. Statistical methods in water resources (04-A3). Retrieved from Reston, VA: http://pubs.er.usgs.gov/publication/twri04A3.

Holschneider, M., Kronland-Martinet, R., Morlet, J., Tchamitchian, P., 1990. A real-time algorithm for signal analysis with the help of the wavelet transform. In: Wavelets. Springer, pp. 286–297.

Hyndman, R.J., Athanasopoulos, G., 2014. Forecasting: Principles and Practice. Heathmont, Vic. OTexts.

Ishii, M., Shouji, A., Sugimoto, S., Matsumoto, T., 2005. Objective analyses of sea-surface temperature and marine meteorological variables for the 20th century using ICOADS and the Kobe Collection. Int. J. Climatol. 25 (7), 865–879. https://doi.org/10.1002/joc.1169.

Jiang, Z., Rashid, M.M., Johnson, F., Sharma, A., 2020a. A wavelet-based tool to modulate variance in predictors: An application to predicting drought anomalies. Environ. Modell. Software 135, 104907. https://doi.org/10.1016/j.envsoft.2020.104907.

Jiang, Z., Sharma, A., Johnson, F., 2020b. Refining Predictor Spectral Representation Using Wavelet Theory for Improved Natural System Modeling. Water Resour. Res. 56 (3) https://doi.org/10.1029/2019WR026962.

Kişi, Ö., 2011. A combined generalized regression neural network wavelet model for monthly streamflow prediction. KSCE J. Civ. Eng. 15 (8), 1469–1479. https://doi.org/10.1007/s12205-011-1004-4.

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

McInerney, D., Thyer, M., Kavetski, D., Lerat, J., Kuczera, G., 2017. Improving probabilistic prediction of daily streamflow by identifying Pareto optimal approaches for modeling heteroscedastic residual errors. Water Resour. Res. 53 (3), 2199–2239. https://doi.org/10.1002/2016wr019168.

Mehrotra, R., Sharma, A., 2006. A nonparametric stochastic downscaling framework for daily rainfall at multiple locations. J. Geophys. Res. Atmos. 111 (D15). https://doi.org/10.1029/2005JD006637.

Mosteller, F., Tukey, J.W., 1977. Data analysis and regression: a second course in statistics. Addison-Wesley Pub. Co, Reading, Mass.

Nason, G.P., Von Sachs, R., 1999. Wavelets in time-series analysis. Philos. Trans. R. Soc. A Mathemat. Phys. Eng. Sci. 357 (1760), 2511–2526. https://doi.org/10.1098/rsta.1999.0445.

Nguyen, H.T., Nabney, I.T., 2010. Short-term electricity demand and gas price forecasts using wavelet transforms and adaptive models. Energy 35 (9), 3674–3685. https://doi.org/10.1016/j.energy.2010.05.013.

Nourani, V., Hosseini Baghanam, A., Adamowski, J., Kisi, O., 2014. Applications of hybrid wavelet-Artificial Intelligence models in hydrology: a review. J. Hydrol. 514, 358–377. https://doi.org/10.1016/j.jhydrol.2014.03.057.

Percival, D.B., Walden, A.T., 2000. Wavelet methods for time series analysis. Cambridge University Press, Cambridge.

Petris, G., 2010. An R Package for Dynamic Linear Models. J. Stat. Softw. 36 (12), 1–16. https://doi.org/10.18637/jss.v036.i12.

Petrova, D., Ballester, J., Koopman, S.J., Rodó, X., 2019. Multiyear statistical prediction of ENSO enhanced by the tropical pacific observing system. J. Clim. 33 (1), 163–174. https://doi.org/10.1175/jcli-d-18-0877.1.

Pui, A., Sharma, A., Santoso, A., Westra, S., 2012. Impact of the El Niño–Southern oscillation, indian ocean dipole, and southern annular mode on daily to subdaily rainfall characteristics in East Australia. Mon. Weather Rev. 140 (5), 1665–1682. https://doi.org/10.1175/MWR-D-11-00238.1.

Quilty, J., Adamowski, J., 2018. 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. J. Hydrol. 563, 336–353. https://doi.org/10.1016/j.jhydrol.2018.05.003.

Rashid, M.M., Johnson, F., Sharma, A., 2018. Identifying sustained drought anomalies in hydrological records: a wavelet approach. J. Geophys. Res. Atmos. 123 (14), 7416–7432. https://doi.org/10.1029/2018jd028455.

Rathinasamy, M., Khosa, R., Adamowski, J., Ch, S., Partheepan, G., Anand, J., Narsimlu, B., 2014. Wavelet-based multiscale performance analysis: an approach to assess and improve hydrological models. Water Resour. Res. 50 (12), 9721–9737. https://doi.org/10.1002/2013wr014650.

Rayner, N., Parker, D.E., Horton, E., Folland, C.K., Alexander, L.V., Rowell, D., Kaplan, A., 2003. Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. J. Geophys. Res. Atmos. 108 (D14) https://doi.org/10.1029/2002jd002670.

Roderick, T.P., Wasko, C., Sharma, A., 2019. Atmospheric moisture measurements explain increases in tropical rainfall extremes. Geophys. Res. Lett. 46 (3), 1375–1382. https://doi.org/10.1029/2018gl080833.

Rössler, O.E., 1976. An equation for continuous chaos. Phys. Lett. A 57 (5), 397–398.

Sang, Y.F., 2013. A review on the applications of wavelet transform in hydrology time series analysis. Atmos. Res. 122, 8–15. https://doi.org/10.1016/j.atmosres.2012.11.003.

Schreiber, T., Schmitz, A., 2000. Surrogate time series. Physica D 142 (3–4), 346–382. https://doi.org/10.1016/S0167-2789(00)00043-9.

Shafaei, M., Kisi, O., 2016. Lake level forecasting using wavelet-SVR, wavelet-ANFIS and wavelet-ARMA conjunction models. Water Resour. Manage. 30 (1), 79–97. https://doi.org/10.1007/s11269-015-1147-z.

Sharma, A., 2000. Seasonal to interannual rainfall probabilistic forecasts for improved water supply management: Part 1 — A strategy for system predictor identification. J. Hydrol. 239 (1), 232–239. https://doi.org/10.1016/S0022-1694(00)00346-2.

Sharma, A., Mehrotra, R., 2014. An information theoretic alternative to model a natural system using observational information alone. Water Resour. Res. 50 (1), 650–660. https://doi.org/10.1002/2013wr013845.

Shensa, M.J., 1992. The discrete wavelet transform: wedding the a trous and Mallat algorithms. IEEE Trans. Signal Process. 40 (10), 2464–2482. https://doi.org/10.1109/78.157290.

Strogatz, S.H., 2000. Nonlinear dynamics and chaos: with applications to physics, biology, chemistry, and engineering, 1st pbk print. ed. Westview Press, Cambridge, MA.

Torrence, C., Compo, G.P., 1998. A practical guide to wavelet analysis. Bull. Am. Meteorol. Soc. 79 (1), 61–78. https://doi.org/10.1175/1520-0477(1998)079<0061:Apgtwa>2.0.Co;2.

Walden, A.T., 2001. In: European Congress of Mathematics. Birkhäuser Basel, Basel, pp. 627–641. https://doi.org/10.1007/978-3-0348-8266-8_56.

Wang, Q.J., Shrestha, D.L., Robertson, D.E., Pokhrel, P., 2012. A log-sinh transformation for data normalization and variance stabilization. Water Resour. Res. 48 (5) https://doi.org/10.1029/2011wr010973.

Wasko, C., Sharma, A., 2017. Continuous rainfall generation for a warmer climate using observed temperature sensitivities. J. Hydrol. 544, 575–590. https://doi.org/10.1016/j.jhydrol.2016.12.002.

Westra, S., Sharma, A., 2010. An upper limit to seasonal rainfall predictability? J. Clim. 23 (12), 3332–3351. https://doi.org/10.1175/2010JCLI3212.1.

Wu, X., Marshall, L., Sharma, A., 2019. The influence of data transformations in simulating Total Suspended Solids using Bayesian inference. Environ. Modell. Software 121, 104493. https://doi.org/10.1016/j.envsoft.2019.104493.