Journal of Hydrology 410 (2011) 1–12

Contents lists available at SciVerse ScienceDirect

Journal of Hydrology

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

Wavelet Auto-Regressive Method (WARM) for multi-site streamflow simulation of data with non-stationary spectra

Kenneth C. Nowak [a,b,c,*], Balaji Rajagopalan [a,b,c], Edith Zagona [a,b]

[a] Department of Civil, Environmental and Architectural Engineering, University of Colorado, Boulder, CO, United States
[b] Center for Advanced Decision Support for Water and Environmental Systems (CADSWES), Boulder, CO, United States
[c] Co-operative Institute for Research in Environmental Sciences (CIRES), University of Colorado, Boulder, CO, United States

ARTICLE INFO

Article history:
Received 2 November 2010
Received in revised form 13 August 2011
Accepted 25 August 2011
Available online 22 September 2011
This manuscript was handled by Andras Bardossy, Editor-in-Chief, with the assistance of Efrat Morin, Associate Editor

Keywords:
Stochastic simulation
Spectral simulation
Wavelet
Non-stationary spectra
Multi-site disaggregation
Colorado River

SUMMARY

Traditional stochastic simulation methods that are crafted to capture measures such as mean, variance and skew fail to reproduce significant spectral properties of the observed data. A growing body of literature indicates that many geo-physical data, especially streamflow, exhibit quasi-periodic and non-stationary variability driven by large scale climate features. Thus, methods which accurately model this behavior, in particular, the time evolution of variability, frequency of wet/dry epochs, etc. are crucial for risk assessment and management of water resources. In this paper, a Wavelet based Auto Regression Modeling (WARM) framework is proposed for data with significant non-stationary spectral features. This approach has four broad steps – (i) the wavelet transform of a time series is reconstructed as several periodic components based on dominant variability frequencies, (ii) scale averaged wavelet power (SAWP) is computed for each band to capture the time varying power and the components are scaled by this, (iii) Auto Regressive (AR) models fit to the scaled components and, (iv) simulations are performed from the AR models, rescaled and combined to obtain simulations of the original time series. Step (ii) is a new and unique departure from the WARM proposed by Kwon et al. (2007). We demonstrate this approach on annual streamflow at the Lee's Ferry gauge on the Colorado River. Furthermore, this is coupled with a spatial disaggregation method to generate streamflow ensembles at multiple locations upstream. We also show that this combination captures the spectral properties at several locations in a parsimonious manner.

Published by Elsevier B.V.

1. Introduction

For some time, it has been recognized that limited historic streamflow data is not sufficient information on which to base important water resources planning and management decisions. As such, approaches to identify a wider range of potential flow scenarios and variability have evolved over the years. Perhaps most rich in methods and application to water management is the field of stochastic streamflow simulation. These methods produce synthetic streamflow traces which reflect the statistical properties of the historic data, while often containing a wider range of flows and drought/surplus periods. Recently, studies examining the drivers of flow variability in the Western United States have identified quasi-periodic climatic forcings (e.g. El Nino Southern Oscillation, etc.) as having some role in modulating flow and other geophysical variability (McCabe et al., 2007; Piechota and Dracup, 1996; Rajagopalan et al., 2000; Tootle et al., 2005). While the understanding of these links is emerging, it is important to consider their influences in stochastic simulation. Of particular importance for water planning and management is the link to persistence of wet/dry epochs. Popular parametric and non-parametric methods (e.g. Auto Regressive Moving Average, K-Nearest Neighbor, etc.) (Lall and Sharma, 1996; Ouarda et al., 1997; Prairie et al., 2006; Salas and Obeysekera, 1982; Thomas and Fiering, 1962) have been developed over the years to capture the distributional and dependence properties of the historical data, but they are poor at capturing the observed data spectrum and consequently, the statistics of wet/dry epochs. Fig. 1 shows the global wavelet spectrum of a hypothetical synthetic streamflow data set with strong variance at the 16–32 year period (developed for demonstration purposes by combining a sine function with non-stationary amplitude and a noise time series). The gray region represents the 5th to 95th percentiles of spectra from 1000 stochastic traces generated to reflect the hypothetical streamflow data using an Auto-Regressive model. This failure to reproduce spectral properties in stochastic data can lead to inaccurate estimation of water resources system reliability and thus affect management and planning decisions.

* Corresponding author. Address: University of Colorado, 421 UCB, Boulder, CO 80309-0421, United States. Tel.: +1 303 492 0892; fax: +1 303 492 1347.
E-mail address: kenneth.nowak@colorado.edu (K.C. Nowak).

0022-1694/$ - see front matter Published by Elsevier B.V.
doi:10.1016/j.jhydrol.2011.08.051

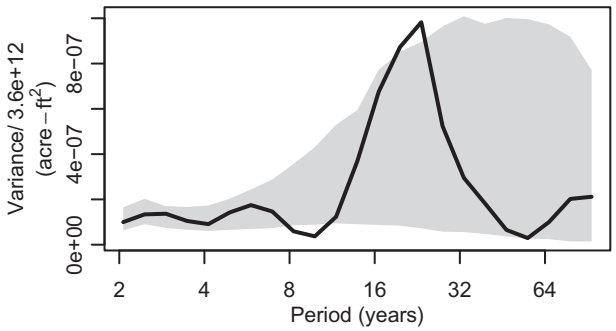

Fig. 1. Global wavelet spectrum for synthetic flow data (black line) and the 5th to 95th percentiles of the global wavelet spectra from 1000 traces simulated with an Auto Regressive (AR) model.

The work of Kwon et al., 2007 offers a frequency domain simulation method that not only reproduces traditional statistical measures of the historic data, but captures spectral characteristics as well. In their paper, the frequency domain method, Wavelet Auto Regressive Method (WARM), is shown to well reproduce the global spectrum of the observed data along with distributional properties such as mean, variance, skew and probability density function (PDF). WARM offers an attractive alternative to traditional time domain stochastic simulation methods. However, for data with non-stationary spectral characteristics, we find that WARM cannot capture this feature and as a result, the traces lack the non-stationarity present in the historical data. Furthermore, there is also a need to provide simulations at multiple nodes in a streamflow network.

In this paper we offer an enhanced WARM framework that provides stochastic simulations of streamflow traces at multiple locations, simultaneously capturing spectral and distributional properties present in the historic data. This generalized framework offers a flexible and robust multi-site streamflow simulation technique, which motivates this research. The paper is organized as follows. Section 2 provides an overview of the original WARM framework, motivation for improvements and the proposed enhancements. Section 3 contains results from application of the

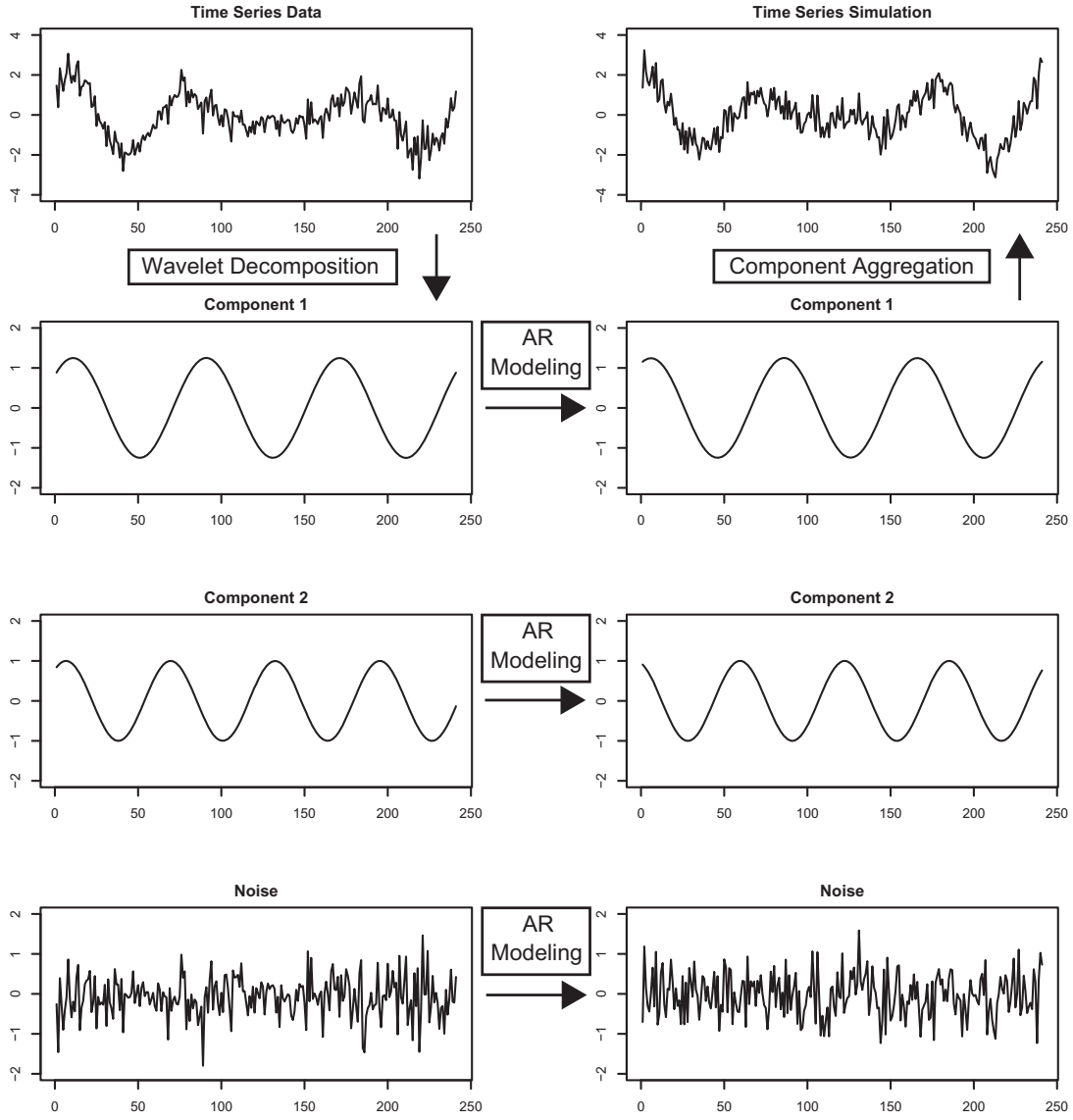

Fig. 2. Example diagram of WARM simulation procedure. The original data is decomposed via continuous wavelet transform using the Morlet wavelet and components are identified based on peaks in the global wavelet spectrum. Auto-Regressive models are fitted to each component as well as the residual noise term and each simulation is produced by combining the modeled component and noise signals.

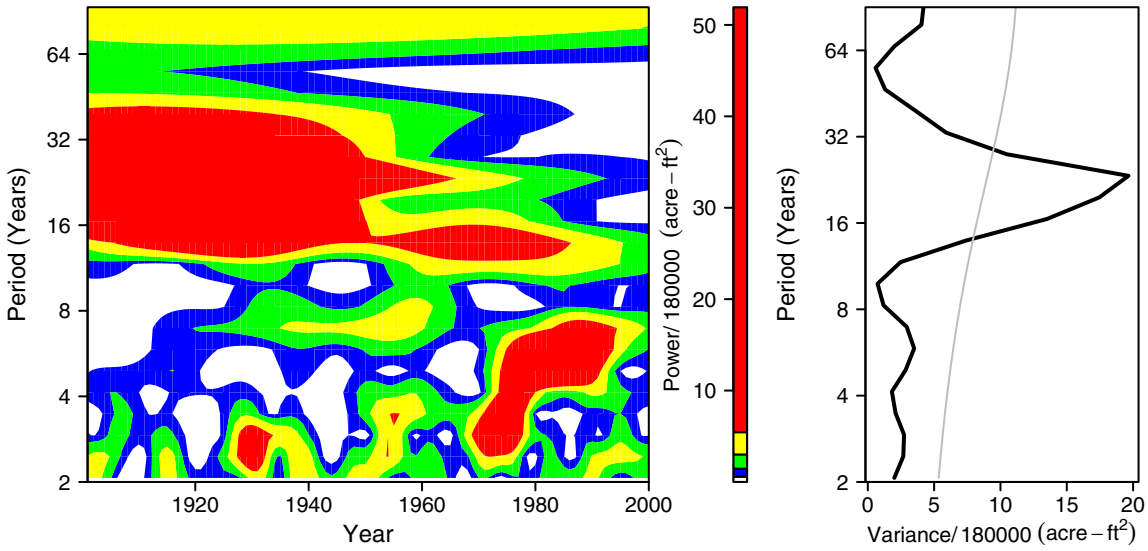

Fig. 3. Wavelet spectrum (left) and global wavelet spectrum (right) for synthetic flow data. The gray line is the 95% significance white noise spectrum.

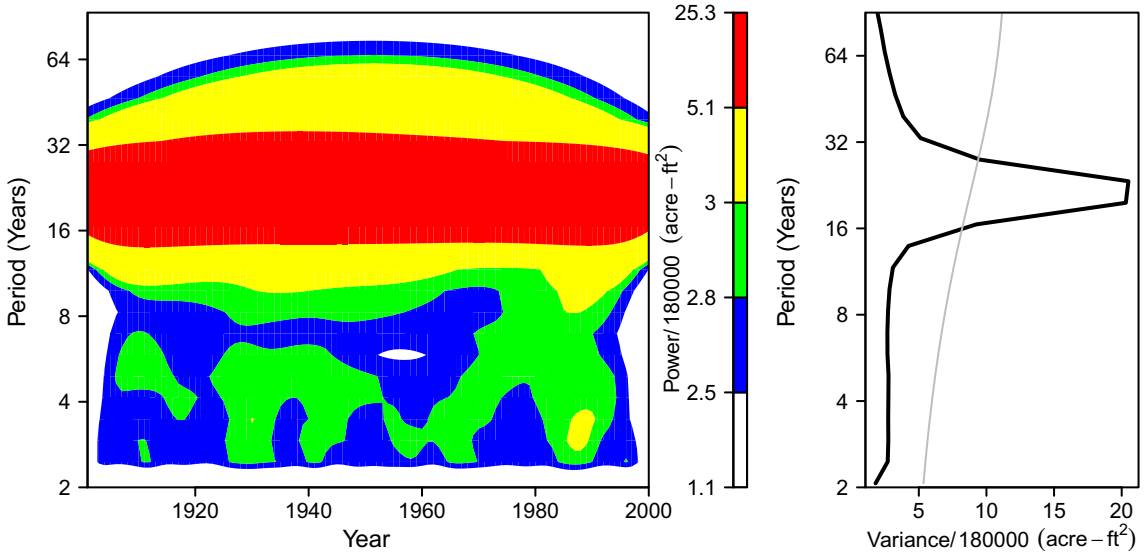

Fig. 4. Average wavelet spectrum (left) and average global wavelet spectrum (right) from 1000 WARM simulations for synthetic flow data. The gray line is the 95% significance white noise spectrum.

new method to streamflow in the Colorado River Basin. Section 4 is a summary discussion of the method and results.

2. Methodology

2.1. Brief introduction to WARM

This section constitutes a brief overview of wavelet spectral analysis and the original WARM framework (see Fig. 2). For additional detail, the reader is referred to Kwon et al., 2007. The use of wavelets in analyzing the time–frequency component of geophysical data has gained considerable popularity (Addison, 2002; Kwon et al., 2009; Torrence and Compo, 1998; Torrence and Webster, 1998). Wavelet methods have several advantages over traditional spectral estimation techniques; an example is the ability capture 'local' features such as amplitude modulation quite effectively. Additionally, wavelet-based methods have been shown to be efficient at revealing quasi-periodic signals in data with considerable noise – which is often the case with real data. For the purpose of introducing the WARM framework and motivating our enhancements, a synthetic flow data set is used, which by design contains underlying periodic variability at various time scales.

To begin, the WARM approach decomposes a time series into various components at a number of frequencies via the wavelet transform, thus giving the power (or variance) of the original data in the frequency and time domains. The continuous wavelet transform for some time series x_t is defined as:

$$X(a,b) = |a|^{-1/2} \int_{-\infty}^{+\infty} x_t \varphi^* \left(\frac{t-b}{a} \right) dt \tag{1}$$

where a is a scale parameter, b is the shift parameter and φ^* is the wavelet function (Kwon et al., 2007). The $(^*)$ denotes complex conjugate. We choose to employ the Morlet wavelet, given by:

$$\varphi(\eta) = \pi^{-1/4} \exp(i\omega_0 \eta) \exp(-\eta^2/2) \tag{2}$$

where ω_0 and η are non-dimensional frequency and time parameters respectively (Torrence and Compo, 1998). By substituting $\left(\frac{t-b}{a}\right)$ for η in Eq. (2), the shifted and dilated form of the mother wavelet is given (Addison, 2002; Kwon et al., 2009; Torrence and Compo, 1998; Torrence and Webster, 1998). Eq. (1) can be thought of as a series of convolutions between the wavelet function (Eq. (2)) and the original time series at all points for a variety of wavelet scales. To simplify the process, all convolutions can be completed simultaneously at a given scale by the convolution theorem. By doing so, the wavelet transform is defined as the inverse Fourier transform of the product of the data and the wave function in the Fourier space:

$$X_t(a) = \sum_{k=0}^{N-1} \hat{x}_k \hat{\varphi}^*(a\omega_k) \exp(i\omega_k t \delta_t) \tag{3}$$

where $\hat{x}_k$ is the discrete Fourier transform of $x(t)$, $\hat{\varphi}(a\omega)$ is the Fourier transform of the wavelet function, N is the number of points in the original data and ω_k is the angular frequency. The exponential term accomplishes the inverse transform (Torrence and Compo, 1998). Additional details on wavelet-based time series spectral estimation can be found in Torrence and Compo (1998).

A contour plot of the wavelet transform ($X_t(a)$) gives the wavelet spectrum at different frequencies over time (Fig. 3). To the right in this figure is the global wavelet spectrum, which shows average variance strength at each frequency across time. The global spectrum shows a strong peak in the 16–32 year period but it can be seen that this band is active only in the early epoch of the data.

In the WARM approach, the global spectrum is next used to identify significant spectral components of the wavelet transformed data. Frequencies corresponding to statistically significant spectral peaks are combined together to create several 'band passed' time series which are then back-transformed to the original flow space. The same is done for the residual frequencies as 'noise.' Significant peaks are often determined based on a white or red noise background global spectrum for a desired confidence level. Here, a 95% confidence white noise spectrum is used (Torrence and Compo, 1998). Alternatively, more subjective criteria could be employed, such as selecting the "El Nino band" given some a priori knowledge that there is a strong link between the data and that particular phenomenon. This choice will likely vary from individual to individual and application to application. The reconstructed (back-transformed) original time series is given as:

$$x_t = \frac{\delta_j \delta_t^{1/2}}{C_\delta \psi_0(0)} \sum_{j=0}^{J} \frac{\Re\{X_t(a_j)\}}{a_j^{1/2}} \tag{4}$$

which constitutes the sum of the real part of the wavelet transform over all scales ($j = 1{:}J$). δ_j and δ_t are the scale averaging coefficient and sampling period respectively. From Eq. (3), $X_t(a)$ is the wavelet transform of the data and a is the scale parameter introduced earlier. The remaining terms (C_s and $\psi_0(0)$) are empirically derived reconstruction factors specific to the Morlet wavelet (Torrence and Compo, 1998). A "band specific" reconstruction is computed by simply limiting the range of scales (j) over which the summation is performed (Eq. (4)).

By design, the 'band passed' time series components add up to the original data. To complete the WARM simulation, each of the components (including residual noise term) are modeled using traditional time series models such as an Auto-Regressive Moving Average (ARMA) model (Bras and Rodríguez-Iturbe, 1985; Salas, 1980). The reconstructed components (except the noise) are smooth and hence can be effectively modeled as lower order AR models. A simulation is then produced by generating traces from each component model and then summing them together. This process is outlined in Fig. 2.

2.2. WARM drawbacks

From the results of Kwon et al. (2007) it can be seen that the WARM framework, reproduces the global wavelet spectrum quite well for a variety of datasets. In order to assess the efficacy of WARM simulations to reproduce key spectral features, a composite wavelet spectrum of 1000 WARM simulations based on the synthetic data is presented in Fig. 4. It can be seen that the simulations capture the global spectrum very well but, the variability contributing to the significant peak in the global spectrum (which is only active in the first half of the century, see Fig. 3) is smoothed across time, thus failing to capture the non-stationary spectral feature. This inability can have significant impact on the epochal behavior of the wet/dry sequences and consequently estimates of water resources system risk, when coupled with a decision model. Thus, enhancing the WARM approach to address this drawback is our first motivation of this study and our proposed enhancement is described below.

2.3. Proposed enhancement to WARM

The reconstruction of the example data in the dominant spectral band, (16–32 year period, Fig. 5) using Eq. (4), shows reduced variability in recent decades. Also shown in this figure is the Scale Averaged Wavelet Power (SAWP) (Kwon et al., 2007) of this dominant band. The SAWP time series reflects the temporal variability of the strength of this band, which is given as:

$$\bar{X}_t^2 = \frac{\delta_j \delta_t}{C_\delta} \sum_{j=j_1}^{j_2} \frac{|X_t(a_j)|^2}{a_j} \tag{5}$$

where j_1 and j_2 are the scales over which averaging is computed.

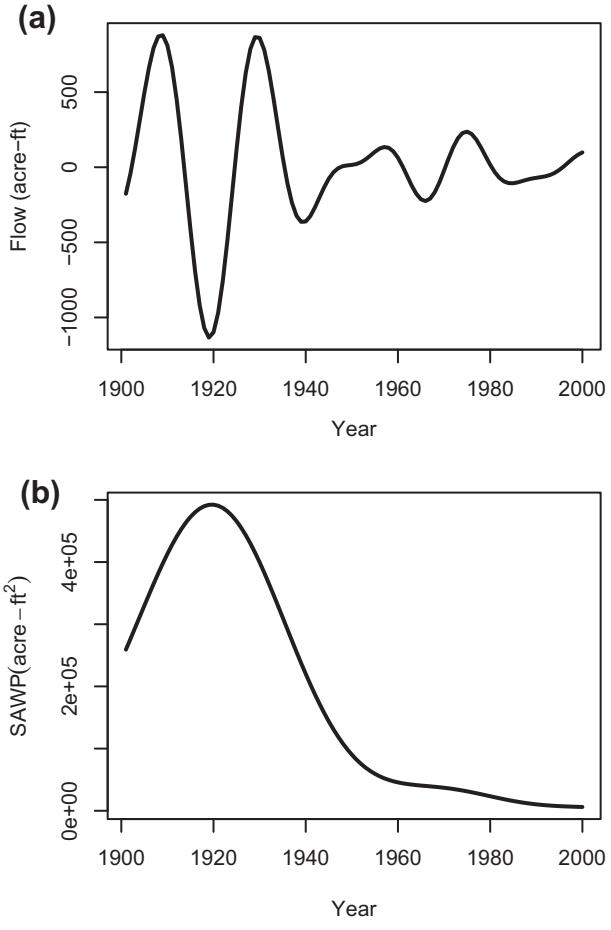

Fig. 5. (a) 25 year Component of synthetic flow data. (b) Scale averaged wavelet power (SAWP) for 25 year component of synthetic flow data.

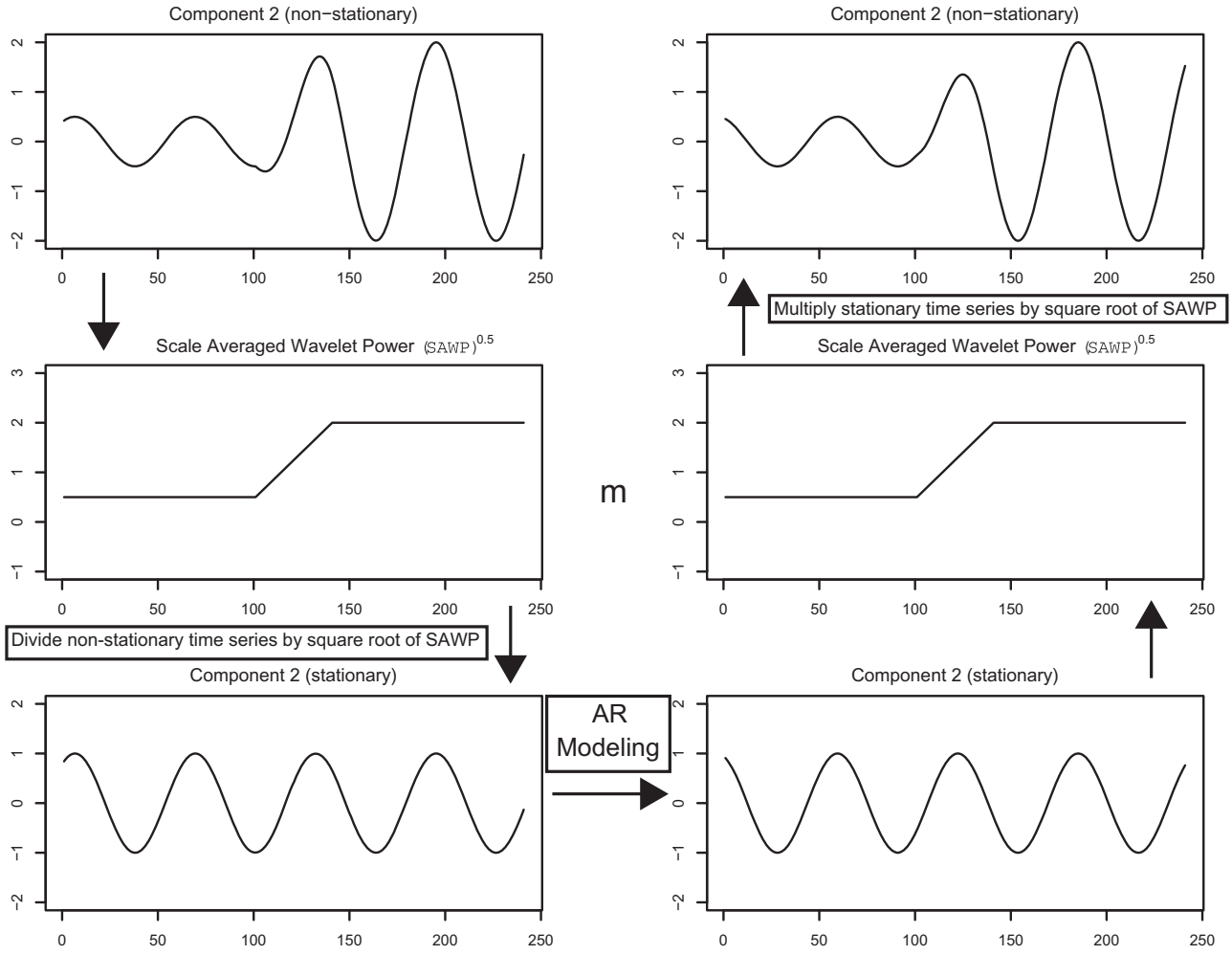

Fig. 6. Schematic of improved WARM simulation procedure, shown for a single, non-stationary component.

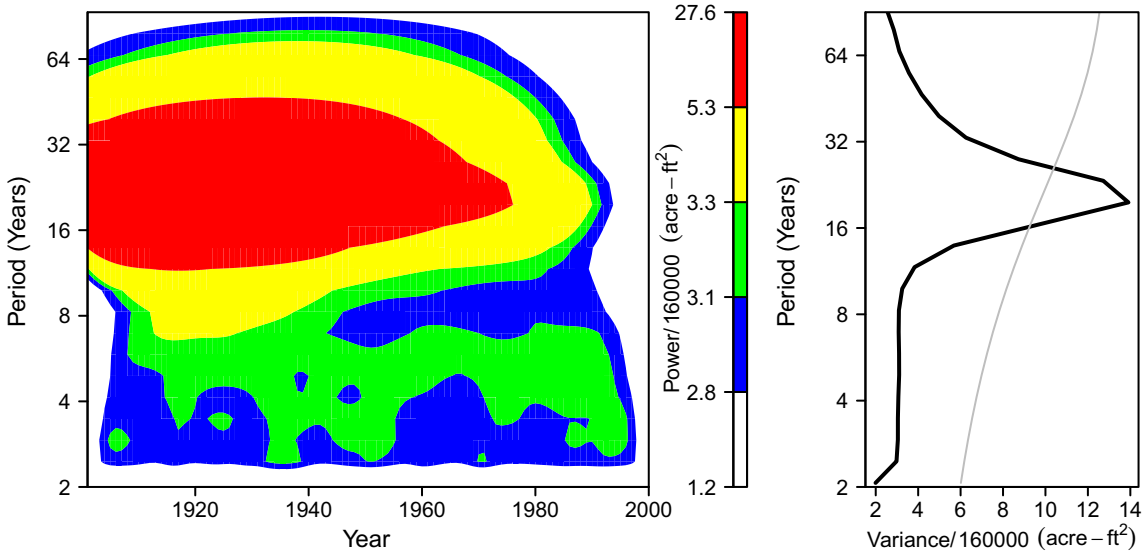

Fig. 7. Average wavelet spectrum (left) and average global wavelet spectrum (right) from 1000 improved WARM simulations for the synthetic flow data. The gray line is the 95% significance white noise spectrum.

To provide the ability to capture non-stationary spectral features we propose the following enhancement to WARM using the SAWP.

(i) The reconstructed, non-stationary components, are divided by its $\bar{X}_t$ (square root of SAWP from Eq. (5)), thus producing a new, stationary time series.

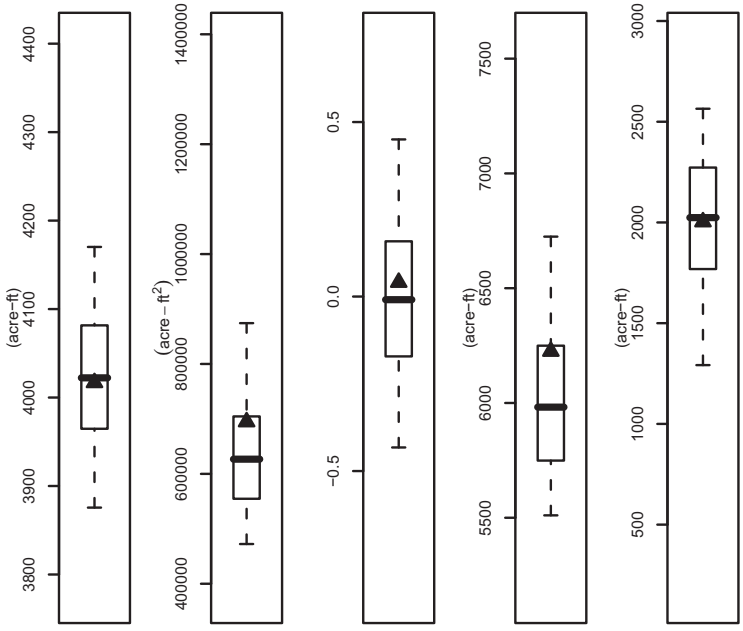

Fig. 8. Boxplots of mean, variance, skew, max and min values from 1000 improved WARM simulations for the synthetic flow data.

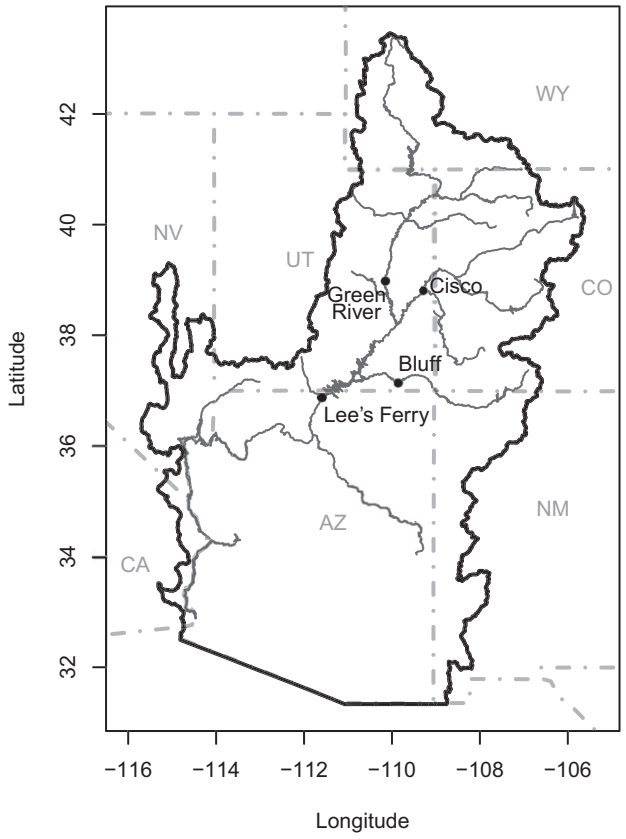

Fig. 9. Colorado River Basin map with gauging locations for disaggregation annotated; Green River at Green River, UT, Colorado River near Cisco, UT, San Juan River near Bluff, UT and Colorado River at Lee's Ferry, AZ.

(ii) These new stationary versions can now be effectively modeled using AR models – as they satisfy the stationary assumptions. The fitted AR models are used to simulate the components – per the WARM approach.

(iii) The non-stationarity is re-introduced via multiplication with $\bar{X}_t$.–essentially reversing step (ii).

These enhancements to the WARM are outlined in Fig. 6. The average global and local wavelet power spectra for the synthetic data based on 1000 "enhanced WARM" traces are seen in Fig. 7. Additionally, Fig. 8 shows basic distributional statistics from the 1000 "enhanced WARM" simulations. The average local power spectrum reflects the non-stationary features seen in Fig. 3 quite well. All of the historic statistics are well reproduced by the boxplots of the 1000 traces, thus demonstrating the ability of this approach to capture the temporal aspect of variability modes without compromising other measures.

2.4. Multisite streamflow simulation

It is well established that for the purpose of water management and planning, particularly on a large river network with competing demands, flow data is required at multiple locations that preserve spatial and temporal dependencies. As such, it is of considerable interest to extend the above described enhanced WARM to simulate streamflow at multiple sites. To accomplish this, the "enhanced WARM" is first applied to an aggregate streamflow (often the downstream gauge on a network). Next, the simulated flows are then disaggregated in space using the proportion method by Nowak et al. (2010).

The disaggregation method of Nowak et al. (2010) is a simple, non-parametric method which has been shown to be effective at simulating streamflow at a variety of spatial and temporal (monthly, daily) resolutions. For the sake of completion we summarize the steps.

(i) Often flow at a downstream gauge is 'aggregate' of the upstream flows. Sometimes a synthetic aggregate gauge is created by summing all the spatial flows.
(ii) For each historic year the 'proportion' vector – i.e., the proportion of aggregate flow at each spatial location, is computed.
(iii) The aggregate flow (at the downstream location) is first generated, e.g., using the WARM approach described above. Analog years for each flow value are identified from the his-

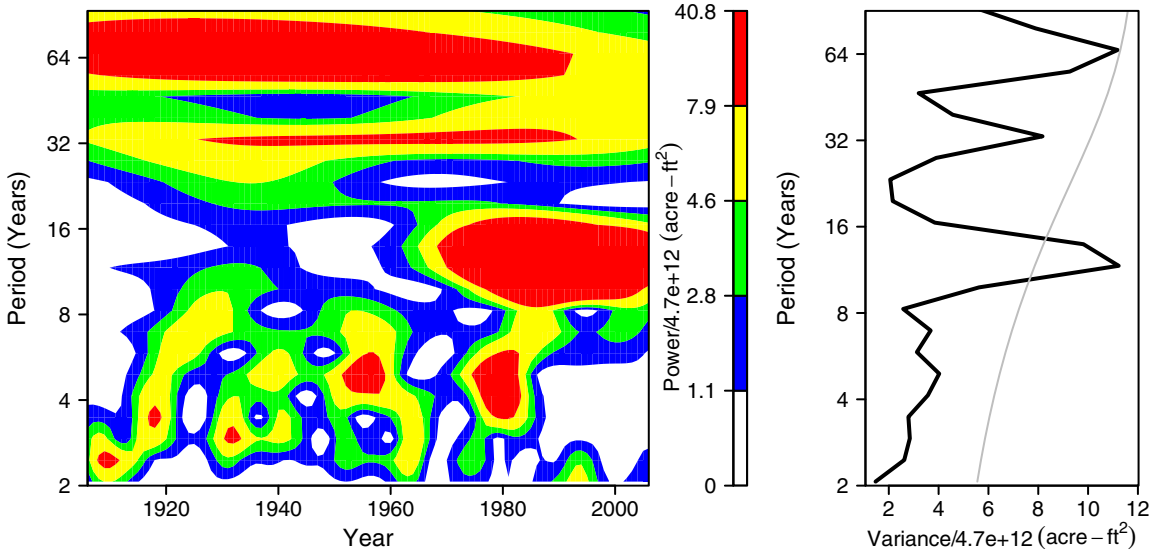

Fig. 10. Wavelet spectrum (left) and global wavelet spectrum (right) for Lee's Ferry, AZ natural flow data. The gray line is the 95% significance white noise spectrum.

toric record and one is resampled using a probability metric that gives high weight to the nearest neighbors and least to the farthest. The proportion vector corresponding to the selected neighbor is then applied to the aggregate flow to obtain disaggregated flows at the spatial locations. We refer the reader to the full paper for additional details on the disaggregation methodology(Nowak et al., 2010).

This produces streamflow scenarios at all the desired locations on the network, which we posit, preserves summability, distributional and spectral properties. We apply this to streamflow from four locations on the Colorado River basin, described below.

3. Results from application to Colorado River Basin

3.1. Data and study area

The study area for the demonstration of the enhanced WARM approach is the Upper Colorado River Basin (Fig. 9). This region incorporates parts of Colorado, Utah, Wyoming and small portions of Arizona and New Mexico. The gauge at Lee's Ferry, AZ aggregates all of the Upper Basin flow. Our proposed methodology is first applied to simulate flow for this location, which is then disaggregated to three major upstream tributaries on the network. These locations are Green River at Green River, UT, Colorado River near Cisco, UT and San Juan River near Bluff, UT (Fig. 9). Each represents a significant drainage of the Upper Colorado River Basin.

In this work, annual water year (October–September) natural streamflow data for the Colorado River Basin from the period 1906–2006 are employed. This dataset is developed and updated regularly by the United States Bureau of Reclamation. Naturalized streamflow are computed by removing anthropogenic impacts (i.e., reservoir regulation, consumptive water use, etc.) from the recorded historic flows. Prairie and Callejo (United States Department of the Interior, 2005) present a detailed description of methods and data used for the computation of natural flows in the Colorado River Basin.

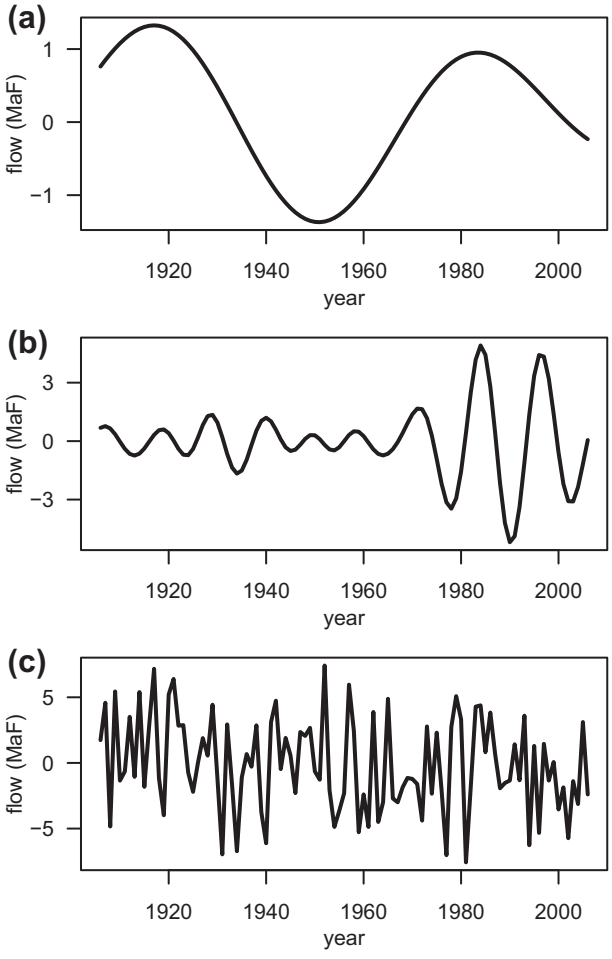

Fig. 11. Re-constructed components (a) low frequency signal, (b) decadal signal and (c) noise.

3.2. Simulations at Lee's Ferry, AZ

The first step, as discussed earlier, is to compute the wavelet spectrum of Lee's Ferry streamflow and identify the significant spectral bands. Fig. 10 shows the wavelet spectrum and we identified significant bands in the 9–16 period and 50–70 year period. The significance of the very low frequency band needs to be taken

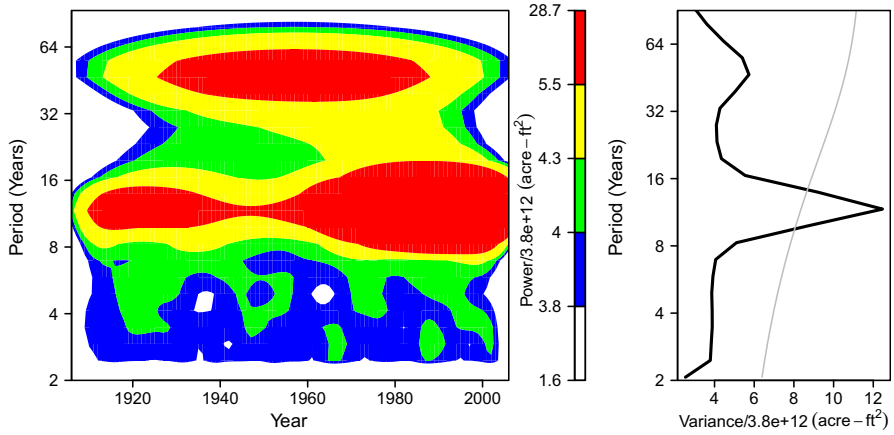

Fig. 12. Average wavelet spectrum (left) and average global wavelet spectrum (right) from 1000 improved WARM simulations for natural flow at Lee's Ferry, AZ. The gray line is the 95% significance white noise spectrum.

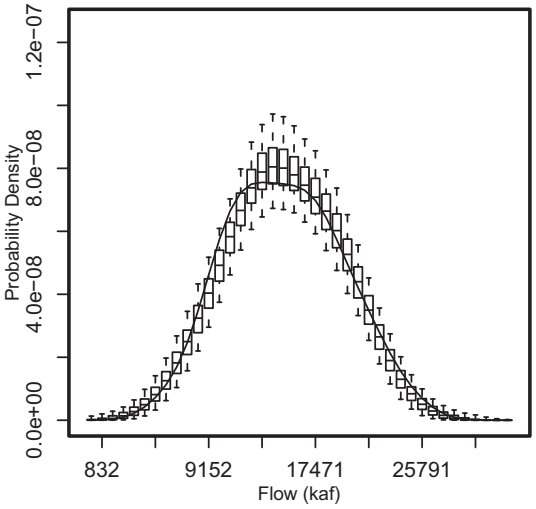

Fig. 13. Improved WARM Lee's Ferry flow PDFs shown as boxplots with historic flow PDF as solid black line.

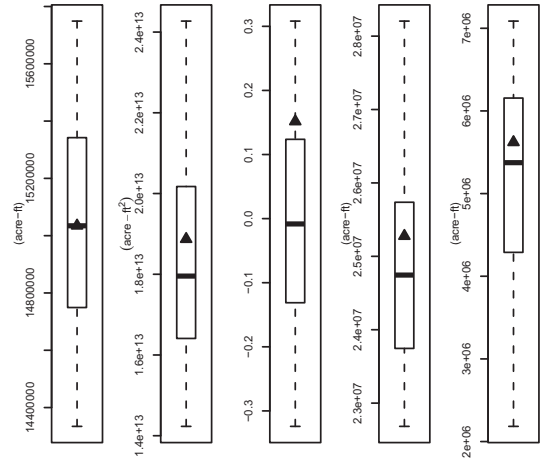

Fig. 14. Boxplots of mean, variance, skew, max and min values from 1000 improved WARM simulations for Lee's Ferry, AZ natural flow. The box of the boxplot represents the interquartile range and whiskers extend from the 5th to 95th percentiles. Black triangles are values from observed data.

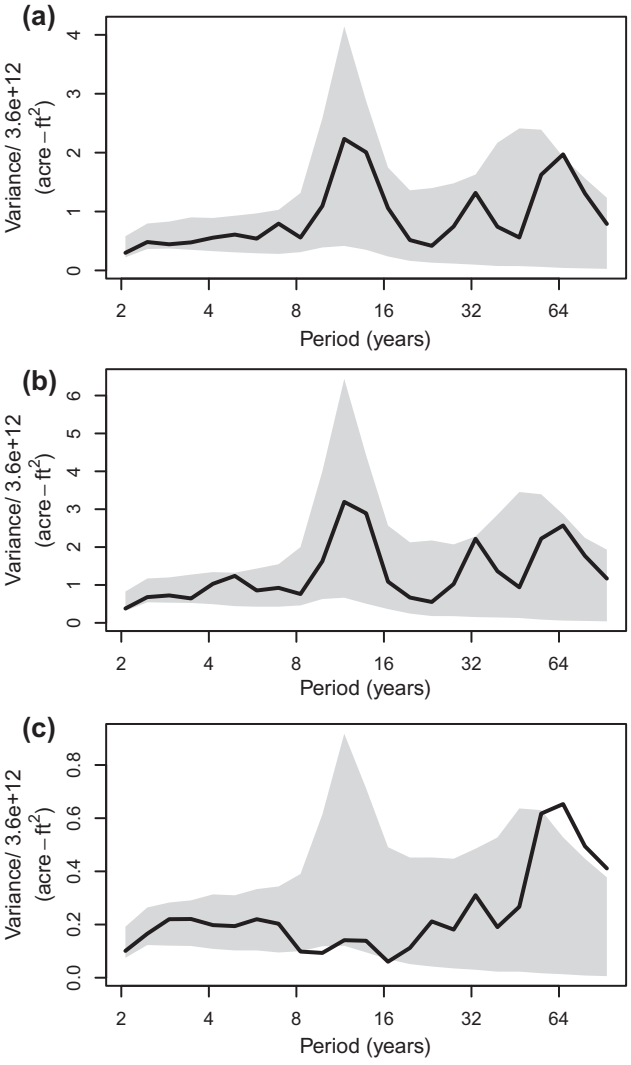

Fig. 15. Global wavelet spectra for (a) Green River at Green River, UT, (b) Colorado River near Cisco, UT and (c) San Juan River near Bluff, UT. Black line is historic spectrum and gray region is 5th to 95th percentile based on spectra from disaggregation of 1000 improved WARM traces.

with caution due to smaller sample size and boundary effects. It can also be seen that the significant decadal band appears just in

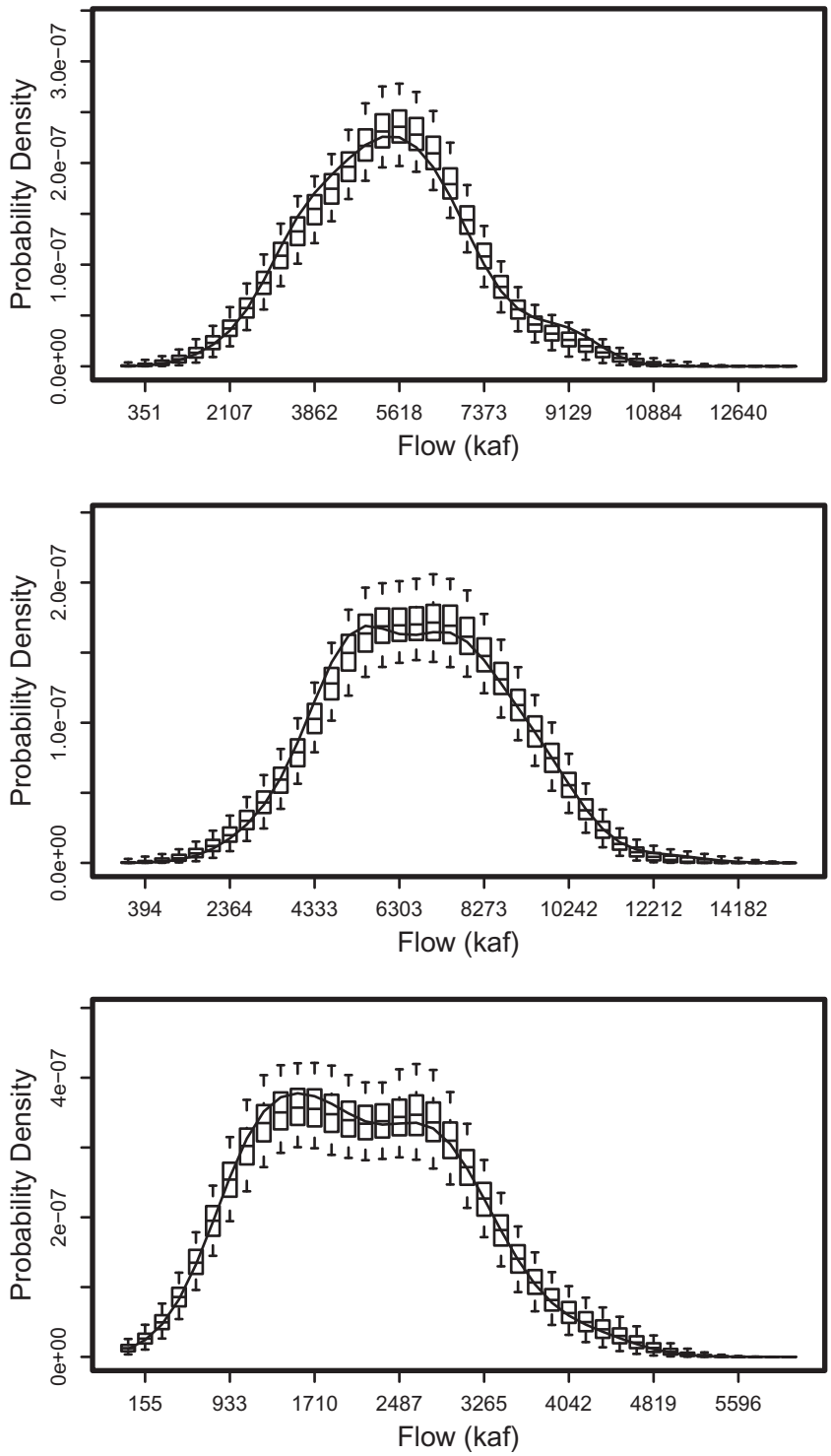

Fig. 16. Disaggregated, improved WARM simulation flow PDFs for (a) Green River at Green River, UT, (b) Colorado River near Cisco, UT and (c) San Juan River near Bluff, UT. The black line is the observed PDF and the boxplots show the range of simulation PDFs.

recent decades, indicating a strong non-stationarity in the spectrum, thus indicating the need for "enhanced WARM." In all, we have three spectral bands including the noise component shown in Fig. 11. The proposed enhanced WARM was applied and 1000 streamflow ensembles 101 years in length were generated (length of Lee's Ferry natural flow data set). The combined spectrum of these simulations can be seen to capture all the non-stationary spectral features presented in the historic data (Fig. 12). Boxplots of probability density functions (PDFs) for annual flow simulations along with that of the historic data are shown in Fig. 13 – it can be seen that they are very well captured by the simulations. Additionally, distributional statistics such as mean, variance, coefficient of skew and maximum/minimum values from the simulations are well reproduced (see boxplots in Fig. 14), with the exception of skew (which is very small) falling just outside the boxplot box. It is well established that Auto-Regressive models do not reproduce non-zero skew without data transformation. Thus, given the very small skewness in the data we did not deem it necessary to explore

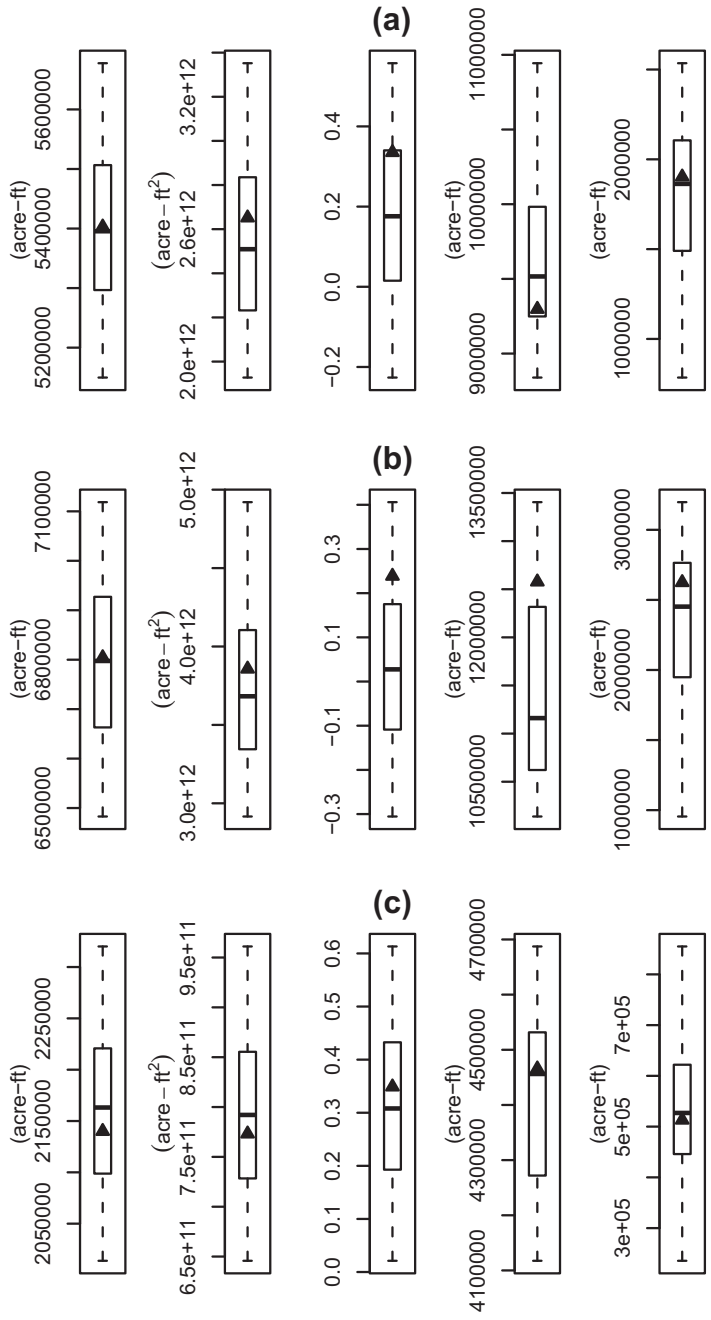

Fig. 17. Statistics for disaggregated improved WARM simulations for (a) Green River at Green River, UT, (b) Colorado River near Cisco, UT and (c) San Juan River near Bluff, UT. Measures are mean, variance, skew, maximum and minimum (left to right). The box of the boxplot represents the interquartile range and whiskers extend from the 5th to 95th percentiles. Black triangles are values from observed data.

this. Perhaps nonlinear AR models for the components might help in correcting this.

3.3. Multi-site simulations

The simulated streamflow at Lee's Ferry from the previous step were disaggregated using the proportion disaggregation method of Nowak et al. (2010) to generate streamflow ensembles at the three upstream locations on the network. The range of global wavelet spectra from the simulations are computed at the three locations and shown in Fig. 15 along with the respective historic spectra. At the Green River and Cisco locations, the observed spectra are very well captured by the simulations. Furthermore, these sites exhibit the non-stationary in the decadal band seen in Fig. 10 in Lee's Ferry flow. These features are also reflected in the disaggregated traces (not shown). The San Juan spectrum, however, is not well reflected by the disaggregated results. The spectral properties of the disaggregated streamflow tend to be inherited from the aggregate flow location. The larger upstream tributaries typically share in the same spectral signature, but smaller tributaries such as the San Juan can differ, particularly when the sub-basin is in a geographically or climatologically different region. The distributional properties on the other hand, are very well preserved at all the three locations as shown by both the PDFs (Fig. 16) and suite of statistical measures (Fig. 17).

4. Summary/discussion

An enhanced Wavelet Auto Regressive Method (WARM) is coupled with a flexible and proven disaggregation approach to

produce stochastic streamflow at multiple locations that accurately reflect the historic spectral and distributional properties. The enhancements are substantial and afford the ability to model and simulate data with non-stationary spectral characteristics, which properly reflect the changing persistence of wet/dry conditions that are important for subsequent risk estimations.

From the results section, the disaggregated spectrum for the San Juan River merits further discussion. It is clear that the observed spectrum is not completely reflected in the simulations. From additional analysis, it was found that the spectral pattern common among the Lee's Ferry, Cisco and Green River locations is present throughout most of the upper Colorado Basin; they are likely impacted by similar large scale climate forcings. San Juan on the other hand, can exhibit variability of both the Upper and Lower Basins and hence a different spectral signature. Thus, it is important to have some a priori knowledge of the characteristics of the basin in which the method is being applied. For most small basins/sub basins the flow variability is likely to be spatially homogeneous and have similar spectral features. However, for a large basin, such as the upper Colorado Basin, which spans five states and encompasses extremely varied terrain, it is possible to have sub regions with substantially different spectral signatures. Apart from imparting "extra" decadal variability, the simulations capture all of the distributional properties and the additional low frequency persistence may be desirable in planning by generating longer wet and dry sequences.

As a potential modification for better capturing the spectral features of a sub basin that does not conform to the "standard basin spectral signature" (e.g. the San Juan), the following might be considered. The flow contribution of this "non-conforming sub basin" to downstream reaches would be removed and modeled independent of the rest of the basin. These traces would then be combined with simulations made for the remainder of the network in order to make the system "whole." This should ensure that historic spectral features are well reflected in the synthetic data. The drawback of this approach is that cross-correlation between the rest of the basin and the independently modeled portion may not be explicitly preserved. However, for a site that does not share in the standard basin spectral signature, cross-correlation with other sites is likely to be weak. As such, preservation of this statistic may not be critical, in favor of improving spectral features.

From working with a variety of data and this simulation approach, a few additional implementation comments are offered. If the noise component exhibits non-normal features it might be better to use a bootstrap approach to capture its distribution. This was the case for the Lee's Ferry data used in this study and the bootstrap was quite effective at improving the overall result.

When combining the noise and component terms to produce a trace, it was found that overall variance can be slightly under-simulated. This was determined to be a result of weak correlation between the original components and noise. As such, when combined, the aggregate variance was greater than the sum of each term's variance. Thus, by simulating each term independently, the weak correlation is lost and total variance may have a tendency to be slightly low. None of the observed correlation values between terms proved to be significant but the small correlations is likely due to the periodic nature of the components. To address this, we look to basic probability theory for correlated random variables (Montgomery and Runger, 2003). The variance of a time series produced by combining two correlated random variables is given by Eq. (6):

$$\sigma^2_{x+y} = \sigma^2_x + \sigma^2_y + 2\sigma_{xy} \tag{6}$$

where x and y are the correlated random variables, ^{2} is variance and $_{xy}$ is their covariance. Thus, once the three simulated components are combined, the new time series is multiplied by a variance correction factor (v_f) given by:

$$v_f = 1 + (2\sigma_{xy}/\sigma^2_{x+y}) \tag{7}$$

which corrects the variance to that of the original data. It is important to note that this may not always be a problem, and is likely dependant upon both the original data and the scales combined during the component reconstruction. However, should this issue present itself, we find this to be a simple and effective way to address the problem – compared to other alternatives such as multivariate Auto-Regressive models which can be unwieldy. Alternatively, exploring the use of a discrete wavelet transform may help with these or similar issues due to the completely orthogonal nature of that transform, whereas the continuous transform can be redundant at times.

The enhancements to the WARM approach offered in this work are unique and substantial improvements to the original framework, particularly for the purpose of water resources planning and management. Additionally the discussion provides insight on applications and addressing potential issues such as under-simulation of variance and failure to re-produce spectral features when coupled with the disaggregation method. Last, the ability to properly model the time-varying strength of significant modes of variability constitutes an effective method by which to simulate changes in low frequency persistence.

Acknowledgements

Funding for this research by Bureau of Reclamation and the Western Water Assessment RISA program at the University of Colorado is gratefully acknowledged. Thanks are also due to the Center for Advanced Decision Support in Water and Environmental Systems (CADSWES) at the University of Colorado, Boulder for use of its facilities and computational support.

References

Addison, P., 2002. The Illustrated Wavelet Transform Handbook. Taylor & Francis, New York.

Bras, R.L., Rodríguez-Iturbe, I., 1985. Random functions and hydrology, vol. xv. Addison-Wesley, Reading, Mass, 559 p.

Kwon, H.H., Lall, U., Khalil, A.F., 2007. Stochastic simulation model for nonstationary time series using an autoregressive wavelet decomposition: applications to rainfall and temperature. Water Resources Research 43 (5).

Kwon, H.H., Lall, U., Obeysekera, J., 2009. Simulation of daily rainfall scenarios with interannual and multidecadal climate cycles for South Florida. Stochastic Environmental Research and Risk Assessment 23 (7), 879–896.

Lall, U., Sharma, A., 1996. A nearest neighbor bootstrap for resampling hydrologic time series. Water Resources Research 32 (3), 679–693.

McCabe, G.J., Betancourt, J.L., Hidalgo, H.G., 2007. Associations of decadal to multidecadal sea-surface temperature variability with Upper Colorado River flow. Journal of the American Water Resources Association 43 (1), 183–192.

Montgomery, D.C., Runger, G.C., 2003. Applied Statistics and Probability for Engineers, vol. xiv. Wiley, New York, 706 p.

Nowak, K., Prairie, J., Rajagopalan, B., Lall, U., 2010. A nonparametric stochastic approach for multisite disaggregation of annual to daily streamflow. Water Resources Research 46.

Ouarda, T., Labadie, J., Fontane, D., 1997. Indexed sequential hydrologic modeling for hydropower capacity estimation. Journal of the American Water Resources Association 33 (6), 1337–1349.

Piechota, T.C., Dracup, J.A., 1996. Drought and regional hydrologic variation in the United States: associations with the El Nino Southern Oscillation. Water Resources Research 32 (5), 1359–1373.

Prairie, J.R., Rajagopalan, B., Fulp, T.J., Zagona, E.A., 2006. Modified K-NN model for stochastic streamflow simulation. Journal of Hydrologic Engineering 11 (4), 371–378.

Rajagopalan, B., Cook, E., Lall, U., Ray, B.K., 2000. Spatiotemporal variability of ENSO and SST teleconnections to summer drought over the United States during the twentieth century. Journal of Climate 13 (24), 4244–4255.

Salas, J.D., 1980. Applied Modeling of Hydrologic Time Series, vol. xiv. Water Resources Publications, Littleton, Colo., 484 p.

Salas, J.D., Obeysekera, J.T.B., 1982. Arma model identification of hydrologic time-series. Water Resources Research 18 (4), 1011–1021.

Thomas, H., Fiering, M., 1962. Mathematical synthesis of streamflow sequences for the analysis of river basins by simulation. In: Maass, A. et al. (Eds.), Design of Water Resource Systems. Harvard University Press, Cambridge.

Tootle, G.A., Piechota, T.C., Singh, A., 2005. Coupled oceanic-atmospheric variability and US streamflow. Water Resources Research 41 (12).

Torrence, C., Compo, G.P., 1998. A practical guide to wavelet analysis. Bulletin of the American Meteorological Society 79 (1), 61–78.

Torrence, C., Webster, P.J., 1998. The annual cycle of persistence in the El Nino Southern Oscillation. Quarterly Journal of the Royal Meteorological Society 124 (550), 1985–2004.

United States Department of the Interior, 2005. Natural flow and salt computation methods, Bureau of Reclamation, Salt Lake City, UT.