Stoch Environ Res Risk Assess (2009) 23:879–896
DOI 10.1007/s00477-008-0270-2

ORIGINAL PAPER

Simulation of daily rainfall scenarios with interannual and multidecadal climate cycles for South Florida

Hyun-Han Kwon · Upmanu Lall · Jayantha Obeysekera

Published online: 2 October 2008

Abstract Concerns about the potential effects of anthropogenic climate change have led to a closer examination of how climate varies in the long run, and how such variations may impact rainfall variations at daily to seasonal time scales. For South Florida in particular, the influences of the low-frequency climate phenomena, such as the El Nino Southern Oscillation (ENSO) and the Atlantic Multi-decadal Oscillation (AMO), have been identified with aggregate annual or seasonal rainfall variations. Since the combined effect of these variations is manifest as persistent multi-year variations in rainfall, the question of modeling these variations at the time and space scales relevant for use with the daily time step-driven hydrologic models in use by the South Florida Water Management District (SFWMD) has arisen. To address this problem, a general methodology for the hierarchical modeling of low- and high-frequency phenomenon at multiple rain gauge locations is developed and illustrated. The essential strategy is to use long-term proxies for regional climate to first develop stochastic scenarios for regional climate that include the low-frequency variations driving the regional rainfall process, and then to use these indicators to condition the concurrent simulation of daily rainfall at all rain gauges under consideration. A newly developed methodology, called Wavelet Autoregressive Modeling (WARM), is used in the first step after suitable climate proxies for regional rainfall are identified. These proxies typically have data available for a century to four centuries so that long-term quasi-periodic climate modes of interest can be identified more reliably. Correlation analyses with seasonal rainfall in the region are used to identify the specific proxies considered as candidates for subsequent conditioning of daily rainfall attributes using a Non-homogeneous hidden Markov model (NHMM). The combined strategy is illustrated for the May–June–July (MJJ) season. The details of the modeling methods and results for the MJJ season are presented in this study.

Keywords Rainfall simulation · Low frequency · Wavelet autoregressive model · Non-homogeneous hidden Markov model · Climate variability

H.-H. Kwon (✉)
Water Resources Division, Korea Institute of Construction Technology, Simindae-Ro, Ilsanseo-Gu, Goyang-Si, Kyeonggi-Do, South Korea
e-mail: hkwon@kict.re.kr

U. Lall
Department of Earth and Environmental Engineering, Columbia University, 918. S.W Mudd Building, West 120th St., New York, NY, USA

J. Obeysekera
Hydrologic and Environmental Systems Modeling, South Florida Water Management District, 3301 Gun Club Road, West Palm Beach, FL, USA

1 Introduction

The South Florida area is considered semi-tropical, with a hot and humid wet season from May to October and a mild dry season from November to April. The wet season is characterized by frequent thunderstorms and severe tropical storms. The wet seasons in South Florida start with intense rainfalls in mid-May or June, followed by a period of minor precipitation events, and then by heavy tropical rainfalls in late August or early September (McPherson and Halley 1996; Van Lent 1993).

Recently, it has been acknowledged that rainfall may correspond to a long memory process (Enfield et al. 2001).

There is evidence of spatio-temporal organization in the form of band-limited, episodic, quasi-oscillatory variations, such as the El Nino Southern Oscillation (ENSO) and the Atlantic Multi-decadal Oscillation (AMO) (Kwon et al. 2006b; Park and Mann 2000). Trimble et al. (1998) and Trimble and Trimble (1998) indicate that ENSO and other climate indices have significant potential to predict net inflow volumes into Lake Okeechobee. ENSO events typically have a 3–7-year cycle, with more pronounced climate effects during the winter months and less certain climate effects during the wet season summer months. In recent years, Artificial Neural Network (ANN) models have been employed by the South Florida Water Management District (SFWMD) for predicting the net inflow volumes into Lake Okeechobee using ENSO and other climatic indices (Trimble et al. 1997; Trimble and Trimble 1998; Zhang and Trimble 1996).

The AMO is a long-range climatic oscillation that causes cyclical changes in the surface temperature of the Atlantic Ocean, between the Equator and Greenland that may persist for years or even decades. AMO describes temperature deviations ("warm" and "cold") in the ocean surface that appear to be related to shifts in South Florida's climate (Enfield et al. 2001). There are significant year-to-year fluctuations in the ocean temperature records, but no clear indications of the exact years when AMO switches from a warm phase to a cold phase.

The rainfall data for the South Florida region exhibit long-memory or regimes with quasi-oscillatory behavior that may derive from low frequency Pacific and Atlantic Ocean climate modes (Kwon et al. 2006b). The connection of these low-frequency climate modes to the Lake Okeechobee inflows and the South Florida rainfall has been noted by Enfield et al. (2001), Schmidt et al. (2001), Trimble et al. (1998), Trimble and Trimble (1998), and Zhang and Trimble (1996). However, the large-scale mechanisms for wet-season rainfall in South Florida are not nearly as well understood as those for the dry season (November–April), and much of the literature suggests that rainfall in this season is dominated by local convection and not by large-scale mechanisms (Kwon et al. 2008). Kwon et al. (2008) investigated the synoptic patterns of extreme rainfalls in the wet season. The patterns suggested that those rainfall events may be related to changes in the easterly trade winds and circulation around the Bermuda high. These changes are coincident with the changes expected in the winds, convection and sea surface temperature (SST) over the Eastern equatorial Pacific and the middle to Western equatorial Atlantic.

Daily and monthly rainfall scenarios are simulated by the SFWMD as part of a risk management strategy for water operations and system-wide planning. These scenarios are based on the historical record, assuming that each year is equally likely to recur randomly. Daily rainfall probabilities, the persistence of wet and dry regimes, and other rainfall statistics can vary substantially over time in a systematic way. Climatic features, such as the ENSO, the AMO and the Pacific Decadal Oscillation (PDO), have been associated with such persistent year-to-year changes in rainfall probabilities. These climate modes also affect rainfall statistics in different ways in each season. Generating daily and monthly rainfall scenarios that depend realistically on the slowly varying climate is a challenge, particularly considering the AMO which seems to have a 50–60-year period. Thus, it is difficult to project whether rainfall in the next few years will continue to follow the emerging AMO trend and, if so, with what probability.

While there is limited historical data to calibrate low-frequency climate variability, especially the AMO, proxy data (e.g., corals) have been used to extend the record of climate indices to the preceding 400–500 years. Proxy indices for ENSO exist for about 500 years, and tree rings located in Florida and Texas, which may have information pertinent to the reconstruction of South Florida rainfall, are also available for durations ranging between 200 and 400 years. However, these reconstructed series have only annual resolution and their interannual variability is usually substantially lower than that of the observed climate index. Further, even if realistic simulations of the ENSO, AMO and PDO indices were generated, one would still need a model to generate daily or monthly rainfall scenarios at multiple locations using the annual or seasonal ENSO, AMO and PDO indices as predictive variables. The proposed study extends a previous study (Kwon et al. 2006a) that demonstrated the feasibility of using long Paleoclimate records to generate annual or seasonal simulations of South Florida rainfall using long proxy data. The study addresses the need to identify and integrate multiple suitable proxy indicators to predict the inter-annual and longer climate state that is influencing the daily rainfall in the season of interest.

Recent concerns on the use of a particular historical dataset dominated by one type of climate cycle (e.g., AMO cold) in modeling have generated a renewed interest in future possible rainfall realizations characterized by shifting climate cycles. This work would help toward possible development of a pragmatic modeling approach for evaluating the performance of future infrastructure improvements under conditions of climate shifts. Therefore, the objective of this study is to extend a previously developed stochastic modeling approach to generate annual or seasonal scenarios for rainfall using multiple proxies of Paleoclimate, and to use these to generate daily rainfall scenarios for South Florida rainfall.

As part of the study toward this objective, an investigation of the relationship between the climate indices and the target rainfall series is also done. A number of

statistical models are developed and applied. The study area and the data used are briefly discussed first. The models and their applications are discussed next in the order of the two key research questions indicated above. Finally, conclusions and recommendations are presented.

2 Study area and dataset description

The primary data considered are the South Florida daily rainfall, the AMO index, the Nino Sea Surface Temperature (as an ENSO surrogate), the North Atlantic Oscillation (NAO), and the Paleoclimatic Series, such as the tree-ring chronologies, and the reconstructed Palmer Drought Severity Index (PDSI). Figure 1 shows the area map with the network of the selected rain gauges and the location of the Paleoclimatology data.

2.1 Rainfall data

The daily rainfall measured over a period of 105 years (1901–2006) at 41 stations are initially considered. However, since the time series have different lengths as well as missing values, only 16 out of these 41 stations, for which long-term reliable data are readily available, are selected. The main rainy season extends from mid-May to October. We select the 92-day period beginning May 1 (MJJ), corresponding to the first mode of the peak rainy season over South Florida, for the period 1924–1998, yielding 75 complete 92-day years (6,900 days).

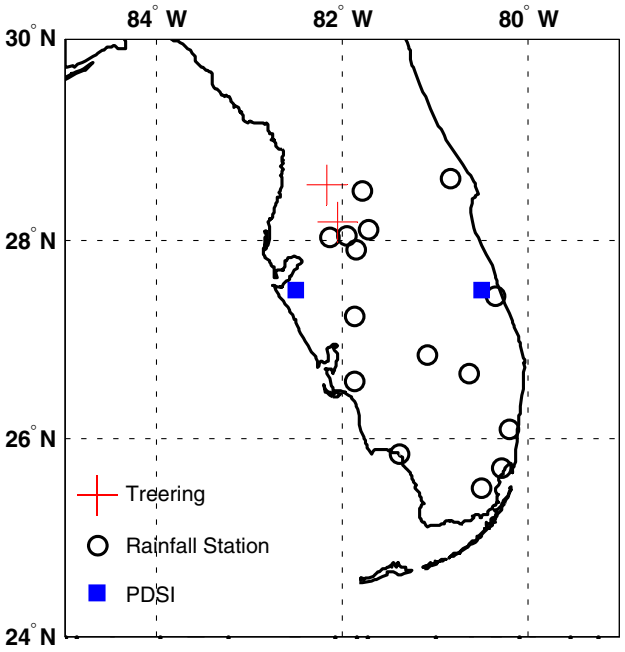


Fig. 1 Area map with the network of the selected rain gauges and the location of Paleoclimatology data. Monthly precipitation amount as aggregated from the 16 stations; 75% of the rain occurs in MJJ (May–July) and ASO (August–October)

2.2 Sea surface temperature

The SST in various regions of the equatorial Eastern Pacific Ocean regions (indicated in brackets) are averaged for each ENSO index: NINO12 (80°W–90°W, 10°S–Equator), NINO3 (120°W–150°W, 5°S–5°N), NINO34 (120°W–170°W, 5°S–5°N) and NINO4 (150°W–180°W, 5°S–5°N). The data set was acquired from Koninklijk Nederlands Meteorologisch Instituut KNMI (http://www.knmi.nl/). The index is presented as a monthly anomaly series or departure from the mean value. For AMO, the time series are obtained from the Kaplan SST dataset (Kaplan et al. 1998), which is updated monthly. It is basically an index of the North Atlantic temperatures. The Kaplan SST data set is stored on a 5° × 5° grid and consists of monthly anomalies from 1856 to present. Anomalies are based on the 1951–1980 time period. The Kaplan data set is available at: http://ingrid.ldeo.columbia.edu/SOURCES/.KAPLAN/.EXTENDED/.ssta/.

2.3 Paleoclimatic data

A tree-ring based AMO index, henceforth the Paleo-AMO index, is also used. It is available from Gray et al. (2004) for the period 1567–1990, that is 424 years. Cook et al. (2004) used tree-ring data to extend records of past drought. The tree-ring reconstructions show a more complete range of drought variability than is provided by the twentieth century instrumental record. The extended records are useful for placing the instrumental period droughts into a longer time frame, and for evaluating the rarity of the major twentieth century droughts, the 1930s and 1950s, in a broader temporal context. The temporal coverage has been expanded to the maximum permitted by the available tree-ring data, extending back nearly 2,000 years for some locations.

After the preliminary analyses, it also became clear that the connection between the AMO and the Florida rainfall series, which we used, was somewhat weak. Consequently, we acquired two tree-ring chronologies from nearby locations (illustrated in Fig. 1). These data were provided by Dr. Edward Cook, from the Tree Ring Laboratory at Columbia University. Tree I has a record for the period 1622–2005 and Tree II for the period 1646–2003.

3 Methodology

3.1 Non-homogeneous hidden Markov chain model

In this section, we summarize the structure of hidden Markov models (HMMs) and the extension of Homogeneous HMM

to Non-homogeneous HMM for multivariate time series. We briefly summarize the NHMM as presented by Robertson et al. (2003, 2004), which in turn based on Hughes and Guttorp (1994) and Hughes et al. (1999). Figure 2 illustrates the general NHMM structure, which is described below.

A hidden Markov model considers a Markovian process to generate simulations of a given time series based on random sampling of the probability distribution functions conditioned on different hidden states.

Let $\mathbf{R}_t = (R_t^1, \ldots, R_t^M)$ be a M-dimensional vector of daily rainfall measurements corresponding to M stations at time t. Let $\mathbf{R}_{1:T} = \mathbf{R}_1, \ldots, \mathbf{R}_T$ denote a vector sequence of length T. The sequence of observed daily rainfall measurements $\mathbf{R}_{1:T}$ is assumed to be characterized by a Markov chain with latent weather states $\mathbf{S}_t = (S_1, \ldots, S_T)$ taking on values from 1 to optimum hidden state K. A HMM defines a joint distribution on $\mathbf{R}_{1:T}$ and $\mathbf{S}_{1:T}$ using two main conditional independence assumptions, as described below.

The sequence of latent states $\mathbf{S}_{1:T}$ depends only on the values of the previous k hidden states. If we consider a first-order model ($k = 1$), then, for a homogenous HMM,

$$p(S_1, \ldots, S_T) = p(S_1) \prod_{t=2}^{T} p(S_t|S_{t-1}) \tag{1}$$

where $p(S_1)$ is the first-state probability vector, and $\mathbf{\Gamma} = \{\gamma_{ij}\}$, $1 \le i, j \le K$ is the state-transition matrix. Individual vector observations $\mathbf{R}_t$ are also conditionally independent of all other variables in the model, given $\mathbf{S}_t$:

$$p(\mathbf{R}_{1:T}|\mathbf{S}_{1:T}) = \prod_{t=1}^{T} p(\mathbf{R}_t|S_t) \tag{2}$$

The joint probability of the data and the hidden states can then be written as follows:

$$p(\mathbf{R}_{1:T}|\mathbf{S}_{1:T}) = \left[p(S_1) \prod_{t=2}^{T} p(S_t|S_{t-1})\right] \left[\prod_{t=1}^{T} p(\mathbf{R}_t|S_t)\right] \tag{3}$$

Rainfall amounts, $\mathbf{R}_t^M$, at time t for M weather stations are considered as conditionally independent of each other given

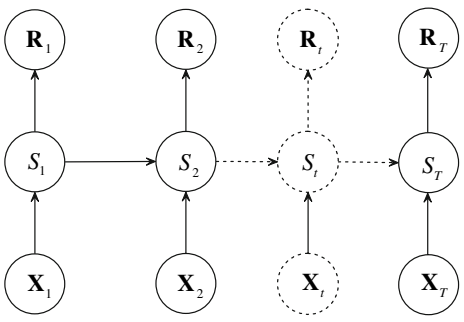


Fig. 2 Non-homogeneous hidden markov model structure. Here, ***R***, ***S*** and ***X*** indicate rainfall, weather state and exogenous variable, respectively

the hidden state S_t. Spatial dependence is hence modeled by the hidden state variable. This may then be presented as:

$$p(\mathbf{R}_t|S_t) = \prod_{m=1}^{M} p\left(\mathbf{R}_t^m|S_t\right) \tag{4}$$

Note that a spatial correlation structure of $\mathbf{R}_t$ across the M sites could also be considered, but is not in the applications reported in this study. The probability distribution at an individual rainfall station $\mathbf{R}_t^m$ is approximated by a delta function to model the zero rainfall, and delta mixture function of C exponential components for the non-zero precipitation as follows:

$$p\left(R_t^m = r|S_t = i\right) = \begin{cases} p_{im0} & r = 0 \\ \sum_{c=1}^{C} p_{imc}\lambda_{imc}e^{-\lambda_{imc}r} & r > 0 \end{cases} \tag{5}$$

with $p_{imc} \ge 0$ and $\sum_{c=1}^{C} p_{imc} = 1$ for all $m = 1, \ldots, M$ and $i = 1, \ldots, K$. The parameters of the model are estimated from the observed rainfall amount data in a standard manner using the expectation-maximization (EM) algorithm. Details of the EM estimation algorithm are available in Robertson et al. (2003) for a model that is similar except that binary precipitation occurrence is modeled instead of rainfall amounts.

The premise of the NHMM is that rainfall is assumed to be generated dependent on a set of observed time series of exogenous variables such as SST, Outgoing Longwave Radiation (OLR), and sea level pressure (SLP). Non-homogeneous HMMs can be formulated by allowing the probability distribution of the output variables to be dependent on observed input variables (Hughes and Guttorp 1994; Hughes et al. 1999). In other words, the rainfall amounts model is extended by incorporating a mixture model for amount into the HMM. A delta function and a mixture of exponentials are employed to model dry day and describe rainfall amounts on wet days, respectively. The EM algorithm is used for estimating the parameters (Dempster et al. 1977).

The assumption of conditional independence is easily visualized as edges in a directed graph of the NHMM, as shown in Fig. 2, with $\mathbf{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_T)$ is a sequence of exogenous input vectors such as global circulation models (GCMs), SST and OLR, one for each data vector. In a non-homogeneous HMM, the state-transition matrix $\mathbf{\Gamma}$ is not stationary. Its evolution over time is a function of multivariate predictor $\mathbf{X}_{1:T}$. The NHMM can be written as:

$$\begin{aligned} p(\mathbf{R}_{1:T}|\mathbf{S}_{1:T}, \mathbf{X}_{1:T}) = &\left[p(S_1|\mathbf{X}_1) \prod_{t=2}^{T} p(S_t|S_{t-1}, \mathbf{X}_t)\right] \\ &\times \left[\prod_{t=1}^{T} p(\mathbf{R}_t|S_t)\right] \end{aligned} \tag{6}$$

We assume that the observed sequences from different years are conditionally independent, given the model.

Under the NHMM model, the conditional log-likelihood $LL(\Theta)$ of the observed precipitation data, given the inputs, is defined as:

$$\begin{aligned} LL(\Theta) &= \ln P(\mathbf{R}_{1:T}|\mathbf{X}_{1:T},\Theta) \\ &= \sum \ln \sum_{S_{1:T}} \left[p(S_1|\mathbf{X}_1,\Theta) \prod_{t=2}^{T} p(S_t|S_{t-1},\mathbf{X}_t,\Theta) \right] \\ &\quad \times \left[\prod_{t=1}^{T} p(\mathbf{R}_t|S_t,\Theta) \right] \end{aligned} \tag{7}$$

The EM algorithm seeks the value of the parameter vector Θ that maximizes this expression. This maximizing value cannot be obtained analytically; however, the EM algorithm provides an iterative method of climbing the $LL(\Theta)$ surface in parameter space Θ.

The hidden state transitions in Eq. 7 are modeled by a multinomial logistic regression:

$$p(S_t = b|S_{t-1} = a, \mathbf{X}_t = \mathbf{x}) = \frac{\exp(\sigma_{ab} + \rho_b'\mathbf{x})}{\sum_{k=1}^{K} \exp(\sigma_{ak} + \rho_k'\mathbf{x})} \tag{8}$$

All σ's are real-valued parameters, while the ρ's are D-dimensional real-valued parameter vectors, where the prime denotes the vector transpose.

3.2 Wavelet-based autoregressive model

Stochastic hydrologic methods have been very useful for a variety of water resources problems where temporal uncertainty needs to be quantified (Kwon et al. 2007). Thus, the literature (Box and Jenkins 1970; Thomas and Fiering 1962; Yevjevich 1972) has developed around autoregressive moving average models and their extensions to consider seasonality through periodic terms. Multi-site models and space–time disaggregation approaches have also been considered (Koutsoyiannis 1994; Stedinger and Vogel 1984; Valencia and Schaake 1973). The wavelet transform has recently been applied to many geophysical time series (Farge 1992; Foufoula-Georgiou and Kumar 1995; Hubbard 1996; Kulkarni 2000; Kwon et al. 2007; Kwon et al. 2006b; Wang and Wang 1996; Weng and Lau 1994). The main advantage cited for wavelet transforms over other spectral methods is that they permit an orthogonal decomposition of the original signal both in the time and frequency domains. This is an attractive property in the context of the additive decomposition suggested in this study.

A wavelet-based autoregressive model (WARM) has recently been initiated by Kwon et al. (2007). One objective of the WARM model is to explore the use of autoregressive models with wavelet decomposition as a time series simulator for systems with quasi-periodic long memory behavior or nonlinear dynamics that may lead to persistent regime like behavior or stochastic regime transitions, without a priori specifying any of these modeling structures. The continuous wavelet transform is applied to decompose a univariate time series into several statistically significant components and then a linear AR model is employed to simulate each component extracted from wavelet transform analysis, as well as the residual "noise" term (Kwon et al. 2007).

Consider a time series x_t, $t = 1, \ldots, N$, recorded at monthly that exhibits quasi-oscillatory, low-frequency variations at intra-seasonal, interannual and longer time scales, as seen in many hydroclimatic time series. Consider the decomposition of this series into K component series R_{kt} that represent "signal" and a residual term ε_t.

$$x_t = \sum_{k=1}^{K} R_{kt} + \varepsilon_t \tag{9}$$

The decomposition in Eq. 9 considers that there are K orthogonal or independent series that carry the low-frequency information, and the residual, ε_t, is a stochastic process. Here, we consider a linear autoregressive model for each term, leading to the following model structure:

$$\begin{aligned} x_t &= \sum_{k=1}^{K} AR(R_{kt}; p_k) + AR(\varepsilon_t; p) \\ &= \sum_{k=1}^{K} \left(\sum_{i=1}^{p_k} \alpha_{k,i} R_{t-i,k} + v_{t,k} \right) + \sum_{j=1}^{p} \beta_j \varepsilon_{t-j} + \omega_t \end{aligned} \tag{10}$$

Here, p_k is the order of an autoregressive model fit to the kth signal, $\alpha_{k,i}$ are the corresponding autoregression coefficients, p is the order of an autoregressive model fit to the stochastic process ε_t, with β_j as the associated autoregression coefficients, and $v_{t,k}$ and ω_t are independent, identically distributed, noise processes. The order of each autoregressive model and the coefficients of each autoregression in Eq. 10 is estimated using standard time series methods (e.g., Maximum Likelihood and Akaike Information Criteria (AIC). For more details regarding model procedures, see Kwon et al. (2007). Here, the presentation of the wavelet transform, as described by Kwon et al. (2007) and Torrence and Compo (1998), is briefly summarized.

The continuous wavelet transform of a square integrable complex-valued signal $x(t)$ is defined by Chui (1992):

$$X(b,a) = |a|^{-1/2} \int_{-\infty}^{+\infty} x(t)\bar{\varphi}((t-b)/a)dt \tag{11}$$

where $X(b,a)$ is the wavelet spectrum, $\varphi(t)$ is a complex-valued wavelet function, b is a translation (shift) parameter and $a \neq 0$ is a scale parameter. Here, the Morlet wavelet has been used, defined as $\varphi(t) = \pi^{-1/4} e^{i\omega_o t} e^{-t^2/2}$, where ω_0 is a

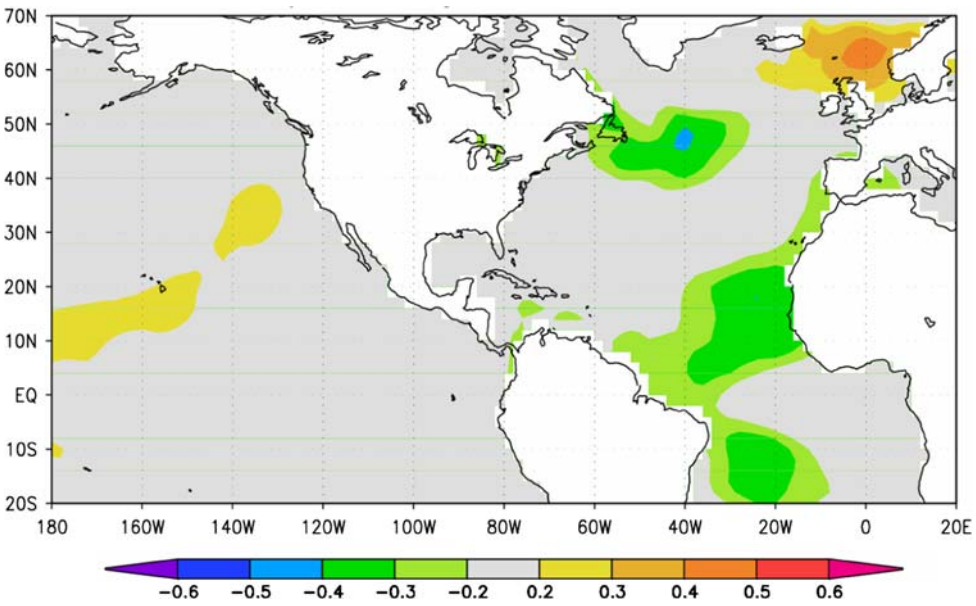


Fig. 3 The spatial pattern of correlations of SST with the MJJ rainfall series from 1924 to 1998

Table 1 Correlation coefficients between the selected daily statistics and climate indices

	RF amount	Wet days	Wet to wet prob.	Dry to dry prob.
RF amount	1.00	0.75	0.67	−0.59
Wet days	0.75	1.00	0.66	−0.81
Wet to wet prob.	0.67	0.66	1.00	−0.26
Dry to dry prob.	−0.59	−0.81	−0.26	1.00
Tree-1	**0.25**	0.06	0.07	−0.03
Tree-2	0.11	0.04	−0.05	−0.05
AMO	−0.06	**−0.26**	−0.16	**0.28**
NAO	0.04	0.11	0.00	−0.09
PDSI-1	**0.28**	**0.19**	0.12	**−0.27**
PDSI-2	**0.26**	**0.18**	0.12	**−0.26**
SST PC-1	**−0.33**	**−0.47**	**−0.36**	**0.44**
NINO12	−0.11	**−0.36**	−0.12	**0.29**

The null hypothesis of zero correlation is rejected at the 5 and 10% (bold) significance levels for each of these estimates

frequency. The wavelet power spectrum can be estimated using a discrete Fourier transform (DFT). The DFT of x_n is given by:

$$\hat{x}_j = \frac{1}{N}\sum_{n=0}^{N-1} x_n \exp^{-2\pi ijn/N} \tag{12}$$

where $j = 0, \ldots, N-1$ is an index that corresponds to the Fourier frequencies, and $n = 0, \ldots, N-1$ is a time index. In the continuous limit, the Fourier transform of a function $\varphi(t/a)$ is given by $\varphi(a\omega)$. By the convolution theorem, the wavelet transform is the inverse Fourier transform of the product:

$$X_n(a) = \sum_{j=0}^{N-1} \hat{x}_j \hat{\varphi} \times (a\omega_j) \exp^{i\omega_j n\delta t} \tag{13}$$

To interpret the wavelet spectrum by time and scale, the Global Wavelet Spectrum (GWP) is used. The GWP is defined as the time integrated variance of energy coefficients at every scale:

$$\text{GWP} : \bar{X}_n^2(a) = \frac{1}{N}\sum_{n=0}^{N-1} |X_n(a)|^2 \tag{14}$$

The GWP for x_t is computed as function of the scale a. A "red noise" significance test for the GWP(x) is then developed following the procedure described by Torrence and Compo (1998). It entails first fitting an AR(1) model to x_t, and then computing its Fourier spectrum, and the associated one-sided 95% confidence limits as a function of frequency. The scales a_j for which the GWP(x) spectrum is higher than the red noise significance level are then selected as candidates for reconstruction.

4 Modeling low-frequency climate variability in South Florida rainfall

A number of tree-ring, coral and other proxy climate data series are now available. There are also a number of global and regional climate indices, such as ENSO, AMO, PDO, NAO (North Atlantic Ocean) that have been reconstructed from such proxy series. The objective of this section is to identify a suitable combination of these indices that is best associated with averaged rainfall in South Florida for May–June–July (MJJ) season. The suitable combination may be

Table 2 Posterior estimates of the regression parameter values for Eq. 17 which estimates the mean of the Normal distribution conditional on each of the predictors listed below

Parameter	Mean	SD	5%	50%	95%
Intercept	20.24	0.39	19.44	20.24	20.97
Tree(5)	4.96	2.53	−0.02	5.01	10.15
NAO(8)	3.86	2.56	−1.18	3.94	8.80
NAO(40)	−9.77	4.36	−17.79	−9.91	−1.05
PDSI(5)	3.45	2.44	−1.35	3.41	8.28
SST PC1(3)	−8.10	2.97	−13.86	−8.11	−2.36
NINO12(6)	−4.37	2.25	−8.78	−4.34	0.16
AMO(60)	1.32	2.45	−3.56	1.33	6.13

The mean, SD and confidence intervals for each parameter are listed

identified as a single index derived as a weighted linear combination of the candidate proxy and index time series, or may be a subset of those series that are most relevant for South Florida rainfall in the chosen season. Relevance is identified herein by looking at a suite of statistics of daily rainfall for the season of interest. For example, these statistics may include the average rainfall, the variance of the daily rainfall over the season, the probability of a wet day following a wet day or a dry day following a dry day, or the length of the longest wet or dry spell.

There exists a relationship between the hydroclimatic patterns of South Florida and ENSO indices and other SST patterns that persists despite differences in flood mechanism, drainage area and season of occurrence. First, global SSTs (for MJJ) from 1924 to 1998 are evaluated as predictors. The spatial pattern of correlations with the MJJ rainfall series is displayed in Fig. 3. This figure shows that anomalous conditions in SSTs are likely to strongly influence the rainfall. While it exhibits patterns of correlations reflecting North Atlantic SSTs as expected, it is notable that the strongest correlations are on the two regions with

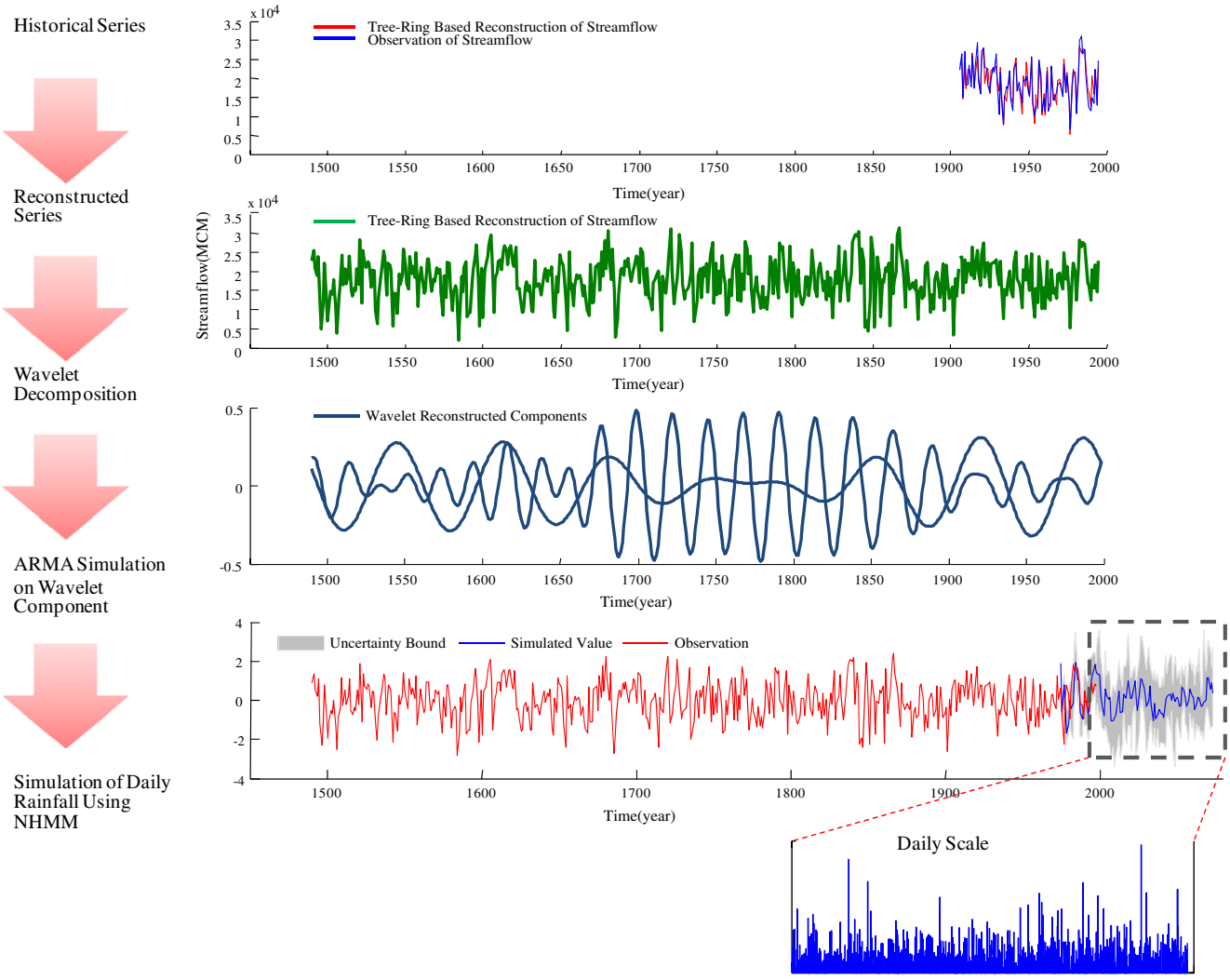


Fig. 4 Schematic representation of generation of daily rainfall scenarios conditional on a scenario of seasonal/annual climate indices

different signs. In the present study, Principal Component Analysis (PCA) is applied to the common signal from SST zones with the highest correlations (60°W–20°E, 35°N–70°N). The correlation across the time series is provided in Table 1. As may be seen, SST PC-1, PDSI and NINO12 index show some correspondence within both wet condition and dry condition. It seems that the SST PC-1 is in good agreement with most of statistics of rainfall in South Florida. We note that the correlations of the climate series with MJJ rainfall are not strong, even where they are statistically significant at the 5 and 10% levels.

As noted above, the raw correlations between rainfall and climate proxies are more or less weak; so, it is not easy to derive a single index using those raw series. Hence, we intend to apply wavelet decomposition approach to the time series so as to characterize the impact of the process underlying the low-frequency variability on rainfall patterns. The qualitative inspection of the wavelet spectrum is insightful and it enables the detection of possible correspondence between time series and significant oscillatory components. Given the spectral structures of rainfall and climate proxies, it seems that rainfall in South Florida is modulated by both inter-annual and multi-decadal variability. In other words, the inter-annual and decadal variability of rainfall is showing a similar spectrum behavior with ENSO while the multi-decadal variability does well match with AMO type oscillation.

With regard to this premise, a Bayesian hierarchical model using the specified wavelet components which the oscillations from Tree, AMO, NAO, PDSI, SST and NINO12 are built for rainfall. Each of the statistically significant wavelet components is first extracted and used as a predictor in a Bayesian regression model for rainfall. The objective of Bayesian inference is to compute the posterior distribution of the desired variables, in this case the parameters of the rainfall distribution. The posterior distribution $p(\theta|x)$ is given by Bayes Theorem as follow:

$$p(\theta|x) = \frac{p(\theta) \times p(x|\theta)}{p(x)} = \frac{p(\theta) \times p(x|\theta)}{\int_{\Theta} p(\theta) \times p(x|\theta) d\theta} \propto p(\theta) \times p(x|\theta) \quad (15)$$

where θ is the vector of parameters of the distribution to be fitted, Θ is the space parameter, $p(x|\theta)$ is the likelihood function, x is the vector of observations and $p(\theta)$ is the prior

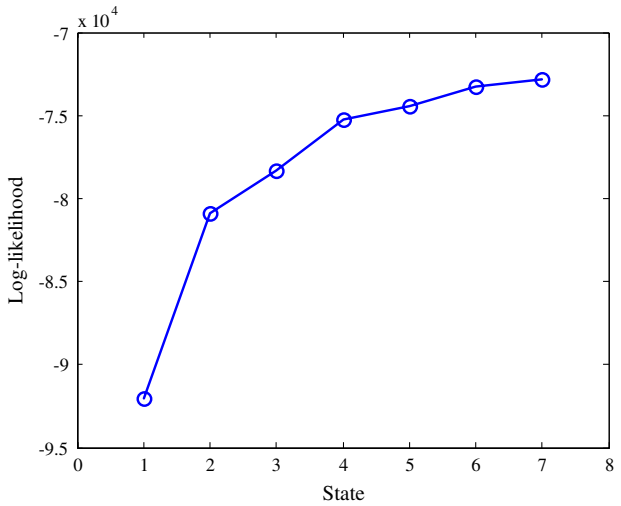


Fig. 5 Log-likelihood estimates for amounts models using mixture of two exponential distributions for rainfall on wet days

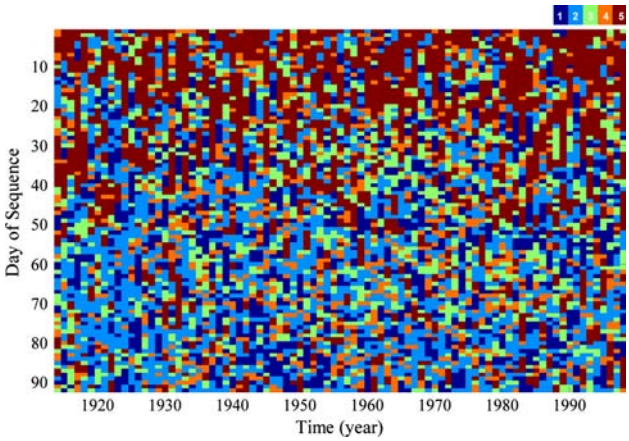
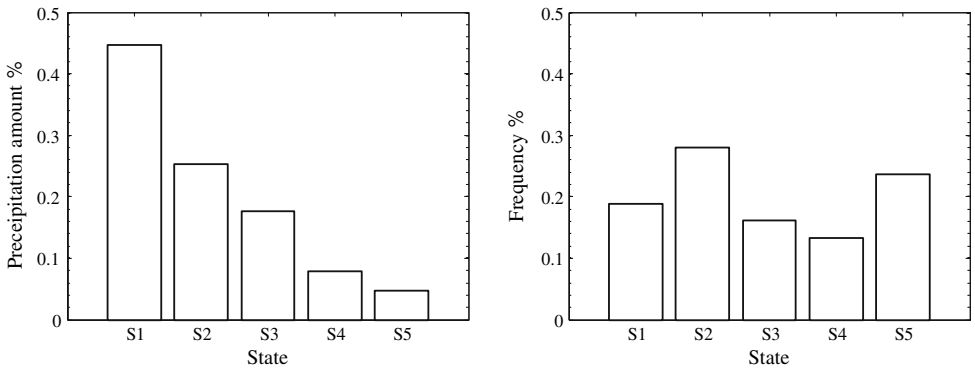


Fig. 6 The estimated state sequence. The number of days falling into the five states for MJJ season are 1,380, 2,048, 1,174, 973 and 2,245

Fig. 7 The estimated percentage of amount and frequency of seasonal rainfall according to each state



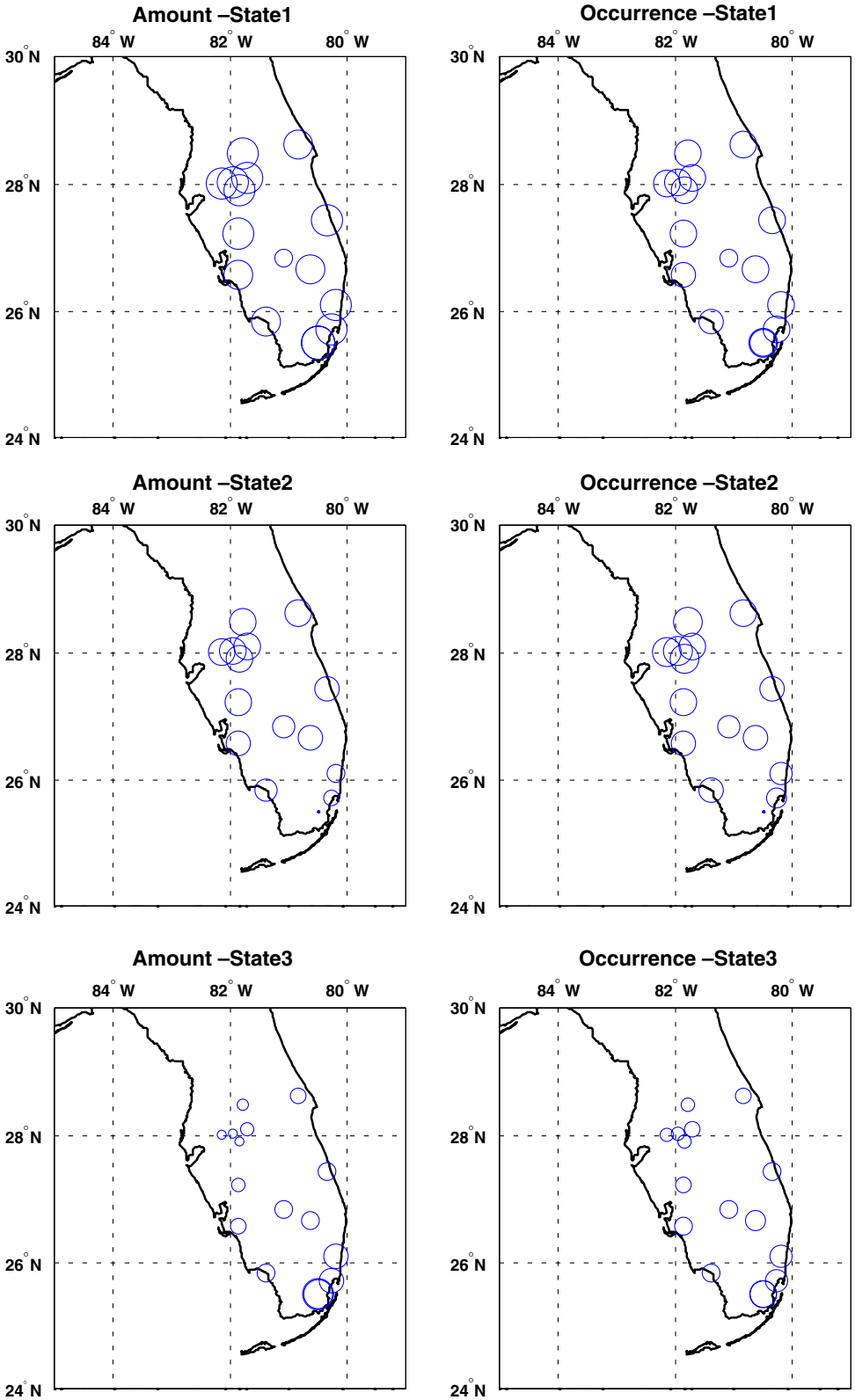
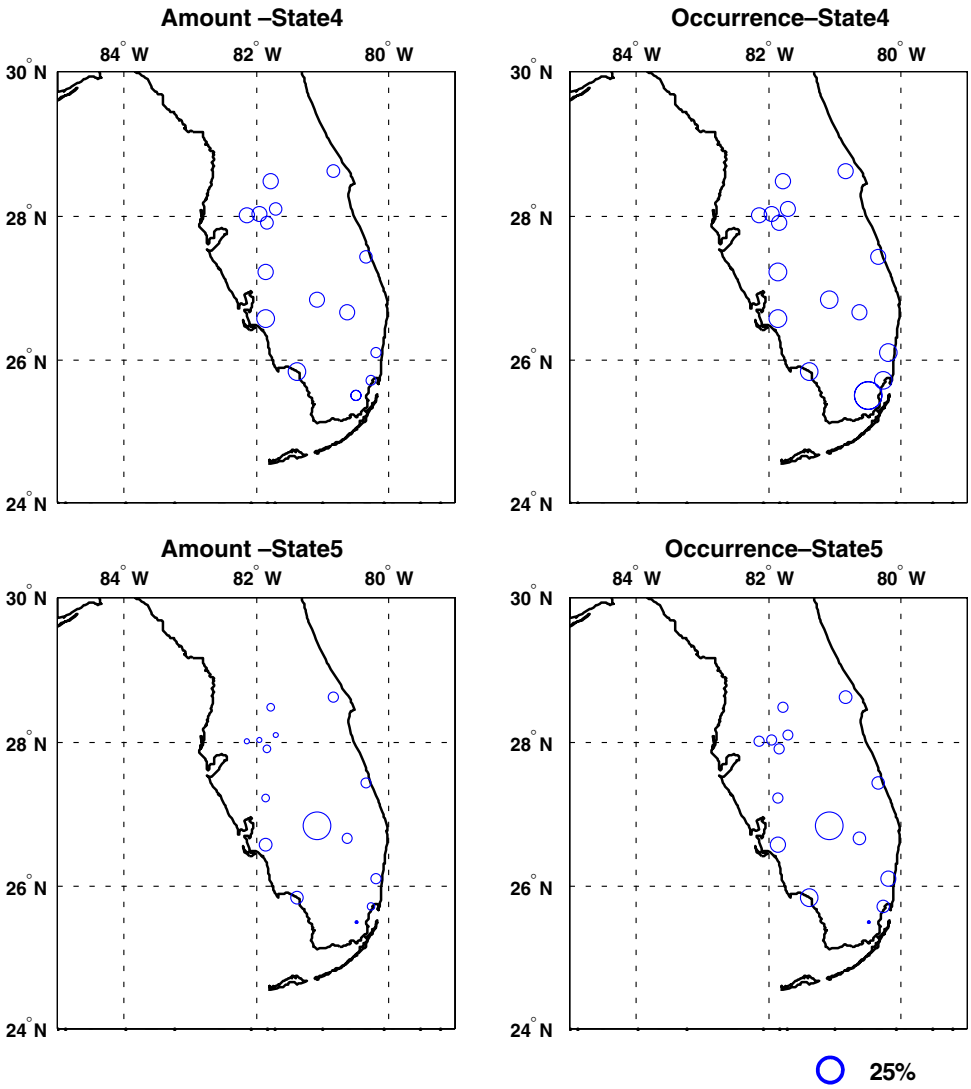


Fig. 8 MJJ NHMM rainfall amount and occurrence probabilities (%) (*Circle size*). The number of days in each state are 1,380, 2,048, 1,174, 973 and 2,245

distribution. Here, we present a method for incorporating climate information into updated estimates of the parameters for the Normal distribution used to represent the seasonal rainfall. The model is expressed in terms of a location (mean) parameter $\mu(t)$, which change with time, t. The parameters are hypothesized to be functions of climate low frequencies that are generally recognized climate phenomena with expected local impacts. Using the

Fig. 8 continued



Normal distribution, the distribution of rainfall $Y(t)$ can be modeled as follows:

$$Y(t) \sim \text{Normal}(\mu(t), \sigma) \tag{16}$$

$$\mu(t) = \beta_0 + \beta_1 \mathbf{X}_{1i}(t) + \cdots + \beta_k \mathbf{X}_{ki}(t) \tag{17}$$

where $\mathbf{X}_i$, wavelet reconstructed components derived by climate indices, are the ith rows of the known design matrices $\mathbf{X}$, and $\boldsymbol{\beta}$ is a vector of regression parameters which is normally distributed, $\beta_k \sim N(\eta_{\beta_k}, \sigma_k)$. Similarly, the hyper-parameters η_{β_k} and σ_k are also assigned to a conjugate normal distribution and half-Cauchy distribution with hyperparameters a and b. Here, we use the Gibbs sampler, which is an effective Markov chain Monte Carlo method, for simulating the posterior probability distribution of the data field conditional on the current choice of parameters (Chen et al. 2005; Gelman et al. 2003; Godsill et al. 2001; Hue et al. 2002; Ridgeway and Madigan 2003; Tsionas 2001; Tucker and Liu 2003).

Table 2 summarizes the results from the Bayesian model using wavelet components as the covariates. Overall performance of this model shows good results, with a correlation coefficient of 0.55. All regression coefficients are statistically significant with their associated uncertainty bounds.

Given the results, the wavelet reconstructed components identified are used for further analyses to produce multiple rainfall scenarios. In other words, the wavelet reconstructed components are simulated using an autoregressive model. The simulated components are then used as predictors with the Non-homogeneous HMM.

5 Case study (May–June–July daily rainfall simulation)

The primary application we consider in this study is the simulation of the MJJ daily rainfall of the South Florida

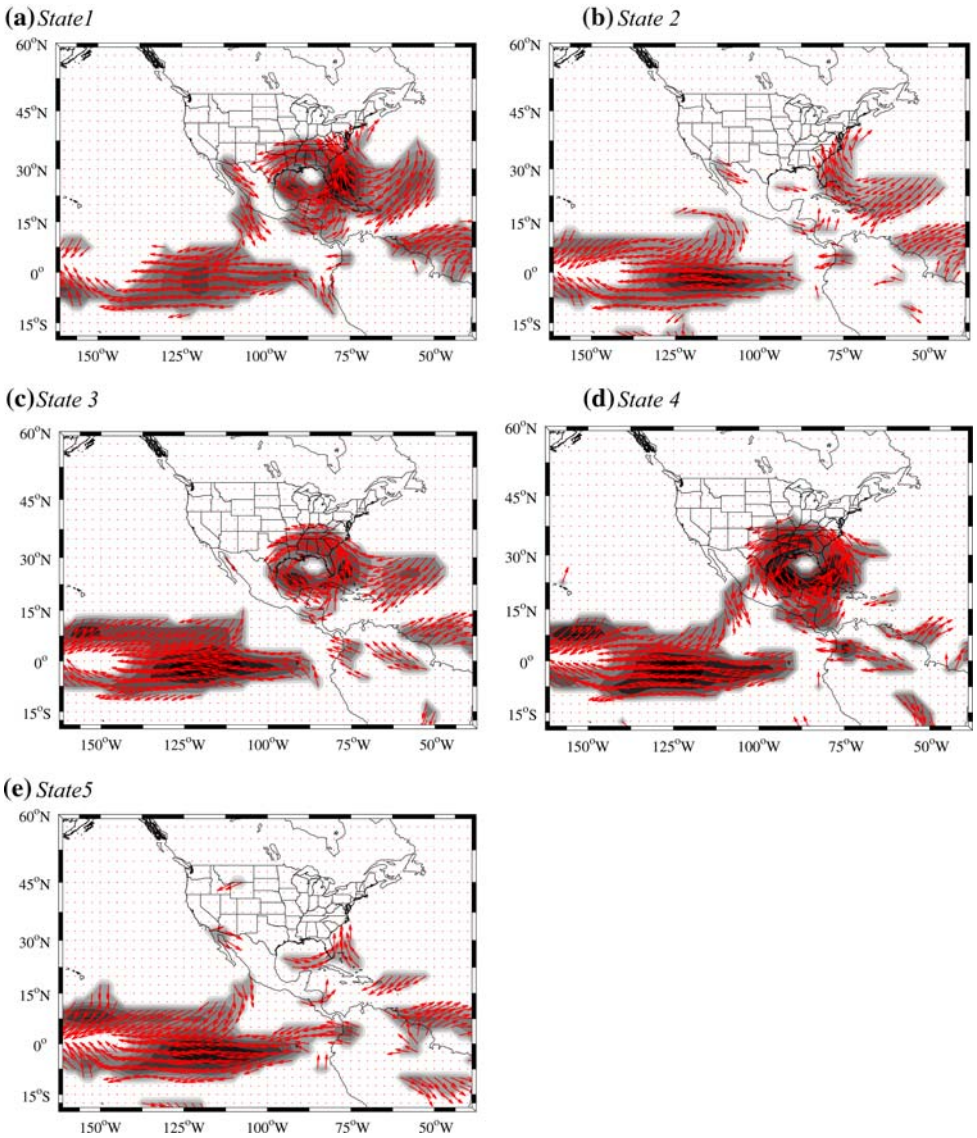


Fig. 9 NHMM state anomaly composites of storm tracks for MJJ season. The composites are defined on the days assigned to each state. The storm track is identified as a connected set of grid boxes that is in the vicinity of the rainfall location of interest and has a vertically integrated moisture flux that is above some threshold (90th percentile of the daily value for that grid box and season)

area at key points in time from 1924 to 1998, using the climate proxies just identified. However, we shall begin by simulation of data using climate division-4 Florida monthly rainfall data (1895–2004) to establish the potential efficacy of the NHMM based simulation scheme.

The main question addressed in this section is how we can generate long-term rainfall scenarios that exhibit the inter-annual and decadal variability seen in longer records, and use these to simulate the daily rainfall. A hierarchical modeling approach for multi-scale simulation using climate proxies and historical data is proposed. First, historical data is extended by using Paleoclimatic climate proxy, and then wavelet transform analysis is carried out to decompose the extended data into several statistically significant components. Third, a linear AR model is employed to simulate each component extracted from wavelet transform analysis, as well as the residual "noise" term. Finally, NHMM model is used for the generation of daily rainfall scenarios conditional on a scenario of seasonal/annual climate indices. A schematic representation of the concept of the modeling procedure is presented in Fig. 4.

5.1 Hidden state determination

Log-likelihood is used to evaluate the quality of the fitted HMMs with different state **K** (Robertson et al. 2004). In each case, the EM algorithm is run ten times from different initial seeds, selecting the run with the highest log-likelihood. The resulting normalized values of the log-likelihood for each model are plotted for **K** = 1–7 in Fig. 5.

The log-likelihood of the models increases monotonically with **K** in an asymptotic manner. Thus, the model does not "overfit" for larger **K**, suggesting that the rainfall

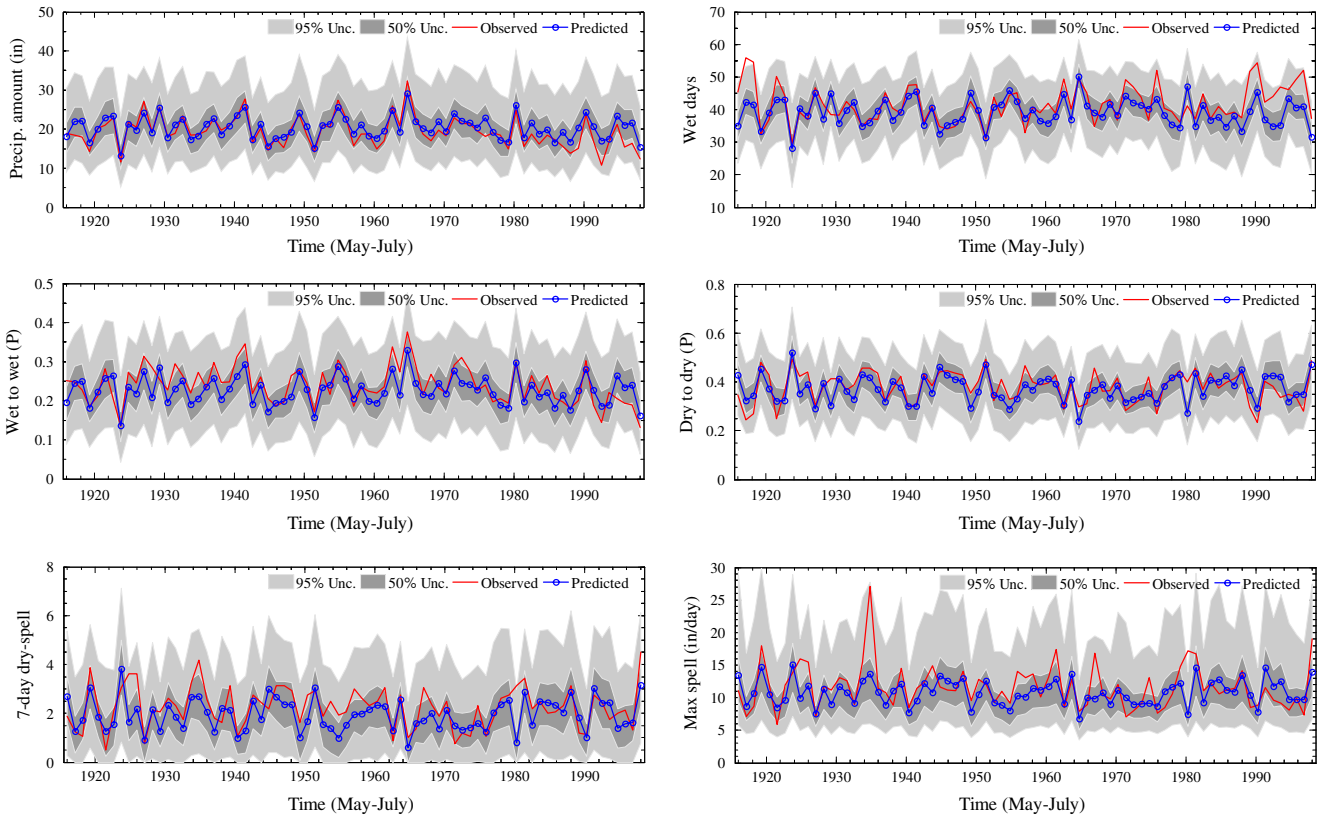


Fig. 10 The MJJ seasonal statistics of **a** precipitation amount, **b** wet days, **c** wet to wet probability, **d** dry to dry probability, **e** 7-day dry spell length, and **f** maximum dry spell length of NHMM-simulated rainfall versus the observed (*solid*) averaged over all 16 stations. Plotted is the median of 50 NHMM simulations (*circle*). All statistics per season are averaged across the 16 stations. The input "predictor" is climate division-4 Florida precipitation. The *error bars* indicate the entire range of the 50 simulations, with the inter-quartile range given by the *shaded area*

data is considerably more complex than our models. For parsimony, we choose **K** = 5. The results discussed below are not sensitive to choosing **K** = 4–7. Choosing larger **K** makes the states more difficult to interpret physically. Thus, **K** = 5 represents a reasonable state between model performance and physical interpretability.

For the observed sequence of daily precipitation, we wish to find the most likely sequence the underlying hidden states might have generated it. We can find the most probable sequence of hidden states by Viterbi algorithm. This algorithm is a dynamic programming algorithm that provides a tractable way of analyzing observations of NHMM to recapture the most likely underlying state sequence (Rabiner 1989; Viterbi 1967). The spatial, temporal and synoptic patterns associated with each state in MJJ are presented below.

The most likely state sequences for MJJ, calculated using the Viterbi algorithm, are shown in Fig. 6. The numbers of days falling into the five states for MJJ season are 1380, 2048, 1174, 973 and 2245. This figure suggests that the rainfall sequences exhibit considerable variability on intra-seasonal as well as inter-annual time scales. Figure 7 presents the frequency and amount of precipitation associated with each state for MJJ. The spatial distribution of rainfall corresponding to each state is presented in Fig. 8.

State 5 is the second most frequent, and represents the dry state. The associated storm track pattern (Fig. 9) indicates the absence of cloudiness over the Florida-Caribbean region, and an associated wind anomaly pattern that suggests atmospheric transport out of the region, consistent with the lack of rainfall. This state is most prominent in early May with incidence decreasing in June and July as the rainy season picks up. The spatial pattern of this dry state occurrence is regional.

State 1 is the wettest, contributing 45% rainfall over MJJ season, and occurs primarily in the mid-May to end-June period. State 1 has a strong storm anomaly signature and a well developed circulation consistent with strong regional convection, and the intense rainfall.

State 2 is a wet condition, contributing 25% rainfall, but also is the most frequent. The storm track pattern is related to a typical climatological pattern in MJJ season over Florida-Caribbean region. States 3 and 4 show a similar

Table 3 The MJJ season statistics of NHMM-simulated rainfall versus the observed values (solid) averaged over all 16 stations using division-4 Florida rainfall as an input

	Amounts	WW prob.	DD prob.	Wet days	7-day dry spell	Max dry spell
Station 1	0.42	0.33	0.38	0.24	0.48	0.43
Station 2	0.51	0.42	0.47	0.48	0.35	0.46
Station 3	0.66	0.65	0.68	0.61	0.57	0.44
Station 4	0.41	0.43	0.48	0.49	0.39	0.36
Station 5	0.59	0.46	0.45	0.45	0.29	0.26
Station 6	0.53	0.28	0.44	0.37	0.39	0.42
Station 7	0.45	0.45	0.50	0.43	0.49	0.34
Station 8	0.72	0.70	0.63	0.70	0.44	0.28
Station 9	0.00	0.05	0.06	0.07	−0.14	0.09
Station 10	0.43	0.37	0.29	0.21	0.29	0.23
Station 11	0.39	0.40	0.47	0.36	0.43	0.36
Station 12	0.38	0.37	0.41	0.38	0.30	0.29
Station 13	0.38	0.34	0.29	0.18	0.49	0.37
Station 14	0.61	0.54	0.27	0.19	0.30	0.11
Station 15	0.46	0.28	0.42	0.39	0.43	0.46
Station 16	0.44	0.27	0.38	0.30	0.46	0.35
Total	0.87	0.76	0.71	0.66	0.66	0.55

pattern of storm track and rainfall frequency, but State 3 is somewhat wetter per event.

5.2 Summer daily rainfall simulations with climate division-4 Florida monthly rainfall data

To test the efficacy of the NHMM model as a daily rainfall simulator, a "perfect forecast" simulation is first implemented by using the climate division-4 Florida monthly rainfall data, which is spatially averaged over the South Florida, as an input vector.

The MJJ season statistics of NHMM-simulated rainfall versus the observed ones (solid) averaged over all 16 stations using division-4 Florida rainfall as an input are illustrated in Fig. 10 and Table 3. All statistics per season are averaged across the 16 stations and, in all cases, 50 realizations of rainfall are generated for the 1924–1998 period. The average statistics over all the stations is estimated as follows: (1) from each realization and for each station, we estimate the desired seasonal statistics; (2) similarly, we compute the same statistics from the observed time series per station; (3) we compute the average time series for all the stations given the seasonally aggregated ones; (4) we compare that to the observed time series; and (5) we compare that to the observed time series that is averaged over all the stations.

The seasonal rainfall amounts simulated by NHMM using climate division-4 Florida precipitation have a correlation of 0.87 with the observed rainfall. This correlation is similar to that found between seasonal division-4 Florida and observed MJJ rainfall from 16 stations, suggesting that the seasonal signal is successfully transmitted at least on average across the 16 stations. The other statistics compared in Fig. 10 are the wet day to wet day transition probability and the dry day to dry day transition probability for each season, the number of wet days in the season, the number of 7-day dry spells in the season, and the maximum dry spell length for the season. The year to year variation of most of these statistics is well represented in the model simulations.

The WARM simulation scheme is applied to the division-4 Florida rainfall for the next 30 years. Figure 11 depicts a projection of the simulation. The global wavelet spectral density for the observed and the simulated rainfall series is shown in Fig. 12. Note that the uncertainty bounds are becoming large at very low frequencies. The spectrum from AR simulation tends to be rather flat over the frequency range considered, while the WARM simulations suggest a much sharper peak in the extreme low frequency end similar to that in the observed spectrum. It is clear that the low-frequency behavior of the simulation is preserved. Finally, daily rainfall scenarios conditioned on the simulated WARM precipitation are produced by NHMM, and the MJJ seasonal statistics of NHMM-simulated rainfall averaged over all 16 stations are illustrated in Fig. 13.

5.3 Multiple combination of climate variability

Each of the statistically significant wavelet components is first extracted and used as a predictor for rainfall in the

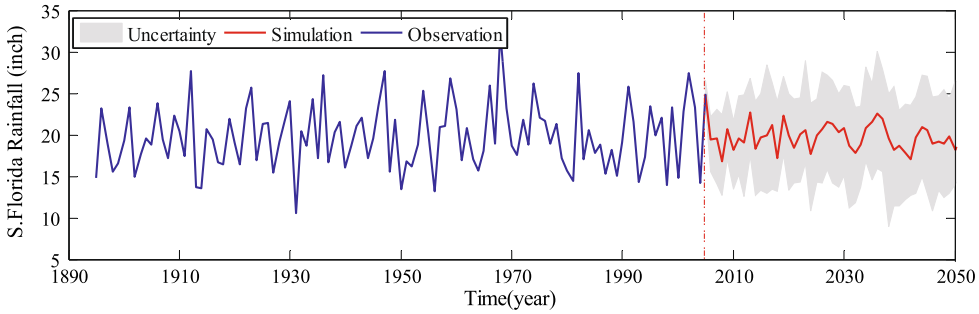


Fig. 11 Climate division-4 Florida rainfall simulation projection using WARM model

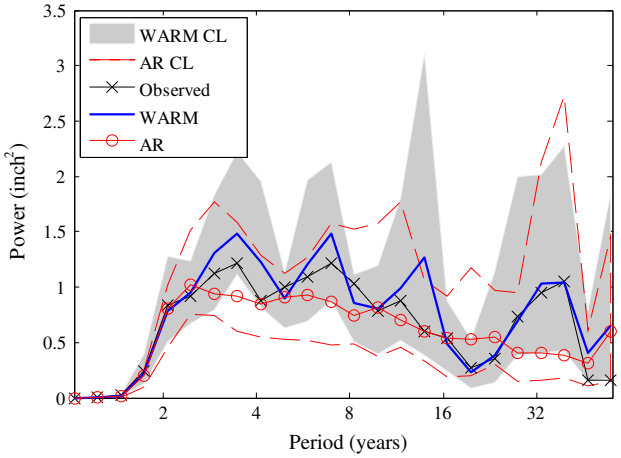


Fig. 12 Comparison of Global wavelet power spectrum for the simulated climate division-4 Florida rainfall series using the AR and the WARM models

NHMM model. The MJJ season statistics of NHMM-simulated rainfall versus the observed ones (solid) averaged over all 16 stations using wavelet components as an input are illustrated in Fig. 14 and Table 4. Most of the statistics from the simulations have a correlation of about 0.5 with the corresponding statistics for observed rainfall. Statistics of dry spell length of 7 days and maximum dry spell length are not reproduced as well as desired. The wavelet components are then simulated forward using an autoregressive model. The simulated components are then used as predictors with the NHMM model, and the results are shown for the post 2003 period in Fig. 15, as a forward simulation.

Since the algorithm development was motivated by the observation that Florida rainfall might exhibit low-frequency variations at inter-annual to decadal time scales, it is useful to examine whether the character of these variations has been captured by the models in terms of the inter-annual and decadal variations of the statistics of daily rainfall for the MJJ season. The global wavelet spectra for the MJJ rainfall for the historical period and for the future period are presented in Fig. 16a, b, respectively. The ability to reproduce the low-frequency variations is demonstrated. Note that the future simulation presented here covers only 30 years and, hence, the decadal and multi-decadal signals are not expected to be seen as significant. The NHMM simulations show a very similar structure to the observed spectrum.

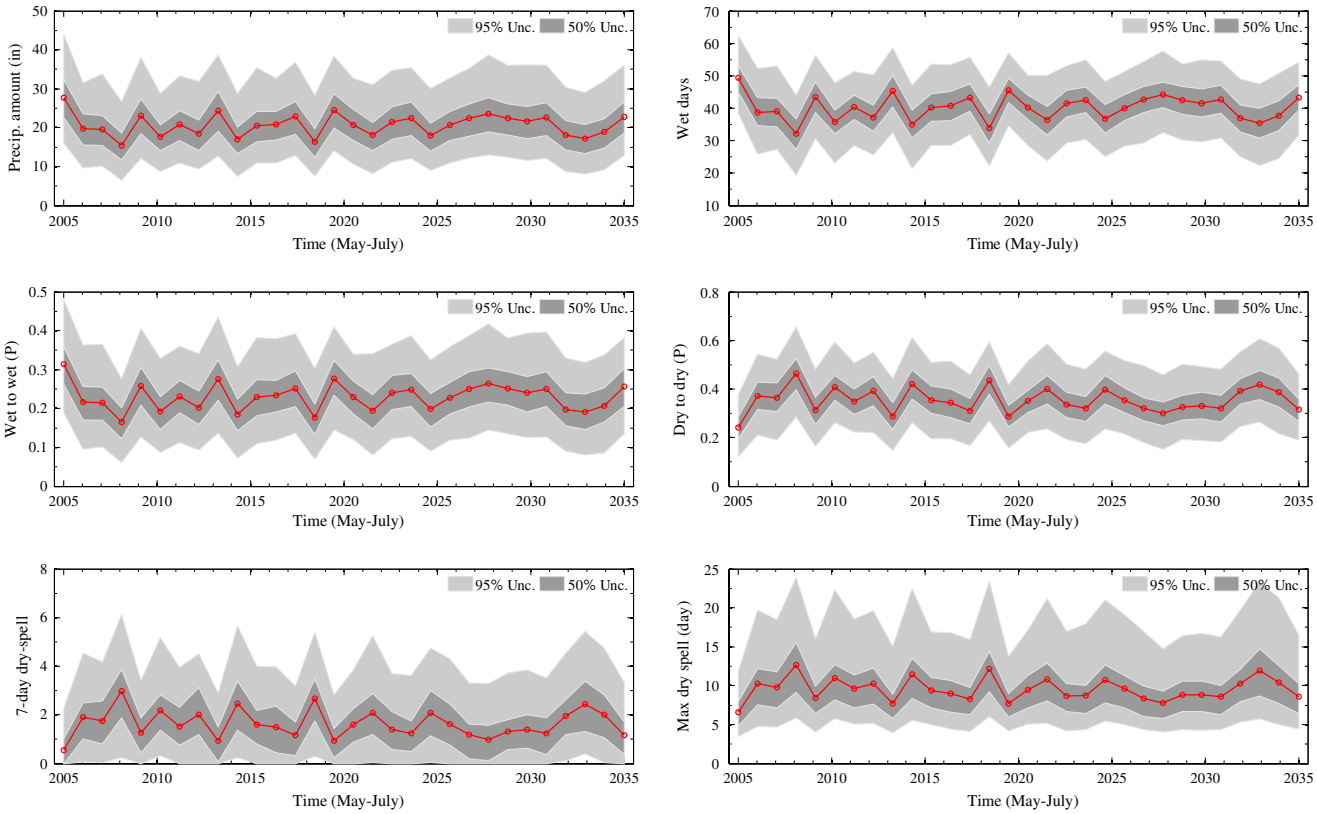


Fig. 13 The MJJ seasonal statistics of NHMM-simulated rainfall averaged over all 16 stations. Plotted is the median of 50 NHMM simulations (*circle*). The input "predictor" is climate division-4 Florida precipitation. The *error bars* indicate the entire range of the 50 simulations, with the inter-quartile range given by the *shaded area*

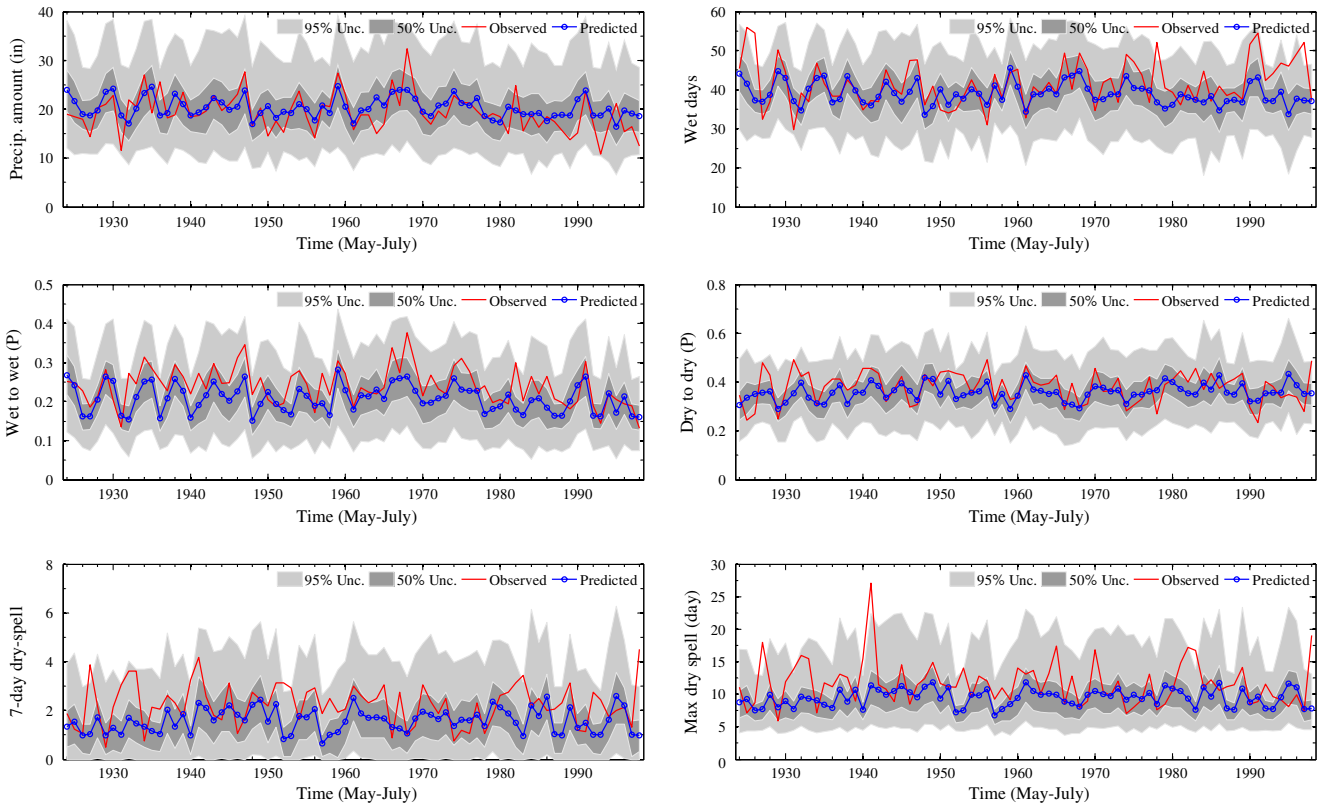


Fig. 14 The MJJ seasonal statistics of **a** precipitation amount, **b** wet days, **c** wet to wet probability, **d** dry to dry probability, **e** 7-day dry spell length, and **f** maximum dry spell length of NHMM-simulated rainfall versus the observed (*solid*) averaged over all 16 stations. Plotted is the median of 50 NHMM simulations (*circle*). All statistics per season are averaged across the 16 stations. The input "predictor" is the wavelet components identified by multiple hydroclimatic proxies. The *error bars* indicate the entire range of the 50 simulations, with the inter-quartile range given by the *shaded area*

Table 4 The MJJ season statistics of NHMM-simulated rainfall versus the observed values (solid) averaged over all 16 stations using wavelet components as an input

	Amounts	WW prob.	DD prob.	Wet days	7-day dry spell	Max dry spell
Station 1	0.20	0.31	0.23	0.29	0.00	0.13
Station 2	0.45	0.47	0.46	0.58	0.18	0.10
Station 3	0.16	0.32	0.29	0.43	−0.15	0.03
Station 4	0.41	0.46	0.46	0.55	0.12	0.08
Station 5	0.40	0.24	0.16	0.23	−0.04	−0.10
Station 6	0.21	0.14	0.24	0.18	0.14	0.06
Station 7	0.26	0.50	0.22	0.26	0.09	0.02
Station8	0.40	0.41	0.29	0.39	0.15	0.11
Station 9	0.02	0.15	−0.09	−0.03	−0.03	0.05
Station 10	0.12	0.22	0.21	0.11	0.06	0.11
Station 11	0.21	0.20	0.17	0.13	−0.03	0.02
Station 12	0.42	0.48	0.47	0.41	0.18	0.08
Station 13	0.42	0.40	0.27	0.12	0.06	0.06
Station 14	0.28	0.45	0.16	0.10	−0.05	0.01
Station 15	0.45	0.30	0.56	0.54	0.20	0.17
Station 16	0.39	0.36	0.11	0.08	0.12	−0.03
Total	0.53	0.62	0.47	0.49	0.13	0.11

6 Conclusion and scope for further research

A key objective of this study was to develop and demonstrate the methods for daily rainfall simulation in South Florida that were conditioned on suitable climate proxies for which long data sets may be available. A hierarchical stochastic modeling system for this purpose was introduced and illustrated with data from 16 long-term rain gauges in the SFWMD region of interest. The demonstration considered the summer wet season. The major features and limitations of the methods presented are itemized below, followed by a brief discussion of the directions for further research.

1. The low-frequency variations of the regional climate state relevant for daily rainfall for the current season are first simulated using WARM. Multiple climate proxies that capture common or unique aspects of low-frequency variation can be simulated here. These may include a long rainfall record in the region, reconstructed SST from appropriate regions, tree ring information, regional PDSI reconstructed from tree rings or other proxies, and/or simulations of ocean,

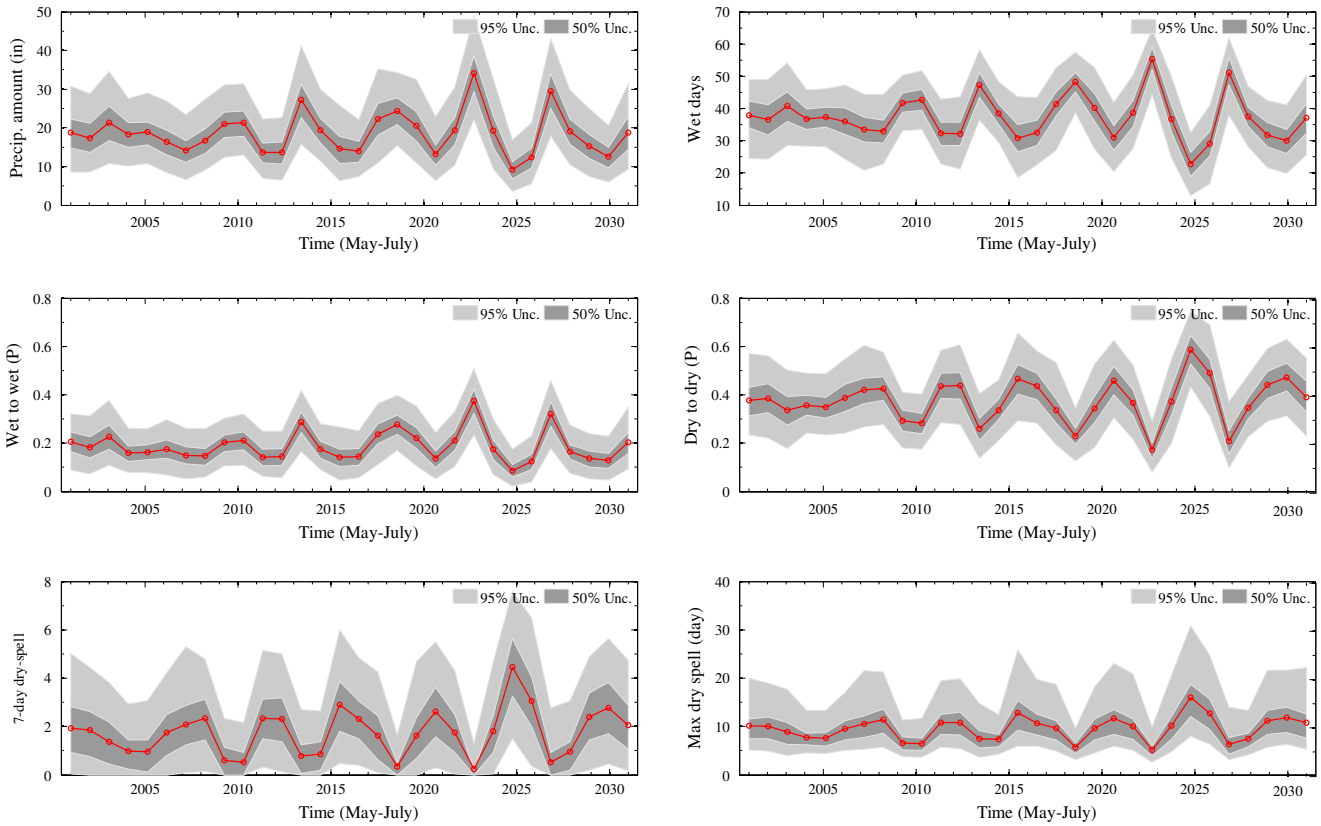


Fig. 15 The MJJ seasonal statistics of NHMM-simulated rainfall averaged over all 16 stations. Plotted is the median of 50 NHMM simulations (*circle*). The input "predictor" is the wavelet components identified by multiple hydroclimatic proxies. The *error bars* indicate the entire range of the 50 simulations, with the inter-quartile range given by the *dark shaded area*. The *red line* is the median of the simulations

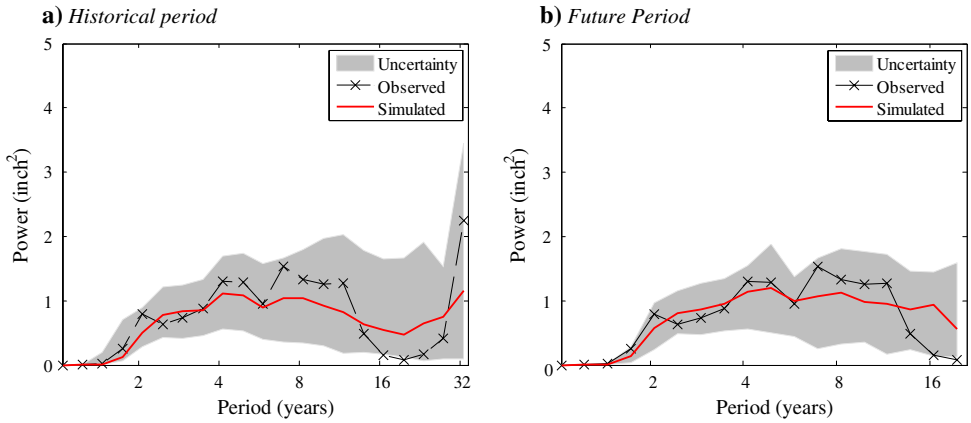


Fig. 16 Comparison of global wavelet power spectrum of 50 simulations for the simulated MJJ rainfall for **a** the historical period (shown in Fig. 14) and **b** the future period (shown in Fig. 15) using the NHMM model

atmosphere or land aspects of climate from numerical models of global or regional climate. The WARM methodology can be applied to any of these proxies individually. The user needs to select the frequency bands whose variations are deemed important to reproduce. This can be done considering all possible frequencies if a discrete wavelet transform based decomposition is used, or using an appropriate statistical significance criterion applied to the global spectrum of the variable of interest. These are necessarily subjective choices—even if a statistical significance test is used, there is subjectivity associated with the particular choice of the null hypothesis and with the significance level selected. Recall that the strategy used here partitions the original time series into a series of components (including one designated as noise) and then all these components are modeled using an autoregressive model. Since all the

components are modeled, to a great extent, the subjectivity applies only to the specific decomposition considered. We have demonstrated that if a single AR model is applied to the raw time series, then it is not always possible to simulate the frequency structure in the original time series. This is largely due to the criteria used to choose the order of the "best" AR model applied. The application of these criteria typically leads to a non-robust choice of the final basis functions used, that is biased toward the high-frequency variations since there is more information available as to those components. The WARM decomposition used here overcomes this limitation by allowing the user to select lower frequency bands specifically based on experience with many other similar time series that are longer, or based on the spectral attributes of the series under investigation. The application of the standard AR approach to each one of these components and the subsequent aggregation of all components then allows for a better reproduction of both high- and low-frequency components in the original spectrum of the series. Of course, the sample spectrum could correspond to a white or red noise process or may exhibit spurious low-frequency variations even if the generating process of that sample is a white noise process. However, if this is the case then averaging over many such sample spectra that are reproduced would still give us the original stochastic process in the region of interest.

2. The simulated climate proxies are then used as predictors of daily rainfall attributes for the season under consideration using the NHMM. Here, the first question is which variables to select as predictors. In this study, we first showed that the regionally averaged seasonal rainfall is a reasonable predictor of the attributes of the daily rainfall at each station that we considered of interest. These include the average daily rainfall amount and frequency, and the Markov chain transition probabilities. Consequently, it was felt that climate proxies that are good predictors of seasonal rainfall may be useful for developing the coupled WARM-NHMM model. These were identified either based on regional proximity (e.g., the two tree ring chronologies and PDSI), or from a correlation analysis using historical data combined with an assessment of the moisture source regions and potential teleconnection controls for South Florida rainfall. These choices are still inherently subjective.
3. The NHMM takes the suite of potential predictors and then uses a posterior likelihood maximization approach to identify the subset of predictors that give the best results in terms of maximizing the likelihood of the daily rainfall sequences observed at all 16 rain gauge locations considered in the analysis. Consequently, to an extent, the effects of the subjective choice of predictors are mitigated by a scoring algorithm that uses predictors based on their performance on daily rainfall attributes. Of course, there is still subjectivity in the entire process, as there is with any statistical modeling approach that includes predictor selection. On the other hand, physically based climate models that could be used for the purpose still require considerable statistical post-processing to produce reasonable results for high-resolution rainfall, and the application of these statistical corrections requires similar subjective choices as those made herein.
4. On the whole, the WARM-NHMM modeling system was remarkably successful at generating the high- and low-frequency attributes of the daily rainfall in the South Florida region for the MJJ season relative to one's prior expectations for this location and season. Extension of this work to other seasons and for anthropogenic change is feasible.

Acknowledgments This work was funded by South Florida Water Management District.

References

Box GEP, Jenkins G (1970) Time series analysis, forecasting and control. Holden-Day, San Francisco

Chen HJ, Lee KC, Murphy-Chutorian E, Triesch J (2005) Toward a unified probabilistic framework for object recognition and segmentation. Adv Visual Comput Proc 3804:108–117

Chui CK (1992) An Introduction to Wavelets, Wavelet Analysis and its Application Vol. 1. Academic Press, Boston

Cook ER, Woodhouse CA, Eakin CM, Meko DM, Stahle DW (2004) Long-term aridity changes in the western United States. Science 306(5698):1015–1018

Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. J Royal Stat Soc B 39:1–38

Enfield DB, Mestas-Nunez AM, Trimble PJ (2001) The Atlantic multidecadal oscillation and its relation to rainfall and river flows in the continental US. Geophys Res Lett 28(10):2077–2080

Farge M (1992) Wavelet transforms and their applications to turbulence. Ann Rev Fluid Mech 24:395–457

Foufoula-Georgiou E, Kumar P (1995) Wavelets in geophysics. Academic Press, New York

Gelman A, Carlin J, Stern H, Rubin D (2003) Bayesian data analysis. Chapman & Hall/CRC, London

Godsill S, Doucet A, West M (2001) Maximum a posteriori sequence estimation using Monte Carlo particle filters. Ann Inst Stat Math 53(1):82–96

Gray ST, Graumlich LJ, Betancourt JL, Pederson GT (2004) A tree-ring based reconstruction of the Atlantic Multidecadal Oscillation since 1567 AD. Geophys Res Lett, 31(12)

Hubbard BB (1996) The World according to wavelets: the story of a mathematical technique in the making. AK Peters, Massachusetts

Hue C, Le Cadre JP, Perez P (2002) Sequential Monte Carlo methods for multiple target tracking and data fusion. IEEE Trans Signal Process 50(2):309–325

Hughes JP, Guttorp P (1994) A class of stochastic models for relating synoptic atmospheric patterns to regional hydrologic phenomena. Water Resour Res 30(5):1535–1546

Hughes JP, Guttorp P, Charles SP (1999) A non-homogeneous hidden Markov model for precipitation occurrence. J R Stat Soc Ser C Appl Stat 48:15–30

Kaplan A, Cane M, Kushnir Y, Clement A, Blumenthal M, Rajagopalan B (1998) Analyses of global sea surface temperature 1856–1991. J Geophys Res 103(C9):18567–18589

Koutsoyiannis D (1994) A stochastic disaggregation method for design storm and flood synthesis. J Hydrol 156(1–4):193–225

Kulkarni JR (2000) Wavelet analysis of the association between the Southern Oscillation and the Indian Summer Monsoon. Int J Climatol 20(1):89–104

Kwon H-H, Khalil AF, Lall U (2006a) Modeling the Atlantic multidecadal oscillation and the associated rainfall variability in S. Florida. Columbia University, New York

Kwon H-H, Khalil AF, Lall U (2008) Non-homogeneous hidden Markov model based daily rainfall simulation using seasonal climate forcing: application to Everglades National Park, Florida. Water Resour Res (in press)

Kwon H-H, Lall U, Khalil AF (2007) Stochastic simulation model for nonstationary time series using an autoregressive wavelet decomposition: Applications to rainfall and temperature. Water Resour Res, 43(W05407), doi:10.1029/2006WR005258

Kwon H-H, Lall U, Moon Y-I, Khalil AF, Ahn H (2006b) Episodic interannual climate oscillations and their influence on seasonal rainfall in the Everglades National Park17. Water Resour Res, 42(W11404), doi:10.1029/2006WR0050

McPherson BF, Halley R (1996) The South Florida environment: a region under stress: National Water-Quality Assessment Program, U.S. Geological Survey Circular 1134

Park J, Mann ME (2000) Interannual temperature events and shifts in global temperature: a multiple wavelet correlation approach. Earth Interact 4:1–53

Rabiner LR (1989) A tutorial on hidden Markov-models and selected applications in speech recognition. Proc IEEE 77(2):257–286

Ridgeway G, Madigan D (2003) A sequential Monte Carlo method for Bayesian analysis of massive datasets. Data Min Knowl Discov 7(3):301–319

Robertson AW, Kirshner S, Smyth PJ (2003) Hidden Markov models for modeling daily rainfall occurrence over Brazil. Technical Report ICS-TR 03–27, Information and Computer Science. University of California, Irvine

Robertson AW, Kirshner S, Smyth P (2004) Downscaling of daily rainfall occurrence over northeast Brazil using a hidden Markov model. J Clim 17(22):4407–4424

Schmidt N, Lipp EK, Rose JB, Luther ME (2001) ENSO influences on seasonal rainfall and river discharge in Florida. J Clim 14(4):615–628

Stedinger JR, Vogel RM (1984) Disaggregation procedures for generating serially correlated flow vectors. Water Resour Res 20(1):47–56

Thomas HA, Fiering MB (1962) Mathematic synthesis of streamflow sequences for analysis of river basins by simulation. In: Mass A (ed) Design of water resources systems. Harvard University press, Massachusetts, pp 459–493

Torrence C, Compo GP (1998) A practical guide to wavelet analysis. Bull Am Meteorol Soc 79(1):61–78

Trimble PJ, Santee ER, Neidrauer CJ (1997) Including the effects of solar activity for more efficient water management: an application of neural networks. In: Second International Workshop on Artificial Intelligence Applications in Solar-Terrestrial Physics, Sweden

Trimble PJ, Santee ER, Neidrauer CJ (1998) A refined approach to lake Okeechobee water management: An application of climate forecast, Special report. South Florida Water Management District, Florida

Trimble PJ, Trimble B (1998) Recognition and predictability of climate variability within south-central Florida. In: 23rd Annual Climate Diagnostics and Prediction Workshop, U.S. Dept. of Comm., NOAA, NWS, NCEP., Rosenstiel School of Marine and Atmospheric Science, University of Miami, Miami

Tsionas EG (2001) Bayesian multivariate Poisson regression. Commun Stat Theory Methods 30(2):243–255

Tucker A, Liu XH (2003) Learning dynamic Bayesian networks from multivariate time series with changing dependencies. Adv Intell Data Anal V 2810:100–110

Valencia D, Schaake JC (1973) Disaggregation Processes in Stochastic Hydrology. Water Resour Res 9(3):580–585

Van Lent TJ (1993). Analysis of the historical Taylor Slough rainfall/flow relationship. University of Virginia, Environmental Sciences Department

Viterbi AJ (1967) Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. IEEE Trans Inform Theory It13(2):260

Wang B, Wang Y (1996) Temporal structure of the Southern Oscillation as revealed by waveform and wavelet analysis. J Clim 9(7):1586–1598

Weng HY, Lau KM (1994) Wavelets, Period-Doubling, and Time-Frequency Localization with Application to Organization of Convection over the Tropical Western Pacific. J Atmos Sci 51(17):2523–2541

Yevjevich V (1972) Stochastic processes in hydrology, Water Resources publication, Fort Collins, 30 CO

Zhang EY, Trimble P (1996) Predicting effects of climate fluctuations for water management by applying neural network. World Resour Rev 8(3):334