Simulating Hydrological Drought Properties at Different Spatial Units in the United States Based on Wavelet–Bayesian Regression Approach

Ashok K. Mishra*, +
Department of Biological and Agricultural Engineering, Texas A&M University, College Station, Texas

Vijay P. Singh
Department of Biological and Agricultural Engineering, and Department of Civil Engineering, Texas A&M University, College Station, Texas

Received 24 March 2012; accepted 27 July 2012

ABSTRACT: Because of their stochastic nature, droughts vary in space and time, and therefore quantifying droughts at different time units is important for water resources planning. The authors investigated the relationship between meteorological variables and hydrological drought properties using the Palmer hydrological drought index (PHDI). Twenty different spatial units were chosen from the unit of a climatic division to a regional unit across the United States. The relationship between meteorological variables and PHDI was investigated using a wavelet–Bayesian regression model, which enhances the modeling

* Current affiliation: Pacific Northwest National Laboratory, Richland, Washington.

+ Corresponding author address: Ashok K. Mishra, Pacific Northwest National Laboratory, P.O. Box 999, Richland, WA 99352.
E-mail address: ashok.mishra@pnnl.gov

DOI: 10.1175/2012EI000453.1
strength of a simple Bayesian regression model. Further, the wavelet–Bayesian regression model was tested for the predictability of global climate models (GCMs) to simulate PHDI, which will also help understand their role for downscaling purposes.

**KEYWORDS:** Palmer hydrological drought index; Bayesian regression; Wavelet analysis; General circulation model

1. Introduction

The demand for water has significantly increased in many parts of the world because of the growth of population and the expansion of agricultural, energy, and industrial sectors (Mishra and Singh 2010). The impact of a drought is well known, extending from water supply to socioeconomic impacts in the region. Increasing occurrences of droughts further worsen growing water demands. For example, in the United States the impacts of droughts have increased with an increased number of droughts or an increase in their severity (Wilhite and Hayes 1998), and based on the National Climatic Data Center (National Climatic Data Center 2002) nearly 10% of the total land area experienced either severe or extreme droughts at any given time during the last century. There has been a variety of concepts (Mishra and Singh 2010) applied to modeling droughts, ranging from simplistic approaches to more complex models (Mishra and Singh 2011).

Drought prediction or forecasting plays a major role in risk management, drought preparedness, and mitigation. Drought indices have been commonly used for prediction and simulation (Mishra and Singh 2011), and they are also used for a variety of applications (Quiring 2009). Several models have been proposed in simulating and forecasting droughts. Using a regression model, Kumar and Panu (Kumar and Panu 1997) predicted agricultural drought characteristics considering the variables affecting the grain yield in the region as explanatory variables for agricultural drought forecasting. Rao and Padmanabhan (Rao and Padmanabhan 1984) used the Palmer drought index (PDI) to forecast and simulate PDI series. Mishra and Desai (Mishra and Desai 2005) developed linear stochastic models [Autoregressive Integrated Moving Average (ARIMA) and Seasonal ARIMA (SARIMA)] for forecasting droughts with the use of the standardized precipitation index (SPI) series as a drought quantifying parameter and Han et al. (Han et al. 2010) used an ARIMA model to forecast droughts in the Guanzhong Plain in China. Improving prediction hybrid models, based on the combination of two or more models, has been explored in several studies. Using a hybrid model combining wavelet decomposition and artificial neural network (ANN), Kim and Valdes (Kim and Valdes 2003) improved the ability to forecast an indexed regional drought. Combining a linear stochastic model and a nonlinear artificial neural network model, Mishra et al. (Mishra et al. 2007) forecasted droughts with an SPI series using the advantages of both stochastic and ANN models. Özger et al. (Özger et al. 2012) developed a wavelet and fuzzy logic (WFL) combination model for long-lead-time drought forecasting using the Palmer drought series.

The spatial distribution of droughts varies largely from a climatic division to a regional unit. Therefore, the potential of drought predictability at spatial units is likely to be affected by the variability in meteorological variables. In
previous studies, prediction and simulation of meteorological and agricultural droughts were carried out using meteorological variables. However, in most cases, hydrological droughts were predicted based on historic streamflow data. Since a hydrologic drought occurs at a later stage in the evolution of the drought, it is largely dependent on precipitation and temperature, which can therefore be used to simulate the hydrological drought. Previous studies focused on the Palmer drought severity index (PDSI) derived from a moisture balance model, using historical records of precipitation, temperature, and the local available water capacity of the soil. However, the Palmer hydrological drought index (PHDI) uses a modification of the PDSI to assess long-term moisture anomalies that affect streamflow, groundwater, and water storage.

The primary difference between PDSI and PHDI is based on the ratio of moisture received to moisture required to definitely terminate a drought. The PDSI abruptly comes back to the near-normal level during the first month in a sequence of months with sufficient moisture to end the drought. On the other hand, PHDI is more gradual to a new spell only when the moisture needs associated with recharge, demand, and runoff have been brought back to normal or above normal (Karl et al. 1987). Therefore, PHDI is used as a hydrological drought index along with precipitation and temperature as influencing variables. Another reason for choosing PHDI, precipitation, and temperature is due to the fact that time series are available for 1900–2000 from the National Oceanic and Atmospheric Administration (NOAA)/National Climatic Data Center.

Based on the above discussion, the objective of this study therefore is (i) to develop a hybrid wavelet–Bayesian regression model and compare with Bayesian regression model, (ii) to simulate the hydrological drought time series considering meteorological variables (precipitation and temperature) at different spatial units and compare drought parameters, and (iii) to evaluate the strength of meteorological variables obtained from GCMs for simulating the hydrological drought time series.

2. Methodology

2.1. PHDI

The Palmer hydrological drought index (PHDI) is used to assess long-term moisture supply. The monthly time series generated indicates the severity of a wet or dry spell based on the balance between moisture supply and demand. The PHDI is suitable to quantify the hydrological impacts of droughts (e.g., reservoir levels, groundwater levels, etc.), which take longer to develop and it takes longer to recover from them (Palmer 1965).

The PHDI generally ranges from $-6$ to $+6$, with negative values denoting dry spells and positive values indicating wet spells. In the present study, we have taken different thresholds for identifying severity levels of different droughts: that is, PHDI values less than 0 include all types of drought; PHDI values less than $-1$ include a range of drought from mild drought to extreme drought; PHDI values less than $-2$ include moderate drought to extreme drought; PHDI values less than $-3$ represent severe drought and extreme droughts; and PHDI values less than $-4$ represent extreme droughts.
2.2. Bayesian linear regression

The Bayesian regression model is summarized from Hoff (Hoff 2009). According to the regression model, the sampling distribution of dependent variable $y$ varies with another variable or sets of independent variable $x = (x_1, \ldots, x_p)$,

$$y_i = \beta^T x_i + \epsilon_i, \quad \text{where} \quad \epsilon_i \sim \text{i.i.d. normal}(0, \sigma^2), \quad (1)$$

where $\beta^T$ represents the vector of coefficients associated with a row vector of independent variables $x_i$ and $\epsilon$ is the independent and identically distributed (i.i.d) random normal term with mean zero and standard deviation $\sigma$.

The joint probability density function of $Y$ with observed data $y_1, \ldots, y_n$ conditional upon $X: \{x_1, \ldots, x_n\}$ and values of $\beta$ and $\sigma^2$ are given by

$$p(y_1, \ldots, y_n; x_1, \ldots, x_n, \beta, \sigma^2) = p(y_1; x_1, \beta, \sigma^2) = \prod_{i=1}^{n} p(y_i; x_i, \beta, \sigma^2)$$

$$= \left(2\pi\sigma^2\right)^{-n/2} \exp \left[ -\frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \beta^T x_i)^2 \right]. \quad (2)$$

Considering $y$ to be the $n$-dimensional column vector $(y_1, \ldots, y_n)^T$ and $X$ to be the $n \times p$ matrix, the normal regression model is given by $\{y \mid X, \beta, \sigma^2 \} \sim \text{multivariate normal } (X, \beta, \sigma^2)$, where $I$ is the identity matrix.

The exponent term in the density function [Equation (2)] is maximized when the sum of squared residuals, $\text{SSR}(\beta) = \sum_{i=1}^{n} (y_i - \beta^T x_i)^2$, is minimized,

$$\text{SSR}(\beta) = \sum_{i=1}^{n} (y_i - \beta^T x_i)^2 = (y - X\beta)^T(y - X\beta) = y^T y - 2\beta^T X^T y + \beta^T X^T X \beta. \quad (3)$$

Therefore,

$$p(y \mid X, \beta, \sigma^2) \propto \exp \left\{ -\frac{1}{2\sigma^2} [y^T y - 2\beta^T X^T y + \beta^T X^T X \beta] \right\}. \quad (4)$$

The distribution of $y$ is multivariate normal. The role played by $\beta$ in the exponent looks very similar to that played by $y$. This suggests a multivariate normal prior distribution for $\beta$. Considering $\beta \sim \text{multivariate normal } (\beta_0, \Sigma_0)$,

$$p(\beta \mid y, X, \sigma^2) \propto p(y \mid X, \beta, \sigma^2) \times p(\beta) \quad \text{and} \quad (5)$$

$$\propto \exp \left\{ -\frac{1}{2\sigma^2} (-2\beta^T X^T y/\sigma^2 + \beta^T X^T X \beta/\sigma^2) - \frac{1}{2} (-2\beta^T \Sigma_0^{-1} \beta_0 + \beta^T \Sigma_0^{-1} \beta) \right\}$$

$$= \exp \left\{ \beta^T (\Sigma_0^{-1} \beta_0 + X^T y/\sigma^2) - \frac{1}{2} \beta^T (\Sigma_0^{-1} + X^T X/\sigma^2) \beta \right\}. \quad (6)$$

This is proportional to a multivariate normal density, with
\[
\text{var}[\mathbf{\beta} \mid \mathbf{y}, \mathbf{X}, \sigma^2] = (\Sigma_0^{-1} + \mathbf{X}^T \mathbf{X} / \sigma^2)^{-1} \quad \text{and} \\
E[\mathbf{\beta} \mid \mathbf{y}, \mathbf{X}, \sigma^2] = (\Sigma_0^{-1} + \mathbf{X}^T \mathbf{X} / \sigma^2)^{-1}(\Sigma_0^{-1} \mathbf{\beta}_0 + \mathbf{X}^T \mathbf{y} / \sigma^2).
\]

The semiconjugate prior distribution for \(\sigma^2\) is an inverse-gamma distribution. Considering \(\gamma = 1 / \sigma^2\) as the measurement precision and \(\gamma \sim \text{gamma}(v_0/2, v_0 \sigma_0^2/2)\),

\[
p(\gamma \mid \mathbf{y}, \mathbf{X}, \mathbf{\beta}) \propto p(\mathbf{y} \mid \mathbf{X}, \mathbf{\beta}) p(\gamma),
\]

where \(v_0\) is the sample size of prior distribution and \(\sigma_0\) is the sample variance, this leads to

\[
p(\gamma \mid \mathbf{y}, \mathbf{X}, \mathbf{\beta}) = \gamma^{(v_0+n)/2-1} \exp\{-\gamma[v_0 \sigma_0^2 + \text{SSR}(\mathbf{\beta})]/2\},
\]

which is recognized as the gamma density, so that

\[
\{\sigma^2 \mid \mathbf{y}, \mathbf{X}, \mathbf{\beta}\} \sim \text{inverse - gamma}\{[v_0 + n]/2, [v_0 \sigma_0^2 + \text{SSR}(\mathbf{\beta})]/2\}.
\]

The Gibbs sampler is used to approximate the joint posterior distribution \(p(\mathbf{\beta}, \sigma^2 \mid \mathbf{y}, \mathbf{X})\).

Given current values \(\{\beta^{(s)}, \sigma^{2(s)}\}\), new values can be generated by the following:

(i) updating \(\mathbf{\beta}\): compute \(V = \text{var}[\mathbf{\beta} \mid \mathbf{y}, \mathbf{X}, \sigma^{2(s)}]\) and \(m = E[\mathbf{\beta} \mid \mathbf{y}, \mathbf{X}, \sigma^{2(s)}]\) then sample \(\beta^{(s+1)} \sim \text{multivariate normal}(m, V)\);

(ii) updating \(\sigma^2\): compute \(\text{SSR}[\beta^{(s+1)}]\) and sample \(\sigma^{2(s+1)} \sim \text{inverse gamma}\{[v_0 + n]/2, [v_0 \sigma_0^2 + \text{SSR}(\beta^{(s+1)})]/2\}\).

### 2.3. Continuous wavelet transform

A continuous wavelet transform (CWT) decomposes a PHDI time series into wavelets and produces coefficients at a given scale. The CWT basis functions are scaled and shifted versions of the time-localized mother wavelet. A Morlet wavelet is one of the many wavelet functions that has a zero mean and is localized in both frequency and time. It provides a good balance between time and frequency localizations and is therefore preferred for application. It can be represented as (Grinsted et al. 2004)

\[
\psi(\eta) = \pi^{-1/4} e^{-i\omega \eta - 0.5 \eta^2},
\]

where \(\omega\) is the dimensionless frequency and \(\eta\) is the dimensionless time parameter. The wavelet is stretched in time \(t\) by varying its scale \((s)\), so that \(\eta = s/t\). When using wavelets for feature extraction purposes, the Morlet wavelet (with \(\omega = 6\)) is a good choice, since it satisfies the admissibility condition (Farge 1992; Torrence and Compo 1998). For a given wavelet \(\psi_0(\eta)\), it is assumed that \(X_j\) is a time series of length \(N\) \((X_j, i = 1, \ldots, N)\) with equal time spacing \(\delta t\). The continuous wavelet transform of a discrete sequence \(X_j\) is defined as convolution of \(X_j\) with the scaled and translated wavelet \(\psi_0(\eta)\).
CWT decomposes a PHDI time series into time–frequency space, enabling the identification of both the dominant modes of variability and how those modes vary with time.

2.4. Goodness-of-fit test

To compare observed and simulated drought time series, the goodness of fit was calculated based on the correlation coefficient (CC), root-mean-square error (RMSE), and mean bias error (MBE). If \( O_i \) and \( S_i \) represent observed and simulated drought time series, then

\[
CC = \frac{n \sum O_i S_i - (\sum O_i)(\sum S_i)}{\sqrt{n(\sum O_i^2) - (\sum O_i)^2} \sqrt{n(\sum S_i^2) - (\sum S_i)^2}}.
\]  

(13)

If \( e_i = O_i - S_i \) denotes individual model-prediction errors, then model performances are based on statistical summaries of \( e_i (i = 1, 2, 3, \ldots, n) \),

\[
RMSE = \left( \frac{1}{n} \sum_{i=1}^{n} |e_i|^2 \right)^{1/2} \quad \text{and}
\]

\[
MBE = \frac{1}{n} \sum_{i=1}^{n} e_i.
\]  

(14)

(15)

3. Study area and data used

Different spatial units were chosen to test the potential of precipitation and temperature to simulate hydrological droughts using PHDI. The definition of spatial units used in this paper is based on the following notion: (i) a climatic division is the smallest unit used in this study, (ii) the state (e.g., Texas) is a combination of several climatic divisions (units), and (iii) the regional unit is a combination of several states. The spatial units include (see Table 1 and Figure 1) all climatic divisions (1–10) of the state of Texas, the state of Texas, and different regions of the United States (northeast, east north-central, central, southeast, west north-central, south, southwest, northwest, and west regions). Monthly precipitation, temperature, and PHDI values that are available from 1900 to 2000 were retrieved from the NOAA/National Climatic Data Center. For both precipitation and temperature, monthly averages within a climatic division were calculated by giving equal weight to stations. To adjust the climatic division average, the model described by Karl et al. (Karl et al. 1986) was used such that all stations end their climatological day at midnight (i.e., climatological day coincides with calendar day).
Statistical properties of precipitation across different spatial units are shown in Table 1. Precipitation in the state of Texas is not evenly distributed, and the mean annual precipitation distribution correlates roughly with longitude and varies little from north to south. The maximum mean annual precipitation was observed for

<table>
<thead>
<tr>
<th>Serial No.</th>
<th>Spatial location</th>
<th>Mean (cm)</th>
<th>Coef of variation (%)</th>
<th>Kurtosis</th>
<th>Skewness</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Climatic division 1 (Texas)</td>
<td>47.92</td>
<td>22.18</td>
<td>5.88</td>
<td>0.71</td>
</tr>
<tr>
<td>2</td>
<td>Climatic division 2 (Texas)</td>
<td>59.85</td>
<td>22.61</td>
<td>4.32</td>
<td>0.45</td>
</tr>
<tr>
<td>3</td>
<td>Climatic division 3 (Texas)</td>
<td>87.23</td>
<td>20.15</td>
<td>2.58</td>
<td>−0.01</td>
</tr>
<tr>
<td>4</td>
<td>Climatic division 4 (Texas)</td>
<td>118.30</td>
<td>18.60</td>
<td>2.54</td>
<td>0.31</td>
</tr>
<tr>
<td>5</td>
<td>Climatic division 5 (Texas)</td>
<td>31.58</td>
<td>29.27</td>
<td>4.81</td>
<td>0.85</td>
</tr>
<tr>
<td>6</td>
<td>Climatic division 6 (Texas)</td>
<td>64.54</td>
<td>26.24</td>
<td>3.05</td>
<td>0.44</td>
</tr>
<tr>
<td>7</td>
<td>Climatic division 7 (Texas)</td>
<td>88.30</td>
<td>24.07</td>
<td>2.49</td>
<td>0.13</td>
</tr>
<tr>
<td>8</td>
<td>Climatic division 8 (Texas)</td>
<td>121.18</td>
<td>21.91</td>
<td>2.67</td>
<td>0.38</td>
</tr>
<tr>
<td>9</td>
<td>Climatic division 9 (Texas)</td>
<td>58.96</td>
<td>24.33</td>
<td>2.54</td>
<td>0.30</td>
</tr>
<tr>
<td>10</td>
<td>Climatic division 10 (Texas)</td>
<td>63.58</td>
<td>22.75</td>
<td>3.59</td>
<td>0.60</td>
</tr>
<tr>
<td>11</td>
<td>State of Texas</td>
<td>71.19</td>
<td>18.86</td>
<td>2.92</td>
<td>0.15</td>
</tr>
<tr>
<td>12</td>
<td>Northeast region (United States)</td>
<td>104.35</td>
<td>9.69</td>
<td>3.66</td>
<td>0.39</td>
</tr>
<tr>
<td>13</td>
<td>East north-central region (United States)</td>
<td>75.63</td>
<td>11.32</td>
<td>2.91</td>
<td>−0.30</td>
</tr>
<tr>
<td>14</td>
<td>Central region (United States)</td>
<td>107.79</td>
<td>10.84</td>
<td>2.88</td>
<td>−0.09</td>
</tr>
<tr>
<td>15</td>
<td>Southeast region (United States)</td>
<td>127.71</td>
<td>10.62</td>
<td>2.45</td>
<td>0.07</td>
</tr>
<tr>
<td>16</td>
<td>West north-central region (United States)</td>
<td>42.99</td>
<td>12.79</td>
<td>3.13</td>
<td>0.13</td>
</tr>
<tr>
<td>17</td>
<td>South region (United States)</td>
<td>90.02</td>
<td>13.78</td>
<td>2.91</td>
<td>−0.08</td>
</tr>
<tr>
<td>18</td>
<td>Southwest region (United States)</td>
<td>34.36</td>
<td>16.59</td>
<td>4.09</td>
<td>0.35</td>
</tr>
<tr>
<td>19</td>
<td>Northwest region (United States)</td>
<td>68.37</td>
<td>14.11</td>
<td>2.63</td>
<td>0.21</td>
</tr>
<tr>
<td>20</td>
<td>West region (United States)</td>
<td>42.82</td>
<td>25.57</td>
<td>3.69</td>
<td>0.87</td>
</tr>
</tbody>
</table>

Statistical properties of precipitation across different spatial units are shown in Table 1. Precipitation in the state of Texas is not evenly distributed, and the mean annual precipitation distribution correlates roughly with longitude and varies little from north to south. The maximum mean annual precipitation was observed for

Figure 1. (a) Nine climatically consistent regions within the contiguous United States identified by the National Climatic Data Center (Karl and Koss 1984). (b) Ten climate divisions located within the state of Texas.
climatic division 8 (121 cm) and climatic division 4 (118 cm) located in the far eastern part of Texas, whereas the minimum was observed in climatic division 5 (32 cm) located in the far western part of Texas. Among climatic divisions in Texas a relatively higher amount of coefficient of variation (CV) was observed in the regions (climatic divisions 5 and 6) and a lower CV was observed in climate division 4.

Based on the regional units, relatively lower CV was observed in the northeast (9.69%), central (10.84), and southeast (10.62%), whereas higher CV was observed in the spatial unit located in the west region (25.57%). In general, the CV for climatic divisions within Texas observed to be higher than regional units, except the west region. Among regional units, the maximum mean annual precipitation was observed to be higher in the southeast region (128 cm) followed by the central region (108 cm) and northeast region (104 cm), whereas the minimum was observed in the southwest region (34 cm) followed by the west and west north-central regions (43 cm). Positive kurtosis and skewness were observed in the maximum number of selected spatial units.

4. Results and discussion

4.1. Comparison of drought properties among spatial units

Good correlation was observed among neighboring climatic divisions (Table 2): for example, the PHDI time series in climatic division 1 had a correlation strength of 0.84 (climatic division 2) and 0.71 (climatic division 5), which happened to be neighboring climatic divisions. This suggests that droughts are regional in nature for most parts of Texas. Therefore, the drought causing variables are likely to share similar association at a regional scale. The PHDI time series for the whole of Texas shares a stronger correlation coefficient (>0.74) with all climatic divisions, except climate division 10, with a correlation coefficient of 0.6. The PHDI time series for larger spatial units beyond Texas: that is, the south region of United States, which includes many states (Texas, Oklahoma, Kansas, Arkansas, Louisiana, and Mississippi), shares a good correlation with climatic divisions of Texas, except climatic divisions of 5, 9, and 10. It is worth noting that the correlation strength remains strong among climatic divisions, state, and region. However, no correlations were observed at higher spatial units that comprise multiple states from different parts of the United States. This suggests geography matters at spatial units when precipitation distribution is considered.

The number of droughts, maximum drought severity, and maximum drought duration were compared among spatial units. These drought parameters were identified by using the theory of runs (Mishra and Singh 2010) by defining a threshold level. In this case, all droughts were considered for which PHDI was consecutively less than $-1.0$. Therefore, those drought events were considered for which the PHDI values were consecutively less than $-1.0$. Based on this criterion, the number of droughts that occurred among climate divisions of Texas included climatic division 1 (61), climatic division 2 (58), climatic division 3 (49), climatic division 4 (65), climatic division 5 (60), climatic division 6 (51), climatic division 7 (50), climatic division 8 (60), climatic division 9 (65), climatic division 10 (58), and the whole of Texas (45). Among the climatic divisions, the number of droughts
Table 2. Correlation matrix between selected spatial units (first row and column represent the serial number of spatial units from Table 1).

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
<th>20</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.00</td>
<td>0.84</td>
<td>0.67</td>
<td>0.51</td>
<td>0.71</td>
<td>0.62</td>
<td>0.59</td>
<td>0.50</td>
<td>0.55</td>
<td>0.43</td>
<td>0.76</td>
<td>0.03</td>
<td>0.27</td>
<td>0.23</td>
<td>0.13</td>
<td>0.18</td>
<td>0.72</td>
<td>0.58</td>
<td>0.08</td>
<td>0.04</td>
</tr>
<tr>
<td>2</td>
<td>0.84</td>
<td>1.00</td>
<td>0.83</td>
<td>0.62</td>
<td>0.71</td>
<td>0.77</td>
<td>0.71</td>
<td>0.61</td>
<td>0.67</td>
<td>0.51</td>
<td>0.87</td>
<td>0.04</td>
<td>0.27</td>
<td>0.22</td>
<td>0.17</td>
<td>0.09</td>
<td>0.78</td>
<td>0.48</td>
<td>0.05</td>
<td>0.01</td>
</tr>
<tr>
<td>3</td>
<td>0.67</td>
<td>0.83</td>
<td>1.00</td>
<td>0.82</td>
<td>0.61</td>
<td>0.78</td>
<td>0.78</td>
<td>0.72</td>
<td>0.64</td>
<td>0.44</td>
<td>0.90</td>
<td>0.03</td>
<td>0.26</td>
<td>0.24</td>
<td>0.21</td>
<td>0.08</td>
<td>0.85</td>
<td>0.39</td>
<td>0.08</td>
<td>0.03</td>
</tr>
<tr>
<td>4</td>
<td>0.51</td>
<td>0.62</td>
<td>0.82</td>
<td>1.00</td>
<td>0.50</td>
<td>0.62</td>
<td>0.75</td>
<td>0.84</td>
<td>0.54</td>
<td>0.39</td>
<td>0.82</td>
<td>0.12</td>
<td>0.29</td>
<td>0.31</td>
<td>0.25</td>
<td>0.11</td>
<td>0.86</td>
<td>0.36</td>
<td>0.04</td>
<td>0.03</td>
</tr>
<tr>
<td>5</td>
<td>0.71</td>
<td>0.71</td>
<td>0.61</td>
<td>0.50</td>
<td>1.00</td>
<td>0.72</td>
<td>0.63</td>
<td>0.53</td>
<td>0.69</td>
<td>0.56</td>
<td>0.74</td>
<td>0.00</td>
<td>0.13</td>
<td>0.03</td>
<td>0.02</td>
<td>-0.09</td>
<td>0.61</td>
<td>0.47</td>
<td>-0.20</td>
<td>-0.06</td>
</tr>
<tr>
<td>6</td>
<td>0.62</td>
<td>0.77</td>
<td>0.78</td>
<td>0.62</td>
<td>0.72</td>
<td>1.00</td>
<td>0.82</td>
<td>0.64</td>
<td>0.79</td>
<td>0.58</td>
<td>0.87</td>
<td>-0.02</td>
<td>0.08</td>
<td>0.06</td>
<td>0.11</td>
<td>-0.09</td>
<td>0.69</td>
<td>0.32</td>
<td>-0.20</td>
<td>-0.05</td>
</tr>
<tr>
<td>7</td>
<td>0.59</td>
<td>0.71</td>
<td>0.78</td>
<td>0.75</td>
<td>0.63</td>
<td>0.82</td>
<td>1.00</td>
<td>0.84</td>
<td>0.83</td>
<td>0.69</td>
<td>0.89</td>
<td>0.02</td>
<td>0.22</td>
<td>0.11</td>
<td>0.18</td>
<td>0.03</td>
<td>0.75</td>
<td>0.37</td>
<td>-0.15</td>
<td>-0.01</td>
</tr>
<tr>
<td>8</td>
<td>0.50</td>
<td>0.61</td>
<td>0.72</td>
<td>0.84</td>
<td>0.53</td>
<td>0.64</td>
<td>0.84</td>
<td>1.00</td>
<td>0.65</td>
<td>0.55</td>
<td>0.81</td>
<td>0.06</td>
<td>0.26</td>
<td>0.18</td>
<td>0.25</td>
<td>0.07</td>
<td>0.67</td>
<td>0.34</td>
<td>-0.07</td>
<td>-0.02</td>
</tr>
<tr>
<td>9</td>
<td>0.55</td>
<td>0.67</td>
<td>0.64</td>
<td>0.54</td>
<td>0.69</td>
<td>0.79</td>
<td>0.83</td>
<td>0.65</td>
<td>1.00</td>
<td>0.77</td>
<td>0.78</td>
<td>0.00</td>
<td>0.10</td>
<td>0.02</td>
<td>0.07</td>
<td>-0.05</td>
<td>0.62</td>
<td>0.30</td>
<td>-0.19</td>
<td>-0.13</td>
</tr>
<tr>
<td>10</td>
<td>0.43</td>
<td>0.51</td>
<td>0.44</td>
<td>0.39</td>
<td>0.56</td>
<td>0.58</td>
<td>0.69</td>
<td>0.55</td>
<td>0.77</td>
<td>1.00</td>
<td>0.60</td>
<td>-0.01</td>
<td>-0.04</td>
<td>-0.09</td>
<td>0.02</td>
<td>-0.09</td>
<td>0.42</td>
<td>0.29</td>
<td>-0.16</td>
<td>-0.10</td>
</tr>
<tr>
<td>11</td>
<td>0.76</td>
<td>0.87</td>
<td>0.90</td>
<td>0.82</td>
<td>0.74</td>
<td>0.87</td>
<td>0.89</td>
<td>0.81</td>
<td>0.78</td>
<td>0.60</td>
<td>1.00</td>
<td>0.01</td>
<td>0.24</td>
<td>0.21</td>
<td>0.19</td>
<td>0.06</td>
<td>0.88</td>
<td>0.49</td>
<td>-0.14</td>
<td>-0.01</td>
</tr>
<tr>
<td>12</td>
<td>0.03</td>
<td>0.04</td>
<td>0.03</td>
<td>0.12</td>
<td>0.00</td>
<td>-0.02</td>
<td>0.02</td>
<td>0.06</td>
<td>0.00</td>
<td>-0.01</td>
<td>0.01</td>
<td>1.00</td>
<td>0.32</td>
<td>0.49</td>
<td>0.28</td>
<td>0.10</td>
<td>0.17</td>
<td>-0.12</td>
<td>0.12</td>
<td>-0.08</td>
</tr>
<tr>
<td>13</td>
<td>0.27</td>
<td>0.27</td>
<td>0.26</td>
<td>0.29</td>
<td>0.13</td>
<td>0.08</td>
<td>0.22</td>
<td>0.26</td>
<td>0.10</td>
<td>-0.04</td>
<td>0.24</td>
<td>0.32</td>
<td>1.00</td>
<td>0.51</td>
<td>0.12</td>
<td>0.54</td>
<td>0.39</td>
<td>0.29</td>
<td>0.28</td>
<td>0.28</td>
</tr>
<tr>
<td>14</td>
<td>0.23</td>
<td>0.22</td>
<td>0.24</td>
<td>0.31</td>
<td>0.03</td>
<td>0.06</td>
<td>0.11</td>
<td>0.18</td>
<td>0.02</td>
<td>-0.09</td>
<td>0.21</td>
<td>0.49</td>
<td>0.51</td>
<td>1.00</td>
<td>0.34</td>
<td>0.42</td>
<td>0.45</td>
<td>0.23</td>
<td>0.18</td>
<td>0.09</td>
</tr>
<tr>
<td>15</td>
<td>0.13</td>
<td>0.17</td>
<td>0.21</td>
<td>0.25</td>
<td>0.02</td>
<td>0.11</td>
<td>0.18</td>
<td>0.25</td>
<td>0.07</td>
<td>0.02</td>
<td>0.19</td>
<td>0.28</td>
<td>0.34</td>
<td>1.00</td>
<td>0.19</td>
<td>0.33</td>
<td>0.06</td>
<td>0.08</td>
<td>-0.03</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>0.18</td>
<td>0.09</td>
<td>0.08</td>
<td>0.11</td>
<td>-0.09</td>
<td>-0.09</td>
<td>0.03</td>
<td>0.06</td>
<td>-0.05</td>
<td>-0.09</td>
<td>0.06</td>
<td>0.10</td>
<td>0.54</td>
<td>0.42</td>
<td>0.19</td>
<td>1.00</td>
<td>0.26</td>
<td>0.41</td>
<td>0.40</td>
<td>0.45</td>
</tr>
<tr>
<td>17</td>
<td>0.72</td>
<td>0.78</td>
<td>0.85</td>
<td>0.86</td>
<td>0.61</td>
<td>0.69</td>
<td>0.75</td>
<td>0.77</td>
<td>0.62</td>
<td>0.42</td>
<td>0.88</td>
<td>0.17</td>
<td>0.39</td>
<td>0.45</td>
<td>0.33</td>
<td>0.26</td>
<td>1.00</td>
<td>0.49</td>
<td>0.01</td>
<td>0.07</td>
</tr>
<tr>
<td>18</td>
<td>0.58</td>
<td>0.48</td>
<td>0.39</td>
<td>0.36</td>
<td>0.47</td>
<td>0.32</td>
<td>0.37</td>
<td>0.34</td>
<td>0.30</td>
<td>0.29</td>
<td>0.49</td>
<td>-0.12</td>
<td>0.29</td>
<td>0.23</td>
<td>0.06</td>
<td>0.41</td>
<td>0.49</td>
<td>1.00</td>
<td>-0.02</td>
<td>0.41</td>
</tr>
<tr>
<td>19</td>
<td>-0.08</td>
<td>-0.05</td>
<td>-0.08</td>
<td>-0.04</td>
<td>-0.20</td>
<td>-0.20</td>
<td>-0.15</td>
<td>-0.07</td>
<td>-0.19</td>
<td>-0.16</td>
<td>-0.14</td>
<td>0.12</td>
<td>0.28</td>
<td>0.18</td>
<td>0.08</td>
<td>0.40</td>
<td>0.01</td>
<td>-0.02</td>
<td>1.00</td>
<td>0.45</td>
</tr>
<tr>
<td>20</td>
<td>0.04</td>
<td>0.01</td>
<td>0.03</td>
<td>0.03</td>
<td>-0.06</td>
<td>-0.05</td>
<td>-0.01</td>
<td>-0.02</td>
<td>-0.13</td>
<td>-0.10</td>
<td>-0.01</td>
<td>-0.08</td>
<td>0.28</td>
<td>0.09</td>
<td>-0.03</td>
<td>0.45</td>
<td>0.07</td>
<td>0.41</td>
<td>0.45</td>
<td>1.00</td>
</tr>
</tbody>
</table>
did not follow any pattern, as the maximum number occurred in all parts of Texas. The lowest number of droughts generally occurred in climatic divisions 3, 6, and 7. Another interesting fact was found to be that the number of droughts based on the state of Texas was less in comparison to any climatic division within the state of Texas. The number of droughts at a regional unit was found to be maximum in the regions located in the northeast (70), southeast (71) and northwest (71) regions of the United States. The minimum number of droughts occurred in the west north-central (36) region of the United States, which happened to be the lowest among all the spatial selected regions.

After knowing the number of drought events, it will be good to look at the maximum drought duration and severity in different spatial units (Figure 2). Based on the number of drought events, maximum duration, and maximum severity, different spatial units were compared. Important observations included the following: (i) The lowest number of drought events occurred in the west north-central region, which happened to have witnessed the maximum drought severity and maximum drought duration (99 months); therefore, chances of getting longer duration droughts are higher for this region. (ii) The maximum number of drought events occurring in the regions (southeast and northwest regions) lay within the first five positions in terms of lowest drought duration or severity in comparison to all other spatial units. (iii) The strong correlation coefficient was observed between maximum duration and maximum severity, whereas the same was not true based on the number of events.

4.2. Comparison between wavelet–Bayesian regression and Bayesian regression model

The wavelet–Bayesian regression model depends on the decomposed PHDI time series and therefore the selection of number of bands which carry significant power
is important for model setup. The spectral bands were obtained according to average wavelet spectra of PHDI. In the current discussion the PHDI time series for all spatial units can be separated into six significant bands so that the lower bands (first and second) show the noisy data, while the upper bands (fourth, fifth, and sixth) stand for low frequency variation of PHDI. All the bands carry specific information related to original time series, for example, the higher level bands contain only information on long time cycles of the concerned variable and exclude other properties, such as noisy data, trends, whereas short times only account for noisy data. Therefore, predicting the homogenous (high and low frequency) time series obtained from wavelet decomposition is more stable, which is its major advantage and enables the Bayesian regression models to simulate with more accuracy.

For comparison, the Bayesian regression and wavelet–Bayesian regression models were applied to climatic division 1 located in Texas considering PHDI as the hydrological drought index and precipitation and temperature as meteorological variables. The length of burn-in period is 10 000 and the number of iteration was chosen as 50 000 during the sampling process of the Bayesian regression analysis. The simulation of PHDI was carried out considering 1900–55 as training period and 1956–2000 as testing period (Figure 3). It was observed that the wavelet–Bayesian regression approach was able to better match the pattern and peaks than the Bayesian regression approach. To further quantify the results obtained from training and testing periods, the goodness of fit was calculated using the correlation coefficient (CC), root-mean-square error (RMSE), and mean bias error (MBE). Based on CC it was 0.35 and 0.27 (Bayesian regression) and 0.78 and 0.65 (wavelet–Bayesian regression) during training and testing periods, respectively. For the Bayesian regression, RMSE was 2.70 and 2.38, whereas for the wavelet–Bayesian regression it was 1.80 and 1.84 during training and testing periods, respectively. Similarly, MBE was 0.07 and 0.34 for the Bayesian regression.
whereas for the wavelet–Bayesian regression it was $-0.014$ and $0.018$ during training and testing periods, respectively. From these three goodness-of-fit tests, it can be observed that the wavelet–Bayesian regression performed better than did the Bayesian regression.

4.3. Application of wavelet–Bayesian regression model to different spatial units

The Bayesian–wavelet regression was applied to 20 selected spatial units and their goodness-of-fit values were calculated for comparison in terms of the predictability of hydrological droughts (Figure 4). Hydrological droughts were
simulated considering 1900–55 as a training period and 1956–2000 as a testing period. The highest (ranked first) CC for hydrologic drought simulation was observed with a CC value of 0.81 in climatic division 5 (Texas) and with a CC value of 0.72 in the northwest region (United States) during training and testing periods, respectively. However, based on the other two goodness-of-fit tests (RMSE and MBE), the northwest region (United States) ranked 3rd and 18th during the testing period. Therefore, it is noted that all three goodness-of-fit tests ranked differently in identifying the regions having better hydrological drought predictability. Based on CC, the observation included the following: (i) Higher correlation coefficients (between 0.7 and 0.8) were observed for the climatic division in Texas for the training period. (ii) During the testing period, the performances were comparatively lower (CC values between 0.5 and 0.6) for climatic divisions in and including the state of Texas. The difference between training and testing periods RMSE value was very small (0–0.04) for climatic division 1 in Texas and the central, southeast, and northwest regions of the United States and comparatively large (climatic divisions 5 and 9) in the southwestern part of Texas. The variability in RMSE values did not follow a clear pattern. Based on MBE, a general pattern was observed in the predictability of hydrological drought using the wavelet Bayesian regression approach with a negative bias in the training period except for a few regions. Within the state of Texas, climatic divisions (9 and 10) had higher positive bias values (0.4 and 0.44), whereas at larger spatial units higher bias occurred in the northeast region (0.48), east north-central region (0.66), central region (0.45), southeast region (0.55), and northwest region (0.49) of the United States.

The autocorrelation function (ACF) coefficient, which is the correlation with its own past values, can be considered as a form of persistence. The ACF was calculated for actual, training, and testing time series and based on observations: (i) the PHDI time series had a stronger persistence with lag-1 ACF values varying between 0.87 and 0.97 and (ii) the lag-1 ACF obtained from the simulated time series was higher than that from the actual time series, leading to higher persistence in the time series.

After quantifying the simulated PHDI time series, it will be appropriate to compare the annual drought characteristics (severity and duration). As was observed in the previous section, the wavelet–Bayesian regression performed better in capturing peaks; therefore, two thresholds (0 and –2) were chosen to further analyze the relationship between observed and simulated annual drought duration and severity using PHDI values. In general, the performance for annual drought severity based on the zero thresholds was slightly higher than that with –2 threshold in most of the regions (Figure 5) based on CC, RMSE, and MBE. Based on the annual drought severity obtained from observed and simulated PHDI time series using CC as a performance measure (Figure 5a), several observations were made, including the following: (i) the CC values differed among spatial units as well as between threshold levels when observed and simulated annual severity values were compared; (ii) based on the two threshold levels for the climatic divisions located within Texas, lower CC values were observed for climatic divisions 2, 3, and 5; (iii) better CC values were observed for climatic divisions 4 and 9 based on the zero threshold level, but the CC values of climatic division 4 reduced significantly when threshold level increased to –2; and (iv) for larger spatial units, higher CC values
were observed for the northeast and southeast regions and poor CC was observed for the southwest region. The performance measured for simulated annual drought severity based on RMSE seemed to be poor when compared with the observed annual drought severity at both threshold levels; however, by comparing among the spatial units, the lower RMSE was observed for the southeastern United States (Figure 5b). Similarly, the performance based on MBE was poor for annual drought severity, except for the spatial units of the northeast, east north-central, central, southeast, and northwest regions (Figure 5c). Positive biases were observed in most of the spatial units with maximum values in the southwest and west regions. This clearly explains (Figure 5c) that the positive bias values were due to the lower simulated values of annual drought severity with respect to the actual values.

Figure 5. Comparison between observed and simulated annual drought severity at two threshold levels for selected spatial locations (simulated period: 1956–2000).
Based on the goodness of fit, comparison between observed and simulated annual drought durations at two threshold levels for selected spatial units are shown in Figure 6. The CC based on the zero threshold was higher than the $-2$ threshold level, with the higher difference being observed for climatic divisions (1–6) located in Texas (Figure 6a). Among regional spatial units, the better performance based on the CC values occurred in the northeast, east north-central, central, and northwest regions and these regions were supplemented by the lower RMSE values (Figure 6b). An important observation was made using MBE at two threshold levels (Figure 6c): (i) using the $-2$ threshold, the positive bias was noted among spatial units; (ii) using the zero threshold level, negative bias values were seen in 17 out of a total 20 spatial units; and (iii) it is worth noting that the bias varied with the type of drought time series used, with the highest values observed in annual drought.
severity followed by annual drought duration and the PHDI time series. This is because the wavelet–Bayesian regression model is able to simulate the PHDI time series better in comparison to annual drought duration and severity. Since the annual drought severity depends on the area under curve in negative PHDI domain and simulated PHDI time series is unable to capture peaks well, larger bias is therefore observed in annual drought severity.

4.4. Spatiotemporal comparison based on annual drought severity and duration

Annual drought severities at two different thresholds (0 and $-2$) are plotted for selected spatial units as shown in Figure 7. Based on these comparison, the
following was observed: (i) The simulated annual drought severity could not capture the actual maximum annual drought severity; for example the maximum annual drought based on the observed PHDI time series reached about 70, whereas based on the simulated data it reached 55. (ii) The simulated annual drought time series could capture the units of actual drought events; however, the limitation included frequent indications of drought events in the simulated time series that did not occur in the observed time series. (iii) Even though all spatial units were not affected by the drought in the observed time series, the simulated time series indicated droughts in a larger number of spatial units. (iv) The performance improved when the threshold of both observed and simulated time series changed to $-2$, as the spatial and temporal units were more pronounced.

Comparison between observed and simulated annual drought durations for different spatiotemporal units is shown in Figure 8. Based on the observed PHDI time series at the zero threshold the maximum drought durations were observed for a larger number of spatial units observed in five time periods (1950–55, 1965–70, 1990–95, and 1995–2000 and in the year 2000). However, based on the simulated time series the maximum annual drought durations for larger number of spatial units were observed more often, including the drought epochs observed from observed data. The accuracy of simulating maximum annual drought duration increased when the threshold level changed to $-2$ as it was able to replicate droughts during the 1965s and the 1990s as noted in the observed time series.

### 4.5. Predictability of PHDI based on GCM outputs

This section discusses the predictability of PHDI using global climate model output and observed precipitation and temperature during the period of 1950–2000. To test the predictability, the available historical data for GCMs was divided into training set (1950–85) and testing set (1986–2000). The projected meteorological variables rely on historical observations and thereby provide information for simulating the PHDI time series.

The chosen study area was climate division 1 located in Texas. The observed monthly temperature, precipitation and PHDI time series were collected from the NOAA/National Climatic Data Center. To test the predictability of PHDI time series, the climate projections for different models for A2 scenarios were obtained from phase 3 of the World Climate Research Programme (WCRP) Coupled Model Intercomparison Project (CMIP3) multimodel dataset. The details can be found in the National Laboratory (LLNL)–Reclamation–Santa Clara University (SCU) multimodel dataset, stored and served out of the LLNL Green Data Oasis (available online at http://gdo-dcp.uccllnl.org/downscaled_cmip3_projections/dcpInterface.html). Each WCRP CMIP3 climate projection was bias corrected and spatially downscaled (Wood et al. 2004; Maurer 2007) using a two-step procedure: (i) bias correction and (ii) spatial scale downscaling. The statistical properties of annual precipitation time series obtained from multiple GCMs are shown in Table 3 during the period of 1950–2000. The mean annual precipitation during the second half of the century did not vary much among GCMs and the values were lower with respect to the observed time series located in climatic division 1. Both higher and lower values of coefficient of variation, kurtosis, and skewness were observed in GCMs precipitation output than in observed precipitation.
Before applying the wavelet–Bayesian regression model the correlation coefficient was plotted between observed precipitation and temperature monthly time series obtained from the National Climatic Data Center (NCDC) website and those obtained from the GCM output for the period 1950–2000 as shown in Figure 9. It is worth noting that CC between observed and different GCM monthly temperatures was found to be consistent and had a value of 0.96, whereas based on the precipitation the CC values fluctuated between 0.35 and 0.5. This preliminary analysis demonstrated that the GCMs were capable of reproducing temperature well, whereas in the case of precipitation they were weak.

Since precipitation plays an important role in characterizing droughts, including PHDI, it will be interesting to look at their predictability using precipitation and temperature obtained from the GCMs output. The PHDI time series was simulated.

**Figure 8.** Comparison between (a),(c) observed and (b),(d) simulated annual drought durations on temporal scale for different spatial locations.
using the selected GCMs for training (1950–80) and testing (1981–2000), as shown in Figure 10. It can be observed from the figure that a pattern is noticed during the training period; however, during the testing period the pattern is missing. This can be quantified using the CC plot during the training and testing periods in simulating the PHDI time series (Figure 11). The performance of GCMs for simulating the PHDI time series based on the CC values was found to be 0.4–0.7 during the training period. However, during the testing period the performance was found to be poor. The GCMs were able to reproduce temperature more accurately than precipitation. Since simulating precipitation is a complex phenomenon dealing with moisture in the atmosphere and the related mechanism to condense it, the GCMs are not able to perform a better job.

Table 3. Different GCMs* and the statistical properties of their annual precipitation located in climatic division 1 of Texas for A2 scenarios during the period of 1950–2000.

<table>
<thead>
<tr>
<th>Serial No.</th>
<th>List of GCMs</th>
<th>Mean (cm)</th>
<th>Coef of variation (%)</th>
<th>Kurtosis</th>
<th>Skewness</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Climatic division 1 (actual data)</td>
<td>47.63</td>
<td>19.03</td>
<td>2.94</td>
<td>-0.22</td>
</tr>
<tr>
<td>2</td>
<td>bccr_bcm2_0.1.sresb1</td>
<td>44.96</td>
<td>20.39</td>
<td>2.33</td>
<td>-0.06</td>
</tr>
<tr>
<td>3</td>
<td>cccma_cgcm3_1.1.sresb1</td>
<td>45.03</td>
<td>19.11</td>
<td>2.14</td>
<td>0.11</td>
</tr>
<tr>
<td>4</td>
<td>cccma_cgcm3_1.2.sresb1</td>
<td>44.86</td>
<td>20.84</td>
<td>2.55</td>
<td>0.13</td>
</tr>
<tr>
<td>5</td>
<td>cccma_cgcm3_1.3.sresb1</td>
<td>45.01</td>
<td>22.67</td>
<td>2.94</td>
<td>0.42</td>
</tr>
<tr>
<td>6</td>
<td>cccma_cgcm3_1.4.sresb1</td>
<td>45.16</td>
<td>21.83</td>
<td>2.42</td>
<td>0.00</td>
</tr>
<tr>
<td>7</td>
<td>cccma_cgcm3_1.5.sresb1</td>
<td>45.21</td>
<td>16.88</td>
<td>2.60</td>
<td>-0.07</td>
</tr>
<tr>
<td>8</td>
<td>cnrm_cm3.1.sresb1</td>
<td>45.01</td>
<td>19.91</td>
<td>2.91</td>
<td>-0.58</td>
</tr>
<tr>
<td>9</td>
<td>csiro_mk3_0.1.sresb1</td>
<td>45.34</td>
<td>25.32</td>
<td>3.41</td>
<td>0.46</td>
</tr>
<tr>
<td>10</td>
<td>gfdl_cm2_0.1.sresb1</td>
<td>44.91</td>
<td>28.47</td>
<td>2.27</td>
<td>0.07</td>
</tr>
<tr>
<td>11</td>
<td>gfdl_cm2_1.1.sresb1</td>
<td>45.39</td>
<td>22.85</td>
<td>2.86</td>
<td>-0.54</td>
</tr>
<tr>
<td>12</td>
<td>giss_model_e_r.1.sresb1</td>
<td>45.21</td>
<td>15.38</td>
<td>2.11</td>
<td>0.34</td>
</tr>
<tr>
<td>13</td>
<td>inmcm3_0.1.sresb1</td>
<td>44.98</td>
<td>20.12</td>
<td>3.44</td>
<td>0.70</td>
</tr>
<tr>
<td>14</td>
<td>ipsl_cm4.1.sresb1</td>
<td>44.88</td>
<td>19.14</td>
<td>2.17</td>
<td>-0.11</td>
</tr>
<tr>
<td>15</td>
<td>miroc3_2_medres.1.sresb1</td>
<td>45.31</td>
<td>18.42</td>
<td>2.30</td>
<td>0.02</td>
</tr>
<tr>
<td>16</td>
<td>miroc3_2_medres.2.sresb1</td>
<td>45.11</td>
<td>20.26</td>
<td>2.33</td>
<td>-0.34</td>
</tr>
<tr>
<td>17</td>
<td>miroc3_2_medres.3.sresb1</td>
<td>45.14</td>
<td>18.46</td>
<td>2.96</td>
<td>0.33</td>
</tr>
<tr>
<td>18</td>
<td>miubi_echo_g.1.sresb1</td>
<td>45.06</td>
<td>17.72</td>
<td>2.93</td>
<td>-0.49</td>
</tr>
<tr>
<td>19</td>
<td>miubi_echo_g.2.sresb1</td>
<td>44.96</td>
<td>16.82</td>
<td>5.00</td>
<td>1.37</td>
</tr>
<tr>
<td>20</td>
<td>miubi_echo_g.3.sresb1</td>
<td>45.24</td>
<td>16.52</td>
<td>2.92</td>
<td>0.13</td>
</tr>
<tr>
<td>21</td>
<td>mpi_echam5.1.sresb1</td>
<td>45.16</td>
<td>24.44</td>
<td>3.45</td>
<td>0.52</td>
</tr>
<tr>
<td>22</td>
<td>mpi_echam5.2.sresb1</td>
<td>44.83</td>
<td>25.29</td>
<td>2.24</td>
<td>0.27</td>
</tr>
<tr>
<td>23</td>
<td>mpi_echam5.3.sresb1</td>
<td>44.98</td>
<td>21.88</td>
<td>2.45</td>
<td>-0.13</td>
</tr>
<tr>
<td>24</td>
<td>mri_cgm2_3_2a.1.sresb1</td>
<td>45.19</td>
<td>18.54</td>
<td>2.90</td>
<td>0.15</td>
</tr>
<tr>
<td>25</td>
<td>mri_cgm2_3_2a.2.sresb1</td>
<td>44.91</td>
<td>18.89</td>
<td>3.76</td>
<td>0.84</td>
</tr>
<tr>
<td>26</td>
<td>mri_cgm2_3_2a.3.sresb1</td>
<td>45.09</td>
<td>24.23</td>
<td>2.53</td>
<td>0.11</td>
</tr>
<tr>
<td>27</td>
<td>mri_cgm2_3_2a.4.sresb1</td>
<td>45.11</td>
<td>18.88</td>
<td>3.04</td>
<td>0.62</td>
</tr>
<tr>
<td>28</td>
<td>mri_cgm2_3_2a.5.sresb1</td>
<td>44.88</td>
<td>20.17</td>
<td>2.30</td>
<td>0.31</td>
</tr>
<tr>
<td>29</td>
<td>ncar_ccsm3.0.1.sresb1</td>
<td>45.19</td>
<td>19.15</td>
<td>2.05</td>
<td>0.07</td>
</tr>
<tr>
<td>30</td>
<td>ncar_ccsm3.0.2.sresb1</td>
<td>45.06</td>
<td>20.80</td>
<td>2.55</td>
<td>0.55</td>
</tr>
<tr>
<td>31</td>
<td>ncar_ccsm3.0.3.sresb1</td>
<td>45.39</td>
<td>21.53</td>
<td>2.23</td>
<td>0.16</td>
</tr>
<tr>
<td>32</td>
<td>ncar_ccsm3.0.4.sresb1</td>
<td>45.21</td>
<td>22.28</td>
<td>1.99</td>
<td>0.06</td>
</tr>
<tr>
<td>33</td>
<td>ncar_ccsm3.0.7.sresb1</td>
<td>45.21</td>
<td>22.88</td>
<td>3.19</td>
<td>0.14</td>
</tr>
<tr>
<td>34</td>
<td>ncar_pcm1.2.sresb1</td>
<td>45.01</td>
<td>21.53</td>
<td>2.26</td>
<td>0.29</td>
</tr>
<tr>
<td>35</td>
<td>ncar_pcm1.3.sresb1</td>
<td>45.09</td>
<td>21.79</td>
<td>2.47</td>
<td>0.00</td>
</tr>
<tr>
<td>36</td>
<td>ukmo_hadcm3.1.sresb1</td>
<td>44.78</td>
<td>22.03</td>
<td>2.76</td>
<td>0.18</td>
</tr>
</tbody>
</table>

* Expansions of the GCM names can be found on the Internet.
Therefore, the lack in simulation of precipitation affects the performance of PHDI time series.

5. Conclusions

The following conclusions are drawn from this study:

1) Common drought information is likely to be observed between neighboring climatic divisions; however, with the increase in spatial units to a state or regional scale, the quality of information sharing reduces. The number of droughts at a regional unit is found to be at a maximum in the regions located in northeast, southeast, and northwest regions of the United States. The minimum number of droughts occurred in the west north-central region of the United States, which happens to be the lowest among all the spatial selected regions.

2) The wavelet–Bayesian regression model performs better than the Bayesian regression model based on different goodness-of-fit results. This advantage results because of the decomposition of PHDI time series into high- and low-frequency individual time series to capture better information.

3) Using the wavelet–Bayesian regression model, the performance of precipitation and temperature for simulating the PHDI time series at different spatial units varies, as judged by the goodness-of-fit tests. The performance during the testing period is better at regional units (the northeast, northwest, east north-central, southeast, and west regions) based on the correlation coefficient, and a clear pattern is not observed using the root-mean-square error measure. The bias in the simulation
time series is observed to be negative during the training period and positive during the testing period at a large number of spatial units. Higher persistence is observed in the simulated PHDI time series than in the observed time series.
4) The performance to evaluate annual drought characteristics (severity and duration) decreases in comparison to simulating only the PHDI time series. Based on the threshold in selecting drought properties, the performance is better at the zero threshold level than at the $-2$ threshold level. A higher bias is observed in simulating annual drought severity except for the spatial units in the northeast, east north-central, southeast, and northwest regions. The maximum positive bias for annual drought severity is observed in the southwest and west regions as the simulated values are lower than the observed values.

5) By changing the threshold levels in identifying the annual drought duration it affects the bias. Using the zero threshold, the positive bias is noted in many spatial units, whereas using the $-2$ threshold negative bias is observed in most of the spatial units.

6) The output from GCMs is able to capture temperature well, and it is not observed in the case of precipitation. Therefore, using these GCMs output in simulating the PHDI time series does not perform well.

Acknowledgments. The authors wish to thank the editor, associate editor, and anonymous reviewers for their constructive suggestions to improve the quality of our manuscript. This work was financially supported in part by the United States Geological Survey (USGS Project ID 2009TX334G) through the project “Hydrological Drought Characterization for Texas under Climate Change, with Implications for Water Resources Planning and Management.”

References


