Water Resources Research

TECHNICAL REPORTS: METHODS
10.1029/2019WR026962

Refining Predictor Spectral Representation Using Wavelet Theory for Improved Natural System Modeling

Ze Jiang[1], Ashish Sharma[1], and Fiona Johnson[1]

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

Key Points:
- A wavelet-based variance transformation is identified for each predictor variable that explains maximal information in the corresponding response
- Additional predictor variables are selected using a stepwise predictor selection rationale
- Results show consistent improvement in predictive accuracy in synthetic experiments from linear, nonlinear, and dynamic systems as well as the real-world example

Supporting Information:
- Supporting Information S1

Correspondence to:
A. Sharma,
a.sharma@unsw.edu.au

Citation:
Jiang, Z., Sharma, A., & Johnson, F. (2020). Refining predictor spectral representation using wavelet theory for improved natural system modeling. *Water Resources Research*, *56*, e2019WR026962. https://doi.org/10.1029/2019WR026962

Received 14 DEC 2019
Accepted 19 FEB 2020
Accepted article online 20 FEB 2020

Abstract Predicting future surpluses or shortages of water is a long-standing problem having considerable ramifications to water management across the world. Any prediction model for a natural system such as one that estimates water surpluses or shortages requires a two-step approach. These are the following: first, identify and select meaningful predictor variables from a large number of potential predictors and second, formulate an accurate, efficient, and robust predictive model between selected predictors and the response. Recognizing that the timescales at which a response may operate is usually different from that of the predictors being identified, we introduce here a wavelet-based unique variance transformation to each of the multiple predictor variables in the system which ensures an improved regression relationship to the modeled response. All existing methods assume no change in predictors even if they characterize variability at markedly different timescales, a deficiency that is addressed using the variance-transformed predictor which can explain maximal information in an associated response. Using this unique variance transformation, additional predictor variables can be selected by assessing their ability to characterize residual information in the response that accounts for the effect of preidentified predictors. We demonstrate the utility of the wavelet-based method using synthetically generated data sets from known linear and nonlinear systems with parametric and nonparametric predictive models. Applications to a dynamic system and a real-world example to downscale a drought indicator over the Sydney region confirm its utility in an applied setting.

1. Introduction

The underlying theme of most regression modeling is that a system response can be characterized using a finite set of predictor variables and a well-defined model. When the response is a natural process as is the case in hydrology, predictors need to be physically justified, with the model based on a strong conceptual understanding of the system. However, for many systems it is difficult to specify the conceptual model, particularly where the response of interest is not directly observed and temporally varying, for example, when considering the severity of drought in a catchment. In these cases, it is even more difficult to select the correct predictors. Prediction approaches range from simple regression models using a range of physically valid predictor variables (Hertig & Tramblay, 2017) to those where complex transformations, including rotations of the predictors, are used (Jiang et al., 2019; Ndehedehe et al., 2016; Vu et al., 2016). One problem is that although predictors may be physically justifiable, they vary at scales that differ substantially from those of the response. For example, daily precipitation is used to predict catchment streamflow, while the variability at seasonal or interannual timescales is of most interest for water managers. This discrepancy in timescale variability is amplified even further when considering groundwater systems.

Such differences in scale are difficult to accommodate in traditional regression modeling approaches. One approach to address this problem is to transfer the modeling problem to the frequency domain. Frequency domain techniques (using Fourier or Wavelet Transform) have been used to characterize and simulate variability in the response (Brunner et al., 2019), to attribute processes that may be causing variability in a certain frequency band (Johnson et al., 2011), and to formulate bias-correction alternatives for multiple variables that exhibit systematic differences in characteristics compared to the observed record (H. Nguyen et al., 2016, 2017, 2019). The current study aims to extend these approaches by developing a generic basis for regression modeling to characterize a response using a set of predictor variables that vary at markedly different timescales from the response and each other.

Our proposed approach stems from the assumption that each predictor can be transformed to exhibit similar spectral attributes as the response while ensuring that the predictability is maximized. This method

essentially reconstructs a predictor by transforming the variance structure of the predictor across the entire frequency domain. Specifically, we propose the following two hypotheses:

1. If the spectral variance structure of the predictor is similar to that of the response, the predictive model using that predictor will exhibit better accuracy than otherwise.
2. If the spectral variance structure of the residual information contained within the predictor and response variable conditioned to the preexisting predictor(s) is similar, the predictive model using that predictor will exhibit better accuracy than otherwise.

Characterizing the spectrum of a time series or a signal is often done using the Fourier transform. Here we adopt the wavelet transform (WT) instead due to the documented drawbacks of Fourier transform in losing time information when transforming to the frequency domain (Daubechies, 1990; Strang & Nguyen, 1996; Torrence & Compo, 1998). The WT is a powerful mathematical tool, which is able to decompose the original time series into separate large-scale (slowly changing, low frequency) and fine-scale (rapidly changing details, high frequency) subtime series. WT has been widely used in the field of hydrological and water resources forecasting (Fahimi et al., 2017; J. Quilty et al., 2019; Rashid et al., 2018; Sang, 2013). Recently, it has been shown that many forecasting applications using wavelets are incorrect (Du et al., 2017; John Quilty & Adamowski, 2018) because the most commonly used discrete wavelet transform (DWT) requires future information to forecast hydrological time series, limiting its use for real-world applications since the future is unknown. Because of this, our proposed approach focuses on models that predict systems retrospectively, for infilling gaps in the response variable record or for projecting derived hydrological variables by using, for example, modeled future projections of the predictors but where the future of the derived response is unknown. An example of the latter is presented in Rashid et al. (2018), in which a drought index can be assessed using predictor variables from climate model projections of the future. More importantly, the proposed method derived from the assumptions mentioned before is different from what has been currently used so far. The two most commonly used approaches are (H. T. Nguyen & Nabney, 2010; John Quilty & Adamowski, 2018) as follows: The first of these directly uses decomposed subtime series to predict the target response (Kişi, 2011; Rashid et al., 2018), known as the direct approach. The second uses decomposed subtime series to forecast the decomposed frequency components of the associated response and then aggregates the separate target series predictions (Rathinasamy et al., 2014; Shafaei & Kisi, 2016), the so-called multicomponent approach. In contrast, our method performs a theoretically derived variance transformation to each decomposed subtime series, the transformation being shown to result in optimal predictive accuracy with respect to a chosen response.

The remainder of the paper is organized as follows. Section 2 reviews the DWT given the need for a good understanding of the fundamentals of this method, introduces the variance transformation technique, and describes the algorithm. In section 3, the utility of the wavelet-based method outlined in section 2 is demonstrated using synthetically generated data sets from known linear and nonlinear systems. Section 4 applies the variance transformation technique to a dynamic data set followed by an application to a real-world example in section 5. Concluding remarks are given in section 6.

2. Background

2.1. DWT

As described previously, we adopt the DWT to characterize the spectrum. Background on relevant wavelet theory is presented here. Consider a time series $x(t)$ and assume that every vector in a vector space can be written as a linear combination of properly formulated basis vectors. DWT involves transforming $x(t)$ by identifying such linear combinations. The synthesis, or the reconstruction to the transformed space followed by the back-transformation to the original space, corresponds to computing an optimal linear combination equation (Halmos, 2017). This concept can be easily generalized to functions and the basis vectors in the wavelet theory expressed as $\psi(\eta)$(Kaiser, 2010) follows:

$$\psi(\eta) = \psi(\tau, s) = \frac{1}{\sqrt{s}}\psi\left(\frac{t-\tau}{s}\right) \tag{1}$$

where τ is the translation and s is scale (equivalent to the inverse of the frequency associated with a Fourier

transformation). The wavelets coefficients $W_x^{\psi}(\tau, s)$ are calculated by the inner product of two functions (Percival & Walden, 2000):

$$W_x^{\psi}(\tau, s) = \frac{1}{\sqrt{s}} \int_{-\infty}^{\infty} x(t) \cdot \psi^* \left(\frac{t-\tau}{s} \right) \mathrm{d}t \tag{2}$$

where star (*) denotes complex conjugate functions of $\psi(\eta)$. The wavelet coefficients measure the similarity of $x(t)$ to the wavelet ψ at the current scale s. With the basis functions and wavelet coefficients, the function $x(t)$ can be reconstructed by the summation of products of basis functions and associated wavelet coefficients. This inverse transform yields (Percival & Walden, 2000):

$$x(t) = \frac{1}{C_{\psi}^2} \int_s \int_{\tau} \left[W_x^{\psi}(\tau, s) \frac{1}{s^2} \psi \left(\frac{t-\tau}{s} \right) \right] \mathrm{d}\tau \cdot \mathrm{d}s \tag{3}$$

where C_{ψ} is a constant that depends on the specific wavelet $\psi(\eta)$ used. This summarizes the rationale behind the continuous wavelet transform. When applying this theory to a discrete data set case with the length of n given by $\mathbf{X} = [x_0, x_1, ..., x_{n-1}]^T$, the DWT analyzes the signal at different scales with different translations (often scales and translations are expressed in powers of two, so-called dyadic scales and translations) by decomposing the signal into wavelet and scaling coefficients. The process can be written in the form of matrix multiplication (Percival & Walden, 2000):

$$\mathbf{W} = \mathbf{w}\mathbf{X} \tag{4}$$

where $\boldsymbol{W}$ is a vector of DWT coefficients and $\boldsymbol{w}$ is the $n \times n$ orthonormal transform matrix (i.e., $\mathbf{w}^T\mathbf{w} = \mathbf{w}\mathbf{w}^T = \mathbf{I_n}$, where $\mathbf{I_n}$ is the $n \times n$ identity matrix and $\langle \mathbf{w}_j, \mathbf{w}_k \rangle = \begin{cases} 1, & when\ j = k \\ 0, & when\ j \neq k \end{cases}$). Due to this property of orthonormality, it implies that $\mathbf{w}^T\mathbf{W} = \mathbf{w}^T\mathbf{w}\mathbf{X} = \mathbf{X}$. Thus, the reconstruction of the original data set can be achieved by the synthesis equation (Aussem et al., 1998; Percival & Walden, 2000):

$$\mathbf{X} = \mathbf{w}^T\mathbf{W} = \sum_{j=1}^{J} \mathbf{w}_j^T \mathbf{D}_j + \mathbf{v}_J^T \mathbf{A}_J \tag{5}$$

where $\boldsymbol{W}$ is partitioned as $\mathbf{W} = [\mathbf{D}_1, ..., \mathbf{D}_J, \mathbf{A}_J]$, including detail coefficients ($\boldsymbol{D}_j$) and approximation coefficient(s) ($\boldsymbol{A}_J$); $\boldsymbol{w}$ can be partitioned into a similar structure as $\boldsymbol{W}$, which is $\mathbf{w} = [\mathbf{w}_1, ..., \mathbf{w}_J, \mathbf{v}_J]^T$ and $\boldsymbol{w}_j$ is $\frac{n}{2^j} \times n$ matrix related to scale $s_j = 2^{j-1}$ and $\boldsymbol{v}_J$ is $\frac{n}{2^J} \times n$ matrix at scale 2^{J-1}. $\boldsymbol{w}_j$ and $\boldsymbol{v}_J$ are also known as wavelet and scaling filters, respectively. The synthesis equation leads to additive decomposition which is known as multiresolution analysis (MRA). Mallat (1989) developed an efficient way to implement this scheme using filters, which is widely used in signal processing (Strang & Nguyen, 1996). In the end, we can simplify the synthesis equation into the following form

$$\mathbf{X} = \sum_{j=1}^{J} \mathbf{d_j} + \mathbf{a_J} \tag{6}$$

where $\mathbf{d}_j = \mathbf{w}_j^{\mathrm{T}} \mathbf{D}_j$ is the reconstructed details, while $\mathbf{a}_J = \mathbf{v}_J^T \mathbf{A}_J$ is the reconstructed approximations. Generally, $\boldsymbol{a}_j$ and $\boldsymbol{d}_j$ represent the large-scale and fine-scale information of the original signal at decomposition level j with associated scale $s = 2^{j-1}$. In the matrix form, the reconstructed time series can be written as

$$\mathbf{X} = \widetilde{\mathbf{R}}\mathbf{I} \tag{7}$$

where $\widetilde{\mathbf{R}}$ is the standardized reconstructions matrix $\mathbf{R} = [\mathbf{d}_1, ..., \mathbf{d_J}, \mathbf{a_J}]$ (i.e., transforming the columns of the reconstructions matrix to zero mean and unit standard deviation) and $\mathbf{I} = [\sigma_{d_1}, ..., \sigma_{d_J}, \sigma_{a_J}]^{\mathrm{T}}$

So far, the basis for DWT has been described. We now consider two additional important aspects before proceeding to our proposed variance transformation. First, boundary conditions (BC) represent the main source of error in wavelet and scaling coefficients and reconstructions (Bakshi, 1999; Maheswaran & Khosa, 2012;

John Quilty & Adamowski, 2018). This is the reason dyadic samples are used in synthetic experiments presented in the next section. Second, the selection of decomposition levels and wavelet family is closely related to BC introduced errors. It is recommended that wide wavelet filters and high decomposition levels be avoided when the sample size is limited (Percival & Walden, 2000; John Quilty & Adamowski, 2018). In this study, Daubechies wavelets were used, which represents a family of orthogonal wavelets and allows the time series to be decomposed and reconstructed perfectly by adding all the details and approximations together. The summation of the variance of details and approximations also equals the variance of the original time series (Burrus et al., 1998; Daubechies, 1992). Another important reason why we chose the Daubechies family here is due to the flexibility of choosing different lengths of filters among the same wavelet family, including db1 (Haar), db2, ... , and db10 (Daubechies, 1992). The proposed method here is a generic method potentially for any natural system modeling; thus, we also want to include a wavelet family which can handle different types of physical processes with varying lengths. In addition, this wavelet family has been well studied in the context of hydroclimatology (Maheswaran & Khosa, 2012; Nalley et al., 2012; Rashid et al., 2018; Rathinasamy et al., 2014; Shafaei & Kisi, 2016). Their results suggest that the Daubechies family has significant advantages over other wavelet families in the field of hydroclimatology. The associated decomposition levels J were determined by the following equation (Kaiser, 2010):

$$J = \frac{\log\left(\frac{n}{2v-1}\right)}{\log(2)} \tag{8}$$

where v is the number of vanishing moments for a given wavelet (e.g., Daubechies 5, db5, has a 10-point filter length with v equals to 5). This maximum decomposition level is the one where the number of data after the last subsampling becomes smaller than the wavelet filter length (de Artigas et al., 2006), which allows us to look at the complete frequency domain. A table of decomposition levels for all the data sets investigated in the study is provided in Table S1 in the supporting information.

2.2. Variance Transformation in the Frequency Domain

As introduced previously, we are aiming to modify a predictor variable to exhibit similar spectral properties as the response and DWT is adopted as a means to characterize the spectrum. The following variance transformation technique was developed to verify our two hypotheses.

Assume that there are n paired centered (i.e., with mean of zero) observations of the predictor variable $\boldsymbol{X}$ and the response variable $\boldsymbol{Y}$, that is, (x_0, y_0),...,(x_{n-1}, y_{n-1}). First, DWT is applied to predictor $\boldsymbol{X}$ resulting in wavelet and scaling coefficients $\mathbf{W} = [\mathbf{D}_1, ..., \mathbf{D}_\mathbf{J}, \mathbf{A}_\mathbf{J}]$. Reconstructed details and approximations using the wavelet and scaling coefficients are denoted as $\mathbf{R} = [\mathbf{d}_1, ..., \mathbf{d}_\mathbf{J}, \mathbf{a}_\mathbf{J}]$ with an associated standard deviation matrix $\mathbf{I} = [\sigma_{d_1}, ..., \sigma_{d_J}, \sigma_{a_J}]^T$. The rationale behind DWT ensures that the summation of the variance of details and approximations also equals the variance of the original time series, which means $\sum_{j=1}^{J+1} \mathbf{I}_j^2 = \frac{1}{n-1}\mathbf{X}^\mathbf{T}\mathbf{X} = \sigma_\mathbf{X}^2$. Therefore, $\boldsymbol{X}$ can be expressed as matrix multiplication $\mathbf{X}=\widetilde{\mathbf{R}}\mathbf{I}$ with the standardized reconstructions $\widetilde{\mathbf{R}}=\left[\widetilde{\mathbf{d}}_1, ..., \widetilde{\mathbf{d}}_\mathbf{J}, \widetilde{\mathbf{a}}_\mathbf{J}\right]$. The variance transformation is achieved by reconstructing a new predictor vector $\boldsymbol{X'}$ with variance structure $\boldsymbol{\alpha}$ aligned with the covariance matrix between the response and predictor reconstructions in the frequency domain. They can be written as

$$\mathbf{X}^{'}=\widetilde{\mathbf{R}}\,\boldsymbol{\alpha} \qquad \boldsymbol{\alpha} = \sigma_X\widetilde{\mathbf{C}} \tag{9}$$

where $\widetilde{\mathbf{C}}$ is the normalized covariance matrix (i.e., with unit norm) for the variable set $(\boldsymbol{Y},\widetilde{\mathbf{R}})$ and the covariance matrix $\boldsymbol{C}$ has the form of

$$\mathbf{C} = \frac{1}{n-1}\mathbf{Y}^T\widetilde{\mathbf{R}} = \left[S_{Y\widetilde{d}_1}, ..., S_{Y\widetilde{d}_J}, S_{Y\widetilde{a}_J}\right] \tag{10}$$

and $\boldsymbol{\alpha}$ is the unique variance transformation matrix given a predictor variable and the corresponding response. Essentially, the reconstructed predictor vector $\boldsymbol{X'}$ is obtained by redistributing the variance in its

spectrum while maintaining the total variance of the original predictor $\boldsymbol{X}$. We cast the above variance transformation process into the following form for ease of illustration in the following sections:

$$\mathbf{X}' = g(\mathbf{X}) \tag{11}$$

where $g(\cdot)$ denotes the variance transformation operation represented in equations (9) and (10). In a linear system using simple linear regression, we can derive the theoretical optimal predictive accuracy of the target response as measured by root-mean-square error (RMSE) as (the detailed derivation refers to Text S1 in the supporting information)

$$RMSE_{\min} = \sqrt{\frac{n-1}{n}\left(\sigma_Y^2 - \|\mathbf{C}\|^2\right)}. \tag{12}$$

In subsequent sections, we illustrate the method by first identifying the significant predictor variables. The significant predictors are selected once they are reconstructed using the variance transformation method, and the reconstructed predictors are then used for response prediction.

2.3. Description of the Algorithm

The wavelet-based variance transformation algorithm adopted partial informational correlation (PIC) to identify significant predictors. PIC is a distribution-independent means of measuring dependence between a given response and a predictor, conditioned to the preexisting predictor(s). A predictor is considered as significant when the estimated PIC is highest among all the remaining candidate predictors, and the maximum PIC is larger than an associated threshold PIC_p at a certain significance level p (we used $p = 0.05$ in this study). Readers are referred to Sharma (2000) and Sharma and Mehrotra (2014) for additional details, as well as to Sharma et al. (2016) for the software needed to estimate the PIC. Given a set of candidate predictor variable $\boldsymbol{X}$, an initially empty predictor vector $\boldsymbol{Z}$, and a target response Y, the algorithm proceeds at each iteration by finding the variance-transformed candidate predictor variable $\boldsymbol{X}$ that maximizes PIC with respect to the response Y, conditioned to the preexisting predictor(s) $\boldsymbol{Z}$ that have been selected. The details of the algorithm are given below:

a Reconstruct a new set of predictors $\boldsymbol{X'}$ by applying the variance transformation $\boldsymbol{X}|\boldsymbol{Z}$ with respect to $Y|\boldsymbol{Z}$. $\boldsymbol{X}|\boldsymbol{Z}$ and $Y|\boldsymbol{Z}$ represent the residual information in the response Y and predictor variable $\boldsymbol{X}$, when the effect of the preexisting predictor(s) $\boldsymbol{Z}$ has been taken into account. We denote these as the partial response and partial predictor variables when $\boldsymbol{Z}$ is not empty.
b Estimate the PIC between the response $Y|\boldsymbol{Z}$ and each new candidate predictor in $\boldsymbol{X'}$. The predictor X^* with maximum PIC^* is selected.
c If $\text{PIC}^* > \text{PIC}_p$, include the significant predictor X^* in predictor vector $\boldsymbol{Z}$ and remove X^* from $\boldsymbol{X}$.
d Repeat Steps a–c, and the algorithm is terminated when PIC^* is smaller than given threshold PIC_p or there is no candidate predictor left in $\boldsymbol{X}$.

An additional step is to estimate the RMSE of the response $Y|\boldsymbol{Z}$ using both original $\boldsymbol{X}|\boldsymbol{Z}$ and variance-transformed $\boldsymbol{X'}$ with a predefined predictive model. The estimated RMSE is used to compare the models in the following sections. It should be noted that the partial response or partial predictor variables above were ascertained using the k-nearest neighbor (knn) regression in the leave-one-out cross-validation setting (in this study, $k = 5$ was used), an efficient alternative that has been found to produce stable outcomes in past applications (Friedman et al., 2001; Mehrotra & Sharma, 2006).

3. Application to Synthetic Data Sets With Known Attributes

The wavelet-based method proposed in this paper was designed to integrate the utility of DWT with predictor variable selection in the predicting response of interest. The aim of this section is to illustrate whether the variance transformation technique can be used to improve predictive performance. To address this, we consider a range of synthetic examples with known attributes, examining both linear and nonlinear relationships between the response and associated predictors. These synthetic data sets were generated in a way that mimics the features of real natural systems: linearity and/or nonlinearity in the underlying process, persistence, and noise in response.

A range of synthetic data sets was generated to evaluate the application of the wavelet-based method. All examples have the same sample size $N = 512$ and the total number of candidate predictors $P = 9$. The amount of the noise in the response and predictors is defined as the standard deviation of the Gaussian, while noise $s\varepsilon$ where $\varepsilon \sim N(0,1)$, and s is the scaling factor used to alter the level of noise in the response and predictors. Three examples are considered, including linear additive sinewave (SW) models with one and three true predictors (denoting as SW1 and SW3, respectively) and a nonlinear hysteresis loop (HL) model. Each of the examples is tested with varying levels of noise in the response and predictors, with three levels of noise used: denoted as low, moderate, and high noise. Examples of variance structure in the spectrum before and after variance transformation for SW1_l and HL_l are given in Figure S1 in the supporting information. The details of the model setup are given in the following sections.

We also investigate the sensitivity of our results to changes in sample size N when the noise levels vary in the response and predictors. The performance of the models is evaluated on both calibration and validation data sets, which is done by partitioning each sample of synthetically generated data set into two complementary subsets. One subset is used as the calibration set, while the other subset is used as the validation set. It should be noted that when applying this method to the situation that the response is unknown (in the validation set), the covariance matrix in the equation (10) from the calibration set will be transferred to the validation set along with the fitted predictive model.

3.1. SW Data Sets

SW models of varying frequencies were added together to generate synthetic data sets. The two groups of data sets were generated by

$$\begin{aligned} x_t &= \sin(2\pi f_i t) + s\varepsilon_t \\ y_t &= \sum_{i=1}^{p} \sin(2\pi f_i t) + s\varepsilon_t \end{aligned} \tag{13}$$

where $f_1, f_2, \ldots, f_p$ are the frequencies of the sine wave and t ranges from 0 to 1. The response of two groups of SW model data sets was generated from y_t with

$$\text{SW1: } p = 1, f = 25 \text{ Hz}$$

$$\text{SW3: } p = 3, f = 15, 25 \text{ and } 30 \text{ Hz}$$

The nine candidate predictors were generated from x_t with varying frequencies from 3, 5, 10, 15, 25, 30, 55, 70, and 95 Hz, respectively, denoted as predictor variable $X_1, \ldots, X_9$. Thus, the true predictor of the response SW1 is predictor variable X_5, while the true predictors of the response SW3 are predictor variables X_4, X_5, and X_6.

Results from synthetic data SW1 using simple linear regression as the predictive model are presented in Table 1. The original predictors and reconstructed predictors with variance transformation are compared. The estimated RMSE with the variance transformation operation always reaches the optimal prediction accuracy with minimized RMSE given by equation (12). Predictor X_5 leads to the lowest RMSE in the response compared with other predictors, which is expected since X_5 is the true predictor for SW1. When the true predictor was identified and considered as the preexisting predictor, the RMSE of the partial response still improved slightly with variance-transformed partial predictor variables as shown in Table 1.

Table 2 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 variance-transformed predictors. According to the results of synthetic model SW1, it can be seen that true predictor X_5 of SW1 is identified 100% of times using both original- and variance-transformed predictors for all three noise levels. However, additional predictor variables are often selected using the wavelet-based method when considering the variable and the response as residual information. This is expected since the aim of the variance transformation is to force the predictor to be similar to the spectrum of the partial response, resulting in higher dependence. The advantage of the proposed approach is thus not evident for the simple SW1 model and low noise levels. However, as shown in the following section when there are multiple predictors in a

Table 1
Comparison of Prediction Accuracy Using Both Original and Reconstructed New Predictor Sets Averaged Across 100 Realizations: SW Data Sets

		Response			Partial response		
SW1	Predictor	Original	Variance transformed	Optimal	Original	Variance transformed	Optimal
s = 0.1	1	0.713	0.703	0.703	0.147	0.146	0.146
	2	0.713	0.708	0.708	0.147	0.147	0.147
	3	0.713	0.712	0.712	0.147	0.146	0.146
	4	0.713	0.713	0.713	0.147	0.146	0.146
	5	0.140	0.116	0.116	—	—	—
	6	0.713	0.712	0.712	0.147	0.146	0.146
	7	0.713	0.703	0.703	0.147	0.146	0.146
	8	0.713	0.703	0.703	0.147	0.146	0.146
	9	0.713	0.711	0.711	0.147	0.146	0.146
s = 0.5	1	0.862	0.852	0.852	0.692	0.687	0.687
	2	0.862	0.853	0.853	0.692	0.687	0.687
	3	0.862	0.858	0.858	0.692	0.687	0.687
	4	0.862	0.859	0.859	0.692	0.688	0.688
	5	0.645	0.539	0.539	—	—	—
	6	0.862	0.860	0.860	0.692	0.688	0.688
	7	0.862	0.853	0.853	0.692	0.687	0.687
	8	0.862	0.855	0.855	0.692	0.687	0.687
	9	0.862	0.855	0.855	0.692	0.687	0.687
s = 1.0	1	1.229	1.219	1.219	1.265	1.256	1.256
	2	1.229	1.218	1.218	1.264	1.256	1.256
	3	1.229	1.220	1.220	1.265	1.256	1.256
	4	1.229	1.221	1.221	1.265	1.256	1.256
	5	1.157	1.051	1.051	—	—	—
	6	1.229	1.222	1.222	1.265	1.256	1.256
	7	1.229	1.218	1.218	1.265	1.255	1.255
	8	1.229	1.217	1.217	1.264	1.254	1.254
	9	1.229	1.219	1.219	1.265	1.255	1.255

linear system and/or a nonlinear system, the variance transformation is more effective. The more complex models are more representative of real natural systems.

The results from this application to synthetic model SW3 with three true predictors (X_4, X_5, and X_6) and higher noise levels (s = 1.0, 2.0, and 3.0) are given in Table 2 as well. The relatively high noise level introduced in this linear system with multiple predictors is used to demonstrate the effectiveness of the variance transformation technique. As the noise in the system increases, the correct predictors are more often selected by the wavelet-based method than the original predictors. The wavelet-based method is able to identify three predictors in 93% (63%) of cases compared to 9% (1%) for the original predictors when data have moderate (high) noise, highlighting the advantage of the proposed approach.

3.2. Hysteresis Loop Data Sets

Hysteresis is a common nonlinear phenomenon in natural systems, for example, storage-discharge relationships in rivers systems (Fovet et al., 2015). We adopt an analytical model to approximate a classical hysteresis loop (Lapshin, 1995), and the synthetic data were generated by

$$\begin{aligned} x_t &= 0.8\cos(2\pi ft) + s\varepsilon_t \\ y_t &= 0.6\cos(2\pi ft)^3 - 0.2\sin(2\pi ft)^5 + s\varepsilon_t \end{aligned} \quad (14)$$

The response of HL model data sets was generated from y_t with f = 25 Hz. The nine candidate predictors were the same as the linear SW models. The true predictor of the HL response is predictor variable X_5.

Results from synthetic model HL using knn regression as a predictive model are presented in Table 3 where the original predictors and reconstructed predictors with variance transformation are compared. The estimated RMSE with variance transformation is smaller than using original predictors in most cases except

Table 2
Frequency of Predictor Selection and the Number of Predictors Selected Using Both Original and Reconstructed New Predictor Sets: SW Data Sets

		Percentage of samples the predictor selected			Percentage of samples total number of predictors identified		
Model	Noise level	Predictor	Original	Variance transformed	Number of predictors	Original	Variance transformed
SW1	s = 0.1	1	0	0	1	100	93
		2	0	2	2	0	6
		3	0	2	3	0	1
		4	0	1	4	0	0
		5	100	100	5	0	0
		6	0	3	6	0	0
		7	0	0	7	0	0
		8	0	0	8	0	0
		9	0	0	9	0	0
	s = 0.5	1	0	11	1	97	49
		2	2	13	2	3	25
		3	0	8	3	0	17
		4	0	6	4	0	7
		5	100	100	5	0	1
		6	0	10	6	0	1
		7	1	14	7	0	0
		8	0	16	8	0	0
		9	0	11	9	0	0
	s = 1.0	1	0	19	1	96	32
		2	0	17	2	3	27
		3	0	16	3	1	26
		4	1	10	4	0	7
		5	100	100	5	0	7
		6	1	12	6	0	1
		7	2	12	7	0	0
		8	1	25	8	0	0
		9	0	22	9	0	0
SW3	s = 1.0	1	2	36	1	0	0
		2	0	25	2	2	0
		3	1	16	3	95	18
		4	100	100	4	3	25
		5	99	100	5	0	31
		6	99	100	6	0	16
		7	0	38	7	0	8
		8	0	27	8	0	2
		9	0	35	9	0	0
	s = 2.0	1	0	22	1	29	1
		2	0	22	2	4	4
		3	1	25	3	1	25
		4	15	99	4	0	20
		5	9	96	5	0	26
		6	13	93	6	0	16
		7	1	25	7	0	6
		8	0	34	8	0	2
		9	1	32	9	0	0
	s = 3.0	1	1	19	1	9	7
		2	0	19	2	2	23
		3	0	26	3	0	18
		4	4	71	4	0	23
		5	1	67	5	0	13
		6	3	63	6	0	7
		7	1	23	7	0	1
		8	1	23	8	0	2
		9	2	18	9	0	0

Table 3
Comparison of Prediction Accuracy Using Both Original and Reconstructed New Predictor Sets Averaged Across 100 Realizations: HL Data Sets

Part (a)	Predictor	Response		Partial response	
		Original	Variance transformed	Original	Variance transformed
s = 0.1	1	0.402	0.393	0.202	0.200
	2	0.338	0.396	0.203	0.201
	3	0.402	0.399	0.203	0.202
	4	0.394	0.397	0.203	0.201
	5	0.185	0.169	—	—
	6	0.400	0.396	0.203	0.202
	7	0.402	0.393	0.203	0.201
	8	0.402	0.391	0.202	0.201
	9	0.400	0.396	0.203	0.201
s = 0.5	1	0.670	0.665	0.677	0.672
	2	0.671	0.663	0.676	0.673
	3	0.672	0.666	0.677	0.671
	4	0.671	0.668	0.676	0.672
	5	0.617	0.582	—	—
	6	0.671	0.667	0.676	0.673
	7	0.673	0.667	0.676	0.671
	8	0.672	0.664	0.677	0.674
	9	0.672	0.665	0.676	0.672
s = 1.0	1	1.163	1.152	1.262	1.250
	2	1.159	1.152	1.260	1.253
	3	1.162	1.154	1.260	1.251
	4	1.160	1.154	1.257	1.248
	5	1.149	1.116	—	—
	6	1.161	1.153	1.258	1.251
	7	1.161	1.154	1.261	1.254
	8	1.158	1.151	1.258	1.255
	9	1.165	1.148	1.259	1.249

for predictor variables X_2 and X_4 at low noise levels. Similarly, in Table 3 predictor variable X_5 has the lowest RMSE compared with other predictors as expected since predictor variable X_5 is the true predictor for HL. When the true predictor was identified and considered as the preexisting predictor, the RMSE of the partial response still improved slightly with variance-transformed partial predictor variables as shown in Table 3.

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 predictor identified using original- and variance-transformed predictors. It can be seen that true predictor X_5 of HL is correctly identified 100% of cases using both original- and variance-transformed predictors when the noise is low or moderate. But once there is high noise, the performance of the original predictors sharply decreases (only 39% of cases), while the variance transformation continues to correctly identify the true predictor in all 100 cases. In summary, considering the results from all the three examples, the advantage of identifying additional predictors using wavelet-based method is well demonstrated.

3.3. Sensitivity Analysis

The sensitivity of our results to changes in sample size N given different levels of noise in the system is shown in Figures 1 and 2. Note that only the improved (i.e., reduced) RMSE of response from SW1 and HL is given here, and the performance of their partial response is similar with slightly smaller improvements. These results presented are also based on 100 realizations for different sample sizes and noise levels. As described previously, the first half of independent samples is used for calibration and another half is used for validation to ascertain the performance of variance transformation in predicting unknown response in the future.

For calibration results, RMSE is improved particularly when sample sizes are low and/or noise levels are high when using variance-transformed predictors. When sample sizes are small, the RMSE is relatively

Table 4
Frequency of Predictor Selection and the Number of Predictors Selected Using Both Original and Reconstructed New Predictor Set: HL Data Sets

	Percentage of samples the predictor selected			Percentage of samples total number of predictors identified		
	Predictor	Original	Variance transformed	Number of predictors	Original	Variance transformed
$s = 0.1$	1	0	7	1	100	82
	2	0	3	2	0	17
	3	0	1	3	0	1
	4	0	2	4	0	0
	5	100	100	5	0	0
	6	0	0	6	0	0
	7	0	4	7	0	0
	8	0	2	8	0	0
	9	0	0	9	0	0
$s = 0.5$	1	0	12	1	96	42
	2	0	21	2	4	22
	3	0	17	3	0	15
	4	0	13	4	0	14
	5	100	100	5	0	5
	6	1	8	6	0	1
	7	2	19	7	0	1
	8	1	19	8	0	0
	9	0	16	9	0	0
s = 1.0	1	0	24	1	40	32
	2	0	26	2	1	23
	3	0	18	3	0	20
	4	1	13	4	0	13
	5	39	100	5	0	8
	6	1	6	6	0	3
	7	0	24	7	0	0
	8	0	18	8	0	1
	9	1	27	9	0	0

large giving more scope for improvements with the proposed new method. For the calibration result of HL, there are some exceptions in some cases of X_2, X_4, and X_8 at low noise level, and it is expected since the method is developed on the basis of linear regression for a linear system. The validation result of SW1 demonstrates that the usage of variance transformation trends to be more superior in terms of predictability in almost all cases (Figure 1b). Figure 2b gives the validation result of HL, which shows relatively less improvement, but it generally demonstrates the capability of the technique in a higher noise system particularly. The largest improvements are seen for the predictor variable X_5 for both calibration and validation for all sample sizes since X_5 is the true predictor for both SW1 and HL.

To sum up, in the linear system, the predictive accuracy can reach the theoretical optimal RMSE using the variance-transformed predictor set. In the nonlinear system, the variance transformation technique still improves the prediction accuracy compared with the original data set in general. The application of the variance transformation technique to a dynamic example given in the following section further demonstrates its capability in a chaotic system.

4. Application to a Dynamic Example

A continuous dynamical system that exhibits chaotic dynamics was proposed by Rössler (1976) in 1976. The Rössler system consists of a set of ordinary differential equations defined as follows:

$$\dot{x} = -y - z \qquad \dot{y} = x + ay \qquad \dot{z} = b + z(x - c) \tag{15}$$

where a, b, and c are parameters that specify the chaotic system. We investigate the system parameters with parameters value $a = 0.2$, $b = 0.2$, and $c = 5.7$, which is known to produce a deterministic chaotic time series (Strogatz, 2018; Harrington & Van Gorder, 2017). The time range is fixed from 0 to 50, and a total number of

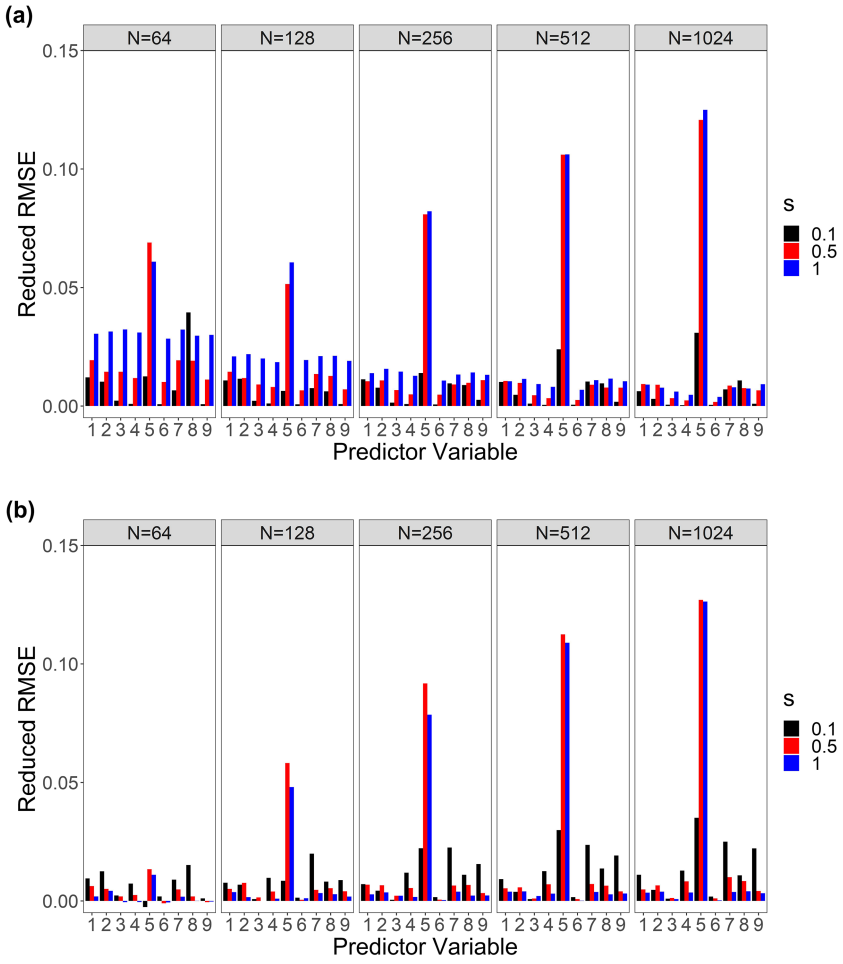

Figure 1. Sensitivity analysis of the prediction accuracy of response, SW1, to changes in sample size *N*. The *y* axis gives the reduced RMSE compared to the model using original predictor variable *X*, and the *x* axis indicates the individual predictor variable. (a) Calibration. (b) Validation.

$N = 100{,}000$ paired observations (x_t, y_t, z_t) was generated from an initial condition of $(-2, -10, 0.2)$ with low noise level ($s = 0.1$). The generated time series of x and y are used as predictors, while z is considered as the response. The knn regression is used as the predictive model. The first half of the data is used for calibration, while the second half of the data is used to verify the variance transformation technique. In Figure 3, the observed and predicted z series with and without variance transformation are given in 3-D view. The wavelet-based model performs substantially better than the model without variance transformation at the calibration stage with errors one magnitude smaller than the model using original predictors (RMSE equals 0.113 using variance-transformed predictors, while RMSE is 1.189 using original predictors). In terms of standard deviation, all three are similar while the proposed model has a marginally higher correlation of 0.999 compared to 0.930. At the validation stage, both models predict small z values well but have difficulties with large z values. The model using original predictors produces larger numbers of high values, while the model using variance-transformed predictors only captures several peaks of high values. However, the overall prediction accuracy of the wavelet-based model is improved, where RMSE is about 2.550 against 4.493 using original predictors. In addition, the correlation of the model with and without variance transformation is 0.56 and 0.36, respectively; the standard deviation between observed response ($\sigma_{obs} = 2.9$) and predicted response using variance-transformed predictors ($\sigma_{pred(vt)} = 2.3$) is equivalent, while the predicted response using original predictors shows much higher standard deviation ($\sigma_{pred} = 4.5$). All three metrics suggest the superiority of the proposed method.

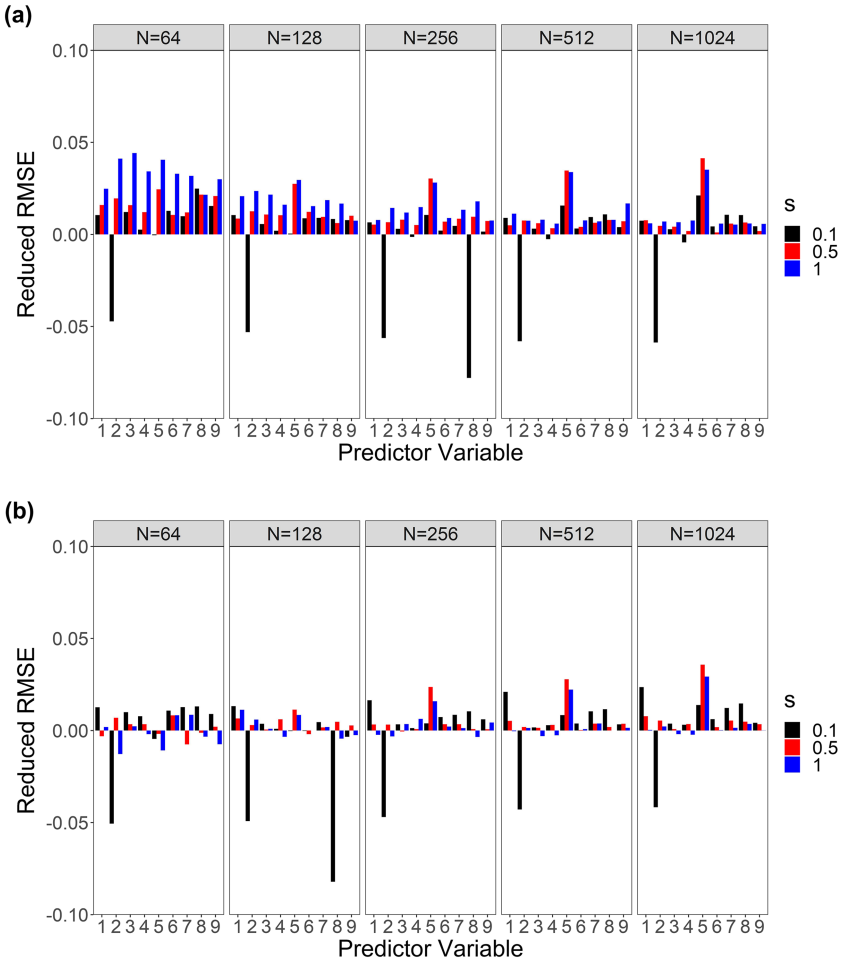

Figure 2. Sensitivity analysis of the prediction accuracy of response, HL, to changes in sample size *N*. The *y* axis gives the reduced RMSE compared to the model using original predictor variable *X*, and the *x* axis indicates the individual predictor variable. (a) Calibration. (b) Validation.

Additional results of the proposed method to other linear, nonlinear, and dynamic systems are given in the supporting information. These include a higher-order autoregressive model (AR4) and a nonlinear threshold autoregressive model (TAR), which were investigated previously by Cryer and Chan (2008), Galelli et al. (2014), and Sharma and Mehrotra (2014). Duffing system (Dignowity et al., 2013; Duffing, 1918) is another example of a dynamical system that exhibits chaotic behavior. In all synthetic experiments, our proposed method exhibits superior accuracy than the reference model using nontransformed predictors in terms of RMSE, correlation, and standard deviation. However, it should be noted that when predictors and the associated response have similar spectral attributes as is the case in these additional experiments where the predictors are actually lagged values of the corresponding response, the improvement in predictability is small.

5. Application to a Real-World Example

The wavelet-based method is applied to downscale a selected drought index (Standardized Precipitation Index, SPI12) over the Sydney region in Australia. We adapted the data set which has been presented in earlier studies by Mehrotra and Sharma (2006, 2010) for downscaling and Mehrotra and Sharma (2015, 2016) and Mehrotra et al. (2004) for bias correction. The data used here consist of seven atmospheric variables as potential predictor variables to ground precipitation. These atmospheric variables are monthly time series at 25 reanalysis grid points (2.5° × 2.5°) over Sydney, Australia, from 1950 to 2009, obtained from the National Center for Environmental Prediction Reanalysis data provided by the NOAA/OAR/ESRL PSD,

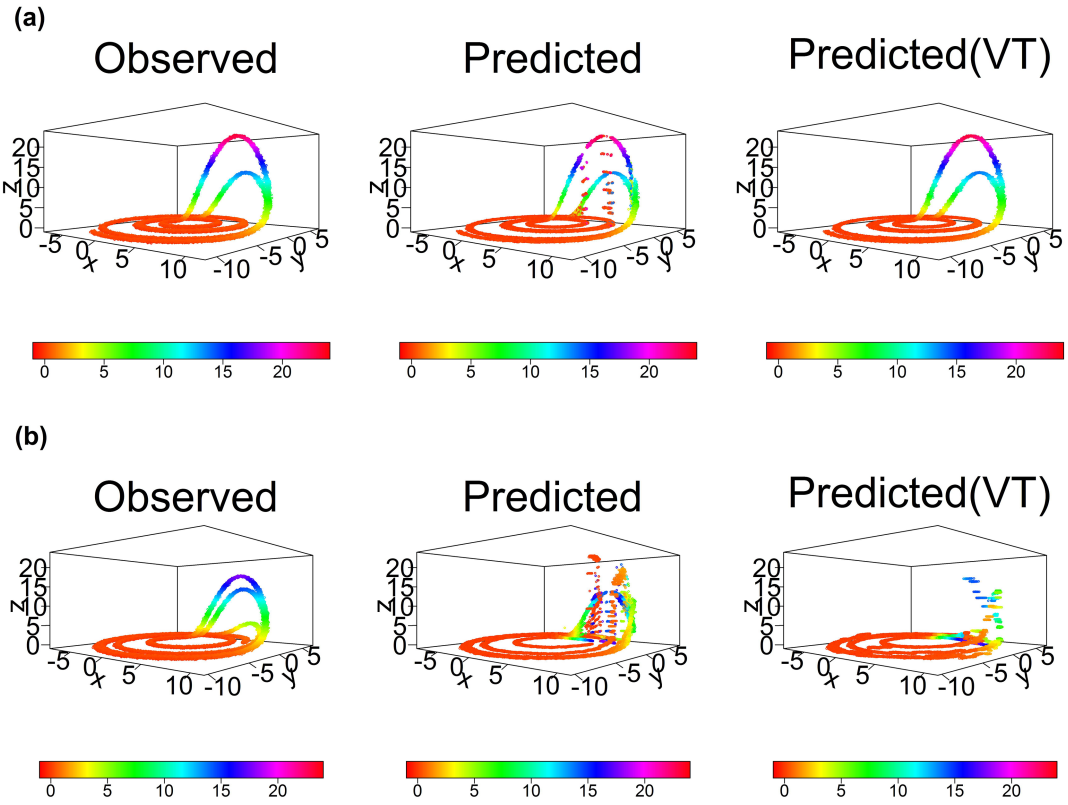

Figure 3. Comparison of observed and predicted z values with and without variance transformation (VT) of x and y. (a) Calibration. (b) Validation.

Boulder, Colorado, USA (https://www.esrl.noaa.gov/psd/). Nonstationarity and persistence are both present in these atmospheric variables, making them a complex and challenging natural system to model. A 61-yr record (1949–2009) of monthly rainfall at 15 stations around Sydney is also obtained, and the observed SPI12 is computed from the observed rainfall accordingly (Guttman, 1999). In addition, to demonstrate the utility of the wavelet-based method, cross validation is performed by partitioning a sample of historical data into two complementary subsets. One subset is used as the calibration set, while the other subset is used as the validation set followed by their order being reversed, leading to results presented and discussed hereafter as cross-validated results for the entire study period. This rationale for using cross validation is that we can assess the model performance with independent data sets (Mehrotra & Sharma, 2006; H. Nguyen et al., 2019) hence indicative of performance under new conditions.

Figures 4 and 5 show the performance of the wavelet-based model. It is clearly demonstrated that the wavelet-based method exhibits better prediction accuracy. First, in Figure 4, the tails of the probability distributions of predicted SPI12 with variance-transformed predictors are considerably closer to observed SPI12 in most cases. This shows that the proposed approach is capable of capturing sustained dry/wet anomalies well. Second, Figure 5a presents the comparison results using a Taylor Diagram (Taylor, 2001), which is typically used to summarize multiple aspects of model performance in a single plot. These metrics include the standard deviation, centered RMSE, and correlation coefficient. It can be seen that the wavelet-based model consistently exhibits higher correlations, lower RMSE, and higher standard deviation. The higher standard deviation is expected since the proposed approach can match sustained dry/wet anomalies better than other alternatives. Lastly, the reduced RMSE of all the investigated rainfall stations relative to the nonwavelet model is shown in Figure 5b. This result is especially pertinent given the expected changes in aridity into the future (Zarch et al., 2015).

In terms of application in real-world cases, there are some points worthy of notice. First, as we stated previously this approach is limited to the condition where the future is "known." However, it can also be applied in a forecasting framework using only past information to predict the future, where the

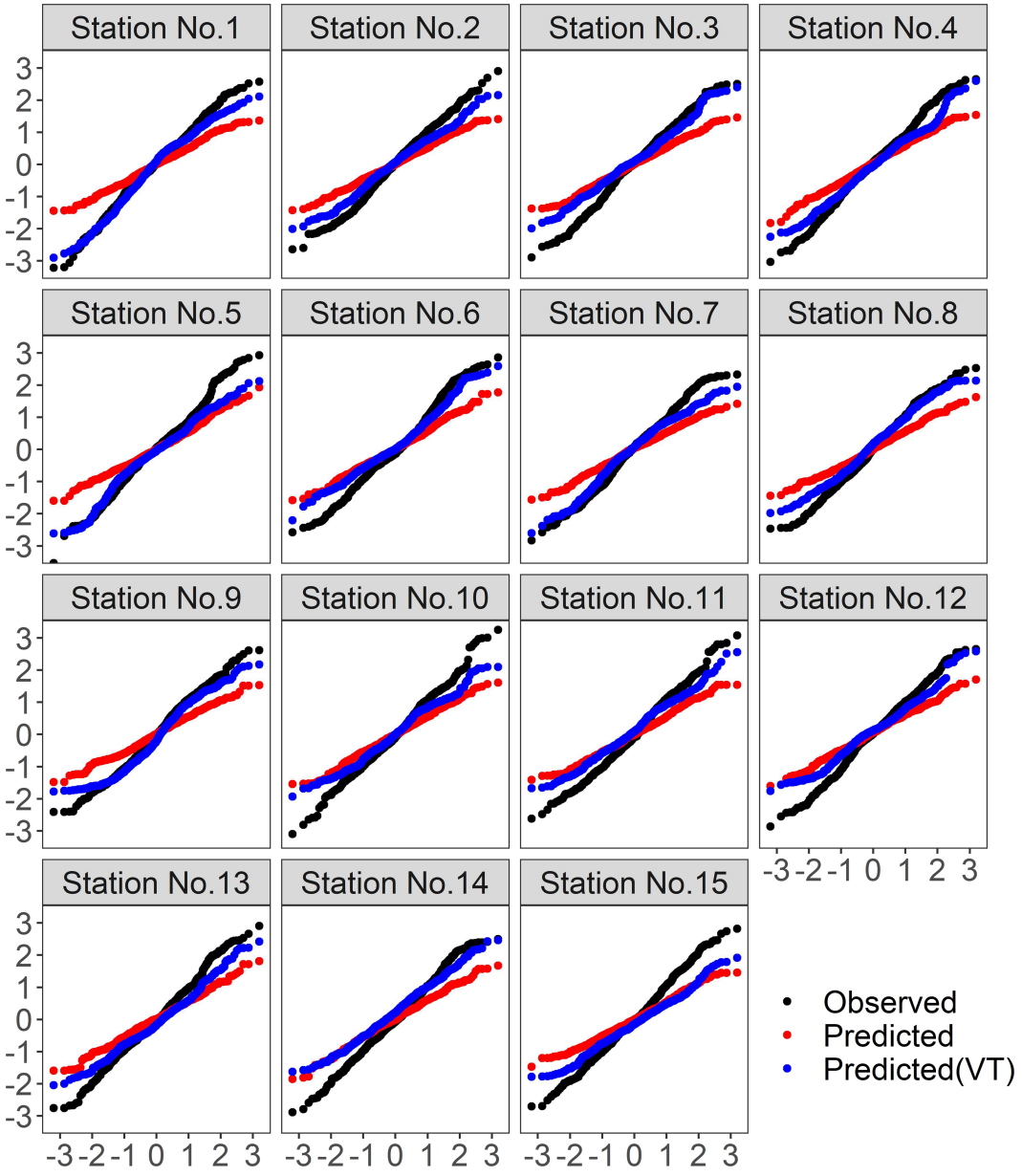

Figure 4. Quantile plots of observed, predicted, and predicted with variance transformation SPI12.

DWT-MRA should be replaced by the Maximal Overlap DWT (MODWT) or à trous algorithm (AT). MODWT and AT overcome the drawbacks of DWT-MRA and allow use when the future is not known (Nason, 2010; John Quilty & Adamowski, 2018). Second, although the variance transformation is mathematically derived based on the linear regression model, the assumptions we proposed are not only valid in a linear system but also can be applied to model nonlinear systems if a nonlinear or a nonparametric model (knn, as implemented in this study) is used for the regression. This is demonstrated by results from both synthetic and real data examples.

Additionally, other issues originating from the DWT should be noted: (1) The selection of the wavelet family, for example, Daubechies family, vanishing moment (ν) controls how well the fast-changing predictor variables information is captured (e.g., as in daily precipitation) (Maheswaran & Khosa, 2012). (2) Due to the requirement of a dyadic sample size in the DWT decomposition and reconstruction (Bakshi, 1999; Torrence & Compo, 1998), there is a slight difference of total variance before and after DWT reconstruction, which can lead to the subtle difference in the optimal RMSE and estimated RMSE in linear regression

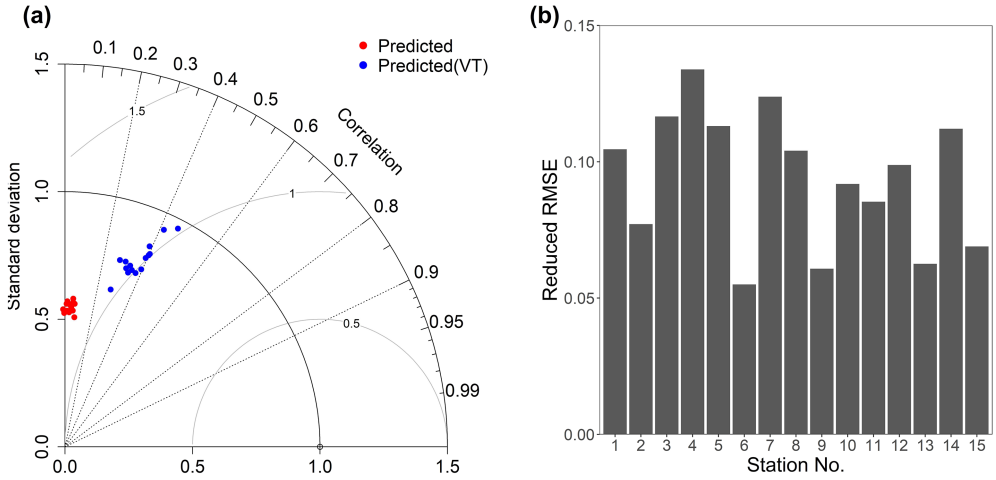

Figure 5. (a) Taylor Diagram of model prediction using wavelet (blue points) and nonwavelet (red points) model in a cross-validation setting. In the diagram normalized standard deviation is on the radial axis; correlation coefficient is on the angular axis, and dashed lines indicate centered RMSE. (b) Bar plot of reduced RMSE compared to the nonwavelet model.

predictive system. (3) The selection of the decomposition level is related to the wavelet filter selected and length of data available. For the most cases of using standard wavelet method lower decomposition levels are preferred, but in terms of the proposed method in this study the decomposition level investigating the complete frequency domain is recommended (systematic assessments on this are provided in Table S2 using synthetic experiments and Figures S2 and S3 using the same real-world example with $J = 2$). (4) BC introduced errors could be further addressed by using boundary corrected wavelet and scaling coefficients using either MODWT or AT as proposed in the work of John Quilty and Adamowski (2018).

6. Conclusions

This paper has presented an information theoretic method to improve the predictive performance of a given response in a natural system based on DWT. This method finds a unique variance transformation in the entire frequency domain of a predictor that can explain maximal information of a given response. We have assessed the utility of the wavelet-based approach using synthetically generated data sets from known linear and nonlinear systems. The two proposed hypotheses have been proven mathematically and using synthetic examples, these being the following: (1) If variance structure of the predictor in the spectrum is similar with the response, the predictive model exhibits improved accuracy (as measured by RMSE). (2) If variance structure between residual information in predictor and response variable conditioned to the preexisting predictor(s) is similar, the predictive model exhibits better accuracy. These assumptions are not only valid in a linear system but also can be applied to modeling nonlinear systems. The results have shown clear improvements in identifying meaningful predictors and improving the predictive performance of the response compared to the use of untransformed predictors, especially in a system with a high level of noise and nonlinearity. The application to the dynamic example demonstrates applying the approach to model a natural system with remarkable different spectral attributes in variance distribution. It is suggested that this method is more effective in real-world situations where the system is often characterized with a higher level of noise and a nonlinear relationship between the response and predictor variables. In conclusion, it has been confirmed that there exists a unique variance transformation in the frequency domain representation of a predictor that can explain maximal information of a corresponding response. Additional predictor variables can be selected using the wavelet-based method by expressing them and the response as residual information when the effect of preidentified predictor variables is considered.

Acknowledgments
This research was funded by the Australian Research Council linkage Grant (LP150100548) and Crown lands & Water Division, Department of Industry, NSW, Australia. Data used in synthetic experiments are generated from known statistical models, and the statistical properties of generated samples are given in the manuscript. NCEP Reanalysis data are provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their Web site (https://www.esrl.noaa.gov/psd/). We are also thankful to four anonymous referees for their constructive comments.

References

de Artigas, M. Z., Elias, A. G., & de Campra, P. F. (2006). Discrete wavelet analysis to assess long-term trends in geomagnetic activity. *Physics and Chemistry of the Earth, Parts A/B/C*, *31*(1-3), 77–80.

Aussem, A., Campbell, J., & Murtagh, F. (1998). Wavelet-based feature extraction and decomposition strategies for financial forecasting. *Journal of Computational Intelligence in Finance*, *6*(2), 5–12.
Bakshi, B. R. (1999). Multiscale analysis and modeling using wavelets. *Journal of Chemometrics*, *13*(3-4), 415–434.
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
Burrus, C. S., Gopinath, R. A., Guo, H., Odegard, J. E., & Selesnick, I. W. (1998). *Introduction to wavelets and wavelet transforms: A primer*, (Vol. 1). New Jersey: Prentice hall.
Cryer, J. D., & Chan, K.-S. (2008). Time series analysis with applications in R second edition Springer science+ business media. In: LLC.
Daubechies, I. (1990). The wavelet transform, time-frequency localization and signal analysis. *IEEE Transactions on Information Theory*, *36*(5), 961–1005.
Daubechies, I. (1992). *Ten lectures on wavelets* (Vol. 61). Philadelphia, PA 19104 USA: Society for Industrial and Applied Mathematics.
Dignowity, D., Wilson, M., Rangel-Fonseca, P., & Aboites, V. (2013). Duffing spatial dynamics induced in a double phase-conjugated resonator. *Laser Physics*, *23*(7), 075002.
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. *Journal of Hydrology*, *552*, 44–51.
Duffing, G. (1918). Erzwungene schwingungen bei veränderlicher eigenfrequenz. *Vieweg u. Sohn, Braunschweig, 7*.
Fahimi, F., Yaseen, Z. M., & El-shafie, A. (2017). Application of soft computing based hybrid models in hydrological variables modeling: A comprehensive review. *Theoretical and Applied Climatology*, *128*(3-4), 875–903. https://doi.org/10.1007/s00704-016-1735-8
Fovet, O., Ruiz, L., Hrachowitz, M., Faucheux, M., & Gascuel-Odoux, C. (2015). Hydrological hysteresis and its value for assessing process consistency in catchment conceptual models. *Hydrology and Earth System Sciences*, *19*(1), 105–123.
Friedman, J., Hastie, T., & Tibshirani, R. (2001). *The elements of statistical learning*, (Vol. 1). NY, USA: Springer series in statistics New York.
Galelli, S., Humphrey, G. B., Maier, H. R., Castelletti, A., Dandy, G. C., & Gibbs, M. S. (2014). An evaluation framework for input variable selection algorithms for environmental data-driven models. *Environmental Modelling and Software*, *62*, 33–51. https://doi.org/10.1016/j.envsoft.2014.08.015
Guttman, N. B. (1999). Accepting the standardized precipitation index: A calculation algorithm. *Journal of the American Water Resources Association*, *35*(2), 311–322. Retrieved from <Go to ISI>://000080900700010
Halmos, P. R. (2017). *Finite-dimensional vector spaces*. Mineola, New York, USA: Courier Dover Publications.
Harrington, H. A., & Van Gorder, R. A. (2017). Reduction of dimension for nonlinear dynamical systems. *Nonlinear Dynamics*, *88*(1), 715–734.
Hertig, E., & Tramblay, Y. (2017). Regional downscaling of Mediterranean droughts under past and future climatic conditions. *Global and Planetary Change*, *151*, 36–48.
Jiang, Z., Sharma, A., & Johnson, F. (2019). Assessing the sensitivity of hydro-climatological change detection methods to model uncertainty and bias. *Advances in Water Resources*, *134*, 103430.
Johnson, F., Westra, S., Sharma, A., & Pitman, A. J. (2011). An assessment of GCM skill in simulating persistence across multiple time scales. *Journal of Climate*, *24*(14), 3609–3623. https://doi.org/10.1175/2011JCLI3732.1
Kaiser, G. (2010). *A friendly guide to wavelets*. Berlin/Heidelberg, Germany: Springer Science & Business Media.
Kişi, Ö. (2011). A combined generalized regression neural network wavelet model for monthly streamflow prediction. *KSCE Journal of Civil Engineering*, *15*(8), 1469–1479. https://doi.org/10.1007/s12205-011-1004-4
Lapshin, R. V. (1995). Analytical model for the approximation of hysteresis loop and its application to the scanning tunneling microscope. *Review of Scientific Instruments*, *66*(9), 4718–4730.
Maheswaran, R., & Khosa, R. (2012). Comparative study of different wavelets for hydrologic forecasting. *Computers and Geosciences*, *46*, 284–295. https://doi.org/10.1016/j.cageo.2011.12.015
Mallat, S. G. (1989). A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *11*(7), 674–693. https://doi.org/10.1109/34.192463
Mehrotra, R., & Sharma, A. (2006). A nonparametric stochastic downscaling framework for daily rainfall at multiple locations. *Journal of Geophysical Research: Atmospheres*, *111*(D15), D15101.
Mehrotra, R., & Sharma, A. (2010). Development and application of a multisite rainfall stochastic downscaling framework for climate change impact assessment. *Water Resources Research*, *46*(7). https://doi.org/10.1029/2009WR008423
Mehrotra, R., & Sharma, A. (2015). Correcting for systematic biases in multiple raw GCM variables across a range of timescales. *Journal of Hydrology*, *520*, 214–223.
Mehrotra, R., & Sharma, A. (2016). A multivariate quantile-matching bias correction approach with auto-and cross-dependence across multiple time scales: Implications for downscaling. *Journal of Climate*, *29*(10), 3519–3539.
Mehrotra, R., Sharma, A., & Cordery, I. (2004). Comparison of two approaches for downscaling synoptic atmospheric patterns to multisite precipitation occurrence. *Journal of Geophysical Research: Atmospheres*, *109*(D14), D14107. https://doi.org/10.1029/2004JD004823
Nalley, D., Adamowski, J., & Khalil, B. (2012). Using discrete wavelet transforms to analyze trends in streamflow and precipitation in Quebec and Ontario (1954–2008). *Journal of Hydrology*, *475*, 204–228. https://doi.org/10.1016/j.jhydrol.2012.09.049
Nason, G. (2010). *Wavelet methods in statistics with R*. Berlin/Heidelberg, Germany: Springer Science & Business Media.
Ndehedehe, C. E., Agutu, N. O., Okwuashi, O., & Ferreira, V. G. (2016). Spatio-temporal variability of droughts and terrestrial water storage over Lake Chad Basin using independent component analysis. *Journal of Hydrology*, *540*, 106–128.
Nguyen, H., Mehrotra, R., & Sharma, A. (2016). Correcting for systematic biases in GCM simulations in the frequency domain. *Journal of Hydrology*, *538*, 117–126.
Nguyen, H., Mehrotra, R., & Sharma, A. (2017). Can the variability in precipitation simulations across GCMs be reduced through sensible bias correction? *Climate dynamics*, *49*(9-10), 3257–3275.
Nguyen, H., Mehrotra, R., & Sharma, A. (2019). Correcting systematic biases across multiple atmospheric variables in the frequency domain. *Climate dynamics*, *52*(1-2), 1283–1298.
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.
Percival, D. B., & Walden, A. T. (2000). *Wavelet methods for time series analysis*, (Vol. 4). Cambridge, England: Cambridge university press.
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. *Journal of hydrology*, *563*, 336–353.

Quilty, J., Adamowski, J., & Boucher, M. A. (2019). A stochastic data-driven ensemble forecasting framework for water resources: A case study using ensemble members derived from a database of deterministic wavelet-based models. *Water Resources Research*, *55*(1), 175–202. https://doi.org/10.1029/2018wr023205

Rashid, M. M., Johnson, F., & Sharma, A. (2018). Identifying sustained drought anomalies in hydrological records: A wavelet approach. *Journal of Geophysical Research: Atmospheres*, *123*(14), 7416–7432.

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 Resources Research*, *50*(12), 9721–9737.

Rössler, O. E. (1976). An equation for continuous chaos. *Physics Letters A*, *57*(5), 397–398.

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

Shafaei, M., & Kisi, O. (2016). Lake level forecasting using wavelet-SVR, wavelet-ANFIS and wavelet-ARMA conjunction models. *Water Resources Management*, *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. *Journal of Hydrology*, *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 Resources Research*, *50*(1), 650–660.

Sharma, A., Mehrotra, R., Li, J., & Jha, S. (2016). A programming tool for nonparametric system prediction using partial informational correlation and partial weights. *Environmental Modelling & Software*, *83*, 271–275.

Strang, G., & Nguyen, T. (1996). *Wavelets and filter banks*. Philadelphia, PA 19104 USA: Society for Industrial and Applied Mathematics.

Strogatz, S. H. (2018). *Nonlinear dynamics and chaos: With applications to physics, biology, chemistry, and engineering*. Boca Raton, Florida: CRC Press.

Taylor, K. E. (2001). Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research: Atmospheres*, *106*(D7), 7183–7192.

Torrence, C., & Compo, G. P. (1998). A practical guide to wavelet analysis. *Bulletin of the American Meteorological Society*, *79*(1), 61–78. https://doi.org/10.1175/1520-0477(1998)079<0061:Apgtwa>2.0.Co;2

Vu, M. T., Aribarg, T., Supratid, S., Raghavan, S. V., & Liong, S.-Y. (2016). Statistical downscaling rainfall using artificial neural network: significantly wetter Bangkok? *Theoretical and applied climatology*, *126*(3-4), 453–467.

Zarch, M. A. A., Sivakumar, B., & Sharma, A. (2015). Droughts in a warming climate: A global assessment of Standardized precipitation index (SPI) and Reconnaissance drought index (RDI). *Journal of Hydrology*, *526*, 183–195.