Investigating the variability of GCMs’ simulations using time series analysis

Babak Zolghadr-Asli, Omid Bozorg-Haddad, Parisa Sarzaeim and Xuefeng Chu

ABSTRACT

The natural vulnerability to the climate change phenomenon due to the unique topographic and climatic conditions in the Middle East adds significance to an already important issue of evaluating the simulations of general circulation models (GCMs) in this region. To this end, this study employed time series analysis to evaluate GCMs’ simulations, in terms of the air temperature variable, with regard to the observed climatic behaviors of Karkheh River basin, Iran. Resultantly, each of the GCMs’ time series was broken down into three principal components (i.e., periodicity, trend, and stochastic component), and each component was analyzed accordingly. Results demonstrated that the simulations from different models significantly differed. Even though some models like CSIR-MK3.5 and INGV-SXG outperformed others in representing an accurate estimation of the historical climatic behavior of the southern plains of the Karkheh River, the GCMs could not provide a realistic simulation of the historical climatic behavior for the topographically challenging areas, like the northern mountainous parts of the basin. It should be noted that while the results from such analyses would shed light on the variability of the GCMs’ simulations in regional-scale studies, the results, under no circumstances, provide evidence indicating that one model is more accurate than another.

KEYWORDS | autoregressive moving average (ARMA) model, climate change, deseasonalizing, detrending, time series analysis, water resources management

INTRODUCTION

Model development is perhaps the core principle of predicting future hydrological events (Koutsoyiannis et al. 2008). Undoubtedly, the accuracy of the predicted results is related to the relative ability and credibility of the models. The general circulation models (GCMs) are sophisticated tools designed to simulate Earth's climate system, by capturing the complex interactions of its components (Reichler & Kim 2008). Subsequently, the state-of-the-art GCMs, which are mostly operated across at least nine orders of magnitude in both space and time (Blöschl & Sivapalan 1995), are the best available tools for future climatic simulations (Suppiah et al. 2007). Such attributes made GCMs an ideal option to assess the global consequences of different emission pathways, which are believed to be the main cause behind the plausible changes in Earth’s climate (Schiermeier 2007; Zolghadr-Asli 2017).

While the early attempts of developing GCMs in the early 1980s could be loosely described as simplistic representations of Earth’s climatic behavior, the performances of GCMs have been significantly enhanced compared to prior generations due to the substantial improvements in the physical basis of climate modeling, comprehensive calibration process supported by reliable observation datasets, and advancements in computational capabilities of supercomputers (Reichler &
Consequently, obtaining the geographically distributed predictions of the future climate, which mostly are core-gridded estimations of the global behavior of climate from GCMs for different climate change scenarios, is a common practice in climate change studies. While the projections of GCMs often portray a similar condition due to a similar initial scenario (Knutti 2008), GCMs’ trajectories can vary from one another, mostly in regional-scale predictions (Shukla et al. 2006).

GCMs’ simulations are primarily driven by assuming a set of external forcing, including the observed or estimated natural and anthropogenic sources, such as variations in solar radiations, volcanic activities, and emission patterns of greenhouse gases. What mainly makes the GCMs’ projections different from one another is the formulation of these external forces, and the interactions between model components (Reichler & Kim 2008; Wilby 2010). Even though a review over the projections from earlier generations of GCMs could reveal the relative comparability with the current state of the GCMs’ predictions on the global scale, given a particular scenario associated with the behavioral pattern of climate drivers (Simmons & Hollingsworth 2002), this does not indicate that the regional-scale predictions would not differ from one model to another (Murphy et al. 2004; Shukla et al. 2006). Consequently, it is plausible that some GCMs are able to provide reasonable projections in regional studies.

Despite the core gridding systems used in the GCMs’ simulations, and the imperfect theoretical assumptions regarding the climatic behaviors in modern GCMs (Reichler & Kim 2008; Ashofteh et al. 2016), the overall performances of GCMs have been considerably enhanced over time (Perkins et al. 2007). A review of the existing work reveals that, while many researchers support the acceptability of GCMs’ projections (e.g., Kalnay et al. 1996; Watterson 1996; Murphy et al. 2004; Min & Hense 2006; Shukla et al. 2006; Perkins et al. 2007; Knutti 2008; Reichler & Kim 2008; Chiew et al. 2009; Ashofteh et al. 2016), some researchers believe that GCMs are yet to be improved to be considered as tools for predicting the future climate conditions (e.g., Koutsoyiannis et al. 2008). In spite of such skepticism toward the performance of GCMs, they seem to be the only viable options for studying the global reactions to the internal and external climatic drivers, which in turn is one of the main stages of climate change studies. Bearing in mind the crucial and indispensable role of GCMs, nowadays, a wide variety of modern, state-of-the-art GCMs are provided by different institutions and organizations all around the world, which in turn, raise some questions: Given an initial condition, do the simulation results of these models differ significantly from one another? If so, in which ways can one expect to see such variations?

Considering that GCMs are, in essence, a set of models that can provide the simulated behaviors of the climate, a comparison-oriented approach that correlates the GCMs’ outcome to a reference condition can be the framework to pursue the answers to the aforementioned questions. Naturally, this reference condition cannot be set in future datasets for they are yet to happen. Therefore, comparing the GCMs’ simulations against the historical data can be considered as a viable approach to test the aforementioned notion regarding the GCMs’ simulated results in regional-scale studies. Such evaluation and inter-comparison of the GCMs’ performances would shed light on the overall capability of the climate system modeling GCMs (Min & Hense 2006). Note that this framework, however, is not without challenges. Mainly, to test the GCMs’ simulation results, reliable and consistently observed climate datasets are required for a long time period. Unfortunately, however, well-equipped climate monitoring stations are not uniformly distributed throughout the world. A crucial point that should be stated at this stage is that while these evaluations can provide vital information about the profile variability in GCMs’ simulation results, such comparison results do not necessarily provide evidence as to one model being more accurate than another, for, in fact, the primary purpose of GCMs is not to reproduce historical climate conditions, but rather to provide a plausible range of future conditions under various climate change scenarios (Reichler & Kim 2008).

Assessments of GCMs’ simulations can be categorized into two major groups: (1) index-based method and (2) non-index-based method. While the index-oriented studies employ and, in some cases, alter different indexes to evaluate and rank the simulation accuracies of GCMs (e.g., Watterson 1996; Murphy et al. 2004; Reichler & Kim 2008), the non-index-based studies evaluate the climatic
and, in some cases, the hydrologic datasets directly (e.g., Kalnay et al. 1996; Min & Hense 2006; Shukla et al. 2006; Perkins et al. 2007; Knutti 2008; Koutsoyiannis et al. 2008; Chiew et al. 2009; Ashofteh et al. 2016). Even though each method has certain merits and drawbacks, the objective of this study is to evaluate GCMs’ simulations using the original climatic modeling results as the basis for analysis, rather than indices.

While evaluating the regional-scale GCMs’ simulations is a necessity, conducting such assessments in water-stressed regions like the Middle East should be considered as one of the managerial priorities for achieving water security. The reason behind such a statement is two-fold. (1) On the one hand, the region with arid to semi-arid climate is already faced with mild to severe water stress (Evans 2009). Despite the climate variability of the region (Krichak et al. 2000; Barth & Steinkohl 2004; Evans et al. 2004; Issar & Zohar 2004; Öñol & Semazzi 2009; Evans 2010), securing both quality and quantity of water resources is considered as a major challenge throughout the region (Evans 2009). Subsequently, such conditions highlight the importance of studying possible influencing factors of water resources (e.g., climate change). On the other hand, numerous studies have demonstrated the potential impacts of climate change in the region. Accordingly, up until now, the climate variability could mask the potential impacts of anthropogenic forces that caused climate changes in the region (Mann 2002). The main concern, however, is that the climate change impacts eventually could amplify the water crisis in the region (Gibelin & Déqué 2003; Kitoh et al. 2008; Giannakopoulos et al. 2009; Black et al. 2010; Evans 2010; Lelieveld et al. 2012). (2) Due to the high interannual climate variability and the unique topography of the region (mountains and inland seas), simulating the regional climatic behavior is considered as a challenge for GCMs (Evans et al. 2004). Consequently, this further emphasizes the importance of investigating the impact of GCM selection on climate change studies in the region. Despite numerous efforts that aimed to investigate the performances of GCMs for different locations around the world (e.g., Australia (Perkins et al. 2007; Chiew et al. 2009), North Pacific Ocean (Overland & Wang 2007), Brazil, Greece, Sudan, Japan, and the United States (Koutsoyiannis et al. 2008)), less attention has been paid to the evaluation of GCMs in the Middle East. Furthermore, few studies have been dedicated to inspecting the GCMs’ simulation in the Middle East region, which has limited the evaluations of the ‘by-products’ of GCMs (e.g., streamflow (Ashofteh et al. 2016)), rather than the original simulation results.

As an attempt to investigate the variability of GCMs’ simulations in the Karkheh River basin, Iran, this study implements a framework based on the time series analysis. Through this systematic analysis, the time series resulting from the GCMs’ simulations are decomposed into their elemental components (i.e., periodicity, trend, and stochastic component), which in turn, facilitates investigation for any signs of variability in the GCMs’ simulation results in this regional-scale study. Furthermore, the proposed study not only facilitates examining the differences in the regional-scale GCMs’ simulation results, but also sheds light on the unique characteristics of the GCMs’ predicted time series that reflect such variations. Note that, while in theory, every climatic variable could be evaluated through the proposed time series-oriented framework, this study emphasizes the investigation on the air temperature variable. The reason behind this is the compatibility of the fluctuations in the temperature variable on the specific case study and the times series analysis. In other words, given that the time series decomposition is the founding principle used in the framework proposed in this case study for characterizing the climatic variables, among the variables that are most likely to influence the water resources in such a water-stressed region, air temperature was the most suitable variable that can be first analyzed to demonstrate the variability in the components of the GCM-generated times series. Undoubtedly, to gain a comprehensive evaluation of the GCMs’ performance, one must conduct a similar procedure for other climatic variables, as well.

MATERIALS AND METHODS

Figure 1 shows the basic framework of the proposed methodology, which involves performing GCM simulations, deseasonalizing and detrending the simulated time series, and eventually regenerating the stochastic component of the time series using time series modeling. In order to compare different GCM models, the historical time series of the region was employed as a reference.
Components of time series

A time series refers to a sequence of data, which are ordered in time. Technically, time series are used to display a phenomenon throughout time. The key concept in time series modeling, which refers to the techniques that relate time series data as dependent variables to the predictors, is understanding the components of time series. A perfectly constructed time series model with well-defined components can be used either for forecasting the upcoming events, or generating synthetic datasets that cannot be obtained by direct measurement for a given situation (Araghinejad 2013).

To analyze a time series, one must account for its unique internal structure (such as autocorrelation, trend, or seasonal variation). A general time series \( (X_t) \) can be described as (Araghinejad 2013):

\[
X_t = P_t + T_t + S_t
\]

(1)

in which \( P_t \) = periodic component; \( T_t \) = linear trend component; and \( S_t \) = stochastic component, which is given by:

\[
S_t = D_t + \epsilon_t
\]

(2)

in which \( D_t \) = deterministic term of the stochastic component and \( \epsilon_t \) = random variable term of the stochastic component.

**Periodic component**

The periodicity or seasonality component can be described as a function, the output of which contains values that are repeated on a regular and constant base, called return period. The most common return periods in the field of water resources and environmental sciences are annual, seasonal, or monthly time intervals. Periodicity is a systematic and chronological component of a time series. The process of removing the periodic component of a time series, which is commonly referred to as ‘deseasonalizing,’ can be achieved by seasonal adjustment of the data. Rearranging the data on an annual, seasonal, or monthly basis can be used to deseasonalize the water resources and climate data series such as the air temperature.

**Trend component**

The trend component refers to any particular ascending or descending patterns in a time series. The linear trend can be expressed as:

\[
T_t = a \times t + b
\]

(3)

in which \( t \) = time as the independent variable; \( a = \text{slope} \); and \( b = \text{intercept} \). It should be noted that \( a \) and \( b \) are considered as the model parameters.

Detecting and in turn estimating the trend component in the time series of a climatic variable (e.g., air temperature) have attracted significant interest in the past decades, mostly due to the climate change phenomena. Subsequently, it is vital to have an accurate estimation of the trend component of a time series from GCMs. Statistical tests are often used to examine: (1) whether a time series illustrates any signs of a trend; and (2) if so, what is the slope (\( a \)) of the projected trend line. In this study, the Mann–Kendall test and Sen’s slope estimator were used to respectively address these two issues.

The rank-based, non-parametric Mann–Kendall test was proposed by Mann (1945) and extended by Kendall (1975).
It has been commonly used to assess the significance of trends in hydro-meteorological time series, including air temperature (Yue et al. 2002). This test is applied to check whether a time series contains an increasing or decreasing trend with regard to a specific level of significance (α). A detailed description about the application of this statistical test can be found in Partal & Kahya (2006) and Tabari & Maroﬁ (2011).

While the Mann–Kendall test is a classical technique to detect the presence of a trend component in a time series, it is unable to quantify the magnitude of this trend (i.e., the trend’s slope (α)). Sen’s slope estimator is, however, a simple non-parametric procedure to estimate the slope of the linear trend (Sen 1968). More information about this statistical test can be found in Partal & Kahya (2006).

Stochastic component

After deseasonalizing and eliminating the trend component, what remains is the stochastic component of a time series. The most common practice to evaluate this component is using time series modeling, which chiefly consists of three major steps: (1) model selection; (2) selection of the order of the model; and (3) determination of the model parameters. It is essential to select a model that is the most suitable to reflect variations of the time series.

In this study, the autoregressive moving average (ARMA) model was considered as an alternative for the time series modeling. Partially due to their mathematical structure, the ARMA models are capable of providing a convenient statistical representation of a time series in terms of a small number of parameters (Katz & Skaggs 1981). Consequently, the ARMA models are ideal alternatives to represent the stochastic component of the deseasonalized and detrended meteorological time series. The ARMA models have been used to reconstruct the stochastic fluctuations in GCMs’ simulation results (Katz 1982; Reed 1986; Wilson & Mitchell 1987; Karl et al. 1991; Bloomﬁeld 1992). Even though some other state-of-the-art time series models can, in some cases, outperform the ARMA models, ARMA is still considered as one of the top choices for synthetic data generation (Araghinejad 2013).

An ARMA model is, in fact, the expanded version of an autoregressive (AR) model, which is based upon a Markov chain. An ARMA(p,q) is a combination of AR(p) or an autoregressive model with the order of p (i.e., the current deterministic term is a function of p previous deterministic terms), and MA(q) or a moving average model with the order of q (i.e., the current random term is a function of q previous random terms), which can be mathematically expressed as:

\[ S_t = \sum_{i=1}^{p} (\varphi_i \times S_{t-i}) - \sum_{j=1}^{q} (\theta_j \times \epsilon_{t-j}) + \zeta_t \quad p, q \geq 0 \]  \hspace{1cm} (4)

in which \( \varphi_i \) is the ith parameter of the autoregressive term of the model; \( \theta_j \) is the jth parameter of the moving average term of the model; and \( \zeta_t \) is error of the model in the tth time step. In order to employ an ARMA model to explain the stochastic component of a time series, the order of the model \( (p \text{ and } q) \), and the model parameters \( (\varphi_i \text{ and } \theta_j) \) must be adjusted, accordingly. Further information regarding the calibration of ARMA models can be found from Salas et al. (1980) and Hipel & McLeod (1994).

Evaluation of the GCMs’ simulation

In order to evaluate the simulations of GCMs, their outcomes must be compared to the historical data. In this study, percent bias (PB) (Gupta et al. 1999) was used to quantify the GCMs’ performances. Such an approach not only enables the modelers to test the acceptability of GCMs in simulating historical time series but also allows relative comparisons among GCMs. This quantitative statistical parameter is defined as:

\[ PB = \left[ \frac{\sum_{t=1}^{N} (T_{obs}^t - T_{sim}^t)}{\sum_{t=1}^{N} T_{obs}^t} \right] \times 100 \]  \hspace{1cm} (5)

where \( T_{obs}^t \) and \( T_{sim}^t \) are observed and simulated air temperatures in the tth time step, respectively; and \( N \) is total number of time steps. For a monthly time step, a PB value between −25 and +25% can be considered satisfactory (Moriai et al. 2007; Adhikari et al. 2016).

Study area

The Karkheh basin, with an area of 51,000 km², which is located in the south-western region of Iran (latitude:
30–35° N; longitude: 46–49° E), is selected for this study. It has a historical background of water use and engineering and is now referred to as ‘Iran’s food basket’, mainly due to its vast under-wheat cultivation lands. Although the basin’s growing industry, booming hydropower, and oil fields’ products have caused a rapid, demographic shift from agriculture towards cities, yet, as a key agricultural region in Iran, there still exists pressure to keep up an increase in the agricultural production of the region (Ahmad & Giordano 2010).

While the basin is dominated by semi-arid to arid climate, the climate of the basin illustrates large variations in the average annual precipitation between the southern and northern regions. The southern parts of the basin receive an average annual rainfall of approximately 150 mm, while in the northern and northeastern parts of the basin, it can reach up to 1,000 mm (Muthuwatta et al. 2019). Additionally, the basin experiences high spatial and temporal variations in temperature, as well. Accordingly, the temperature in this area ranges from −25 to 50°C. While in the northern parts of the region, even in the hot months, the maximum temperature fluctuates from 23 to 36°C, in the southern parts of the basin and during the summer seasons, the temperature can exceed 40°C. In order to properly cover the temperature fluctuations of the entire basin, the basin was divided into three major sections, each of which was represented by a climate station. The information about these stations and their locations are summarized in Table 1. Additionally, the recorded average temperatures by the opted climate stations are shown in Figure 2.

To investigate the variations in the GCMs’ projections, the air temperature time series simulated by seven GCMs were obtained for each of the three selected climate stations, including (1) the second version of Bjerknes Centre for Climate Research (BCM V.2), (2) the coupled global climate model, version T47 (CGCM3T47), (3) the third version of Centre National de Recherches Météorologiques coupled model (CNRM-CM3), (4) Council for Scientific and Industrial Research model, version MK3.5 (CSIR-MK3.5), (5) Hamburg atmosphere–ocean coupled circulation model (ECHO-G), (6) the Goddard Institution for Space's automation object model (GIS-AOM), and (7) Istituto Nazionale di Geofisica e Vulcanologia (INGV-SXG).

Table 1 | Three climate stations in the region

<table>
<thead>
<tr>
<th>Station</th>
<th>Elevation (m.a.s.l.)</th>
<th>Latitude (N)</th>
<th>Longitude (E)</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>Polchehr</td>
<td>1,275</td>
<td>34°20'</td>
<td>47°26'</td>
<td>Northern</td>
</tr>
<tr>
<td>Chamgez</td>
<td>380</td>
<td>32°56'</td>
<td>47°35'</td>
<td>Central</td>
</tr>
<tr>
<td>Abdolkhan</td>
<td>40</td>
<td>31°50'</td>
<td>48°23'</td>
<td>Southern</td>
</tr>
</tbody>
</table>

m.a.s.l. = meters above sea level.

Figure 2 | Average annual temperatures at three selected climate stations.
RESULTS AND DISCUSSION

As stated earlier, GCMs play a crucial role in climate change studies. While the core principles in GCMs are more or less the same, this does not necessarily indicate that their regional-scale simulations of climatic behaviors would result in a similar projection, as well. While this alone can be a good reason to conduct a thorough evaluation on GCMs’ variability in terms of their performances, given the roles of these models in climate change studies, some regions like the Middle East should place priority on such studies due to their unique characteristics, including their vulnerability towards climate change and their diverse climate conditions. This section focuses on evaluation of the three components of the time series simulated by GCMs and the variability of their simulations.

Evaluation of the periodicity component

To make the GCMs’ simulation comparable, first, the gridded temperature time series were downscaled for each selected climate station in the basin. In order to downscale the model outputs to a finer spatial scale, the grid points nearest to each selected station were used to make inferences at a regional scale, from coarser climate model’s gridded trajectories (Georgakakos 2003; Koutsoyiannis et al. 2008).

As previously stated, most meteorological data contain a periodicity component. Rearranging the data in a monthly interval was used as a deseasonalizing technique. Furthermore, reconstructing the time series in a monthly interval revealed which GCM was potentially the best candidate to represent the historical climate behavior of the region. Consequently, to remove the periodicity of the temperature time series, they were broken down and rearranged in a monthly interval. As an example, Figure 3 shows the comparisons of the monthly average temperatures simulated by the GCMs against the observed temperatures in the southern parts of the region. Although such comparisons showed the relative acceptability of the GCMs in producing the historical time series of air temperature, to have a systematic and comprehensive assessment, PB was calculated to quantitatively evaluate

![Figure 3](https://iwaponline.com/jwcc/article-pdf/10/3/449/598425/jwc0100449.pdf)
the performances of GCMs by using the corresponding historical data. The *PB* values for the northern, central, and southern parts of the region are shown in Figures 4–6, respectively. The overall performances of GCMs for the central and southern parts of the region were better than that for the northern parts. Additionally, analyzing the *PBs* in Figure 4 revealed that, although none of the GCMs were able to accurately simulate the observed temperatures in the northern parts of the region in January, February, and December, the INGV-SXG and ECHO-G

<table>
<thead>
<tr>
<th></th>
<th>Jan</th>
<th>Feb</th>
<th>Mar</th>
<th>Apr</th>
<th>May</th>
<th>Jun</th>
<th>Jul</th>
<th>Aug</th>
<th>Sep</th>
<th>Oct</th>
<th>Nov</th>
<th>Dec</th>
</tr>
</thead>
<tbody>
<tr>
<td>BCM V2</td>
<td>215</td>
<td>225</td>
<td>109</td>
<td>45</td>
<td>25</td>
<td>12</td>
<td>10</td>
<td>12</td>
<td>19</td>
<td>32</td>
<td>69</td>
<td>133</td>
</tr>
<tr>
<td>CGCM3.147</td>
<td>204</td>
<td>137</td>
<td>71</td>
<td>43</td>
<td>25</td>
<td>4</td>
<td>-4</td>
<td>0</td>
<td>10</td>
<td>25</td>
<td>69</td>
<td>140</td>
</tr>
<tr>
<td>CNRM-CM3</td>
<td>240</td>
<td>159</td>
<td>46</td>
<td>25</td>
<td>4</td>
<td>-3</td>
<td>-2</td>
<td>4</td>
<td>8</td>
<td>29</td>
<td>72</td>
<td>159</td>
</tr>
<tr>
<td>CSIR-MK3.5</td>
<td>4</td>
<td>-37</td>
<td>-41</td>
<td>-32</td>
<td>-23</td>
<td>-30</td>
<td>-24</td>
<td>-19</td>
<td>-16</td>
<td>-11</td>
<td>18</td>
<td>34</td>
</tr>
<tr>
<td>ECHO-G</td>
<td>108</td>
<td>64</td>
<td>28</td>
<td>12</td>
<td>4</td>
<td>-2</td>
<td>-1</td>
<td>-1</td>
<td>-2</td>
<td>3</td>
<td>22</td>
<td>57</td>
</tr>
<tr>
<td>INGV-SXG</td>
<td>114</td>
<td>59</td>
<td>5</td>
<td>-12</td>
<td>-21</td>
<td>-21</td>
<td>-15</td>
<td>-15</td>
<td>-14</td>
<td>-7</td>
<td>15</td>
<td>74</td>
</tr>
</tbody>
</table>

Figure 4 | *PB* (percent bias) values for the deseasonalized time series in the northern parts of the region.

<table>
<thead>
<tr>
<th></th>
<th>Jan</th>
<th>Feb</th>
<th>Mar</th>
<th>Apr</th>
<th>May</th>
<th>Jun</th>
<th>Jul</th>
<th>Aug</th>
<th>Sep</th>
<th>Oct</th>
<th>Nov</th>
<th>Dec</th>
</tr>
</thead>
<tbody>
<tr>
<td>BCM V2</td>
<td>68</td>
<td>58</td>
<td>25</td>
<td>15</td>
<td>16</td>
<td>13</td>
<td>12</td>
<td>14</td>
<td>19</td>
<td>23</td>
<td>37</td>
<td>57</td>
</tr>
<tr>
<td>CGCM3.147</td>
<td>88</td>
<td>62</td>
<td>34</td>
<td>28</td>
<td>23</td>
<td>13</td>
<td>14</td>
<td>17</td>
<td>22</td>
<td>25</td>
<td>47</td>
<td>80</td>
</tr>
<tr>
<td>CNRM-CM3</td>
<td>77</td>
<td>44</td>
<td>11</td>
<td>11</td>
<td>7</td>
<td>4</td>
<td>6</td>
<td>9</td>
<td>11</td>
<td>21</td>
<td>37</td>
<td>66</td>
</tr>
<tr>
<td>CSIR-MK3.5</td>
<td>18</td>
<td>1</td>
<td>-13</td>
<td>-5</td>
<td>3</td>
<td>-1</td>
<td>2</td>
<td>4</td>
<td>7</td>
<td>10</td>
<td>21</td>
<td>30</td>
</tr>
<tr>
<td>ECHO-G</td>
<td>53</td>
<td>38</td>
<td>21</td>
<td>20</td>
<td>21</td>
<td>19</td>
<td>20</td>
<td>22</td>
<td>23</td>
<td>23</td>
<td>29</td>
<td>41</td>
</tr>
<tr>
<td>GIS-AOM</td>
<td>-46</td>
<td>-53</td>
<td>-43</td>
<td>-18</td>
<td>-7</td>
<td>-6</td>
<td>-4</td>
<td>-5</td>
<td>-7</td>
<td>-7</td>
<td>-13</td>
<td>-29</td>
</tr>
<tr>
<td>INGV-SXG</td>
<td>41</td>
<td>28</td>
<td>7</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>7</td>
<td>6</td>
<td>7</td>
<td>10</td>
<td>16</td>
<td>37</td>
</tr>
</tbody>
</table>

Figure 5 | *PB* (percent bias) values for the deseasonalized time series in the central parts of the region.
models did stand out, by accurately estimating 9 and 8 months, respectively. However, analyzing the PBs in Figure 5 indicated that GCMs had relatively improved performances in the central parts of the region. CSIR-MK3.5 was able to accurately simulate 11 months (all except December). INGV-SXG with accurate simulations for nine months (all except January, February, and December) was the second best-performing GCM. Finally, the PBs in Figure 6 indicated that two GCMs, CSIR-MK3.5 and INGV-SXG, accurately simulated historical temperatures for the southern parts of the region in all months. It is worth mentioning that the GCMs’ inaccurate simulations of the historical temperatures for the northern parts of the region, and perhaps even relative superiority of the simulations for the southern parts over those in the northern and central parts, can be attributed to the topography of the region. The northern parts of the region are highly dense mountainsides, while the southern parts are relatively flat plains. Hilly topography is a natural obstacle for GCMs with coarsely gridded networks. An alternative to overcome this complication is perhaps using certain dynamic downscaling techniques, that is, utilizing a GCM that is recalibrated for a specific regional scale and a finer gridded network (Zolghadr-Asli 2017; Zolghadr-Asli et al. 2018).

### Evaluation of the trend component

Although analyzing the deseasonalized time series could shed light on the capabilities of a GCM, in order to achieve an acceptable simulation of the historical data, all three components of the time series must be accurately represented in the GCM. The trend component, specifically, is of major interest in climate change studies. Although an accurate estimation of the trend of a time series does not necessarily guarantee the overall performance of a GCM, the trend component is, indeed, the bedrock of the time series. In long-term climate studies, inaccurate estimation of this component could be problematic. In this study, the Mann–Kendall test and Sen’s slope estimator were used to detect and quantify the slope of the linear trend in the deseasonalized time series ($\alpha = 5\%$). Figures 7–9 show the results for the northern, central, and southern parts of the region, respectively. The CGCM3T47 and ECHO-G models accurately estimated the trend component of the historical time series for the northern parts, while the BCM V2, GIS-AOM, and INGV-SXG models did a better job in estimating the trend component of the aforementioned time series for the central parts. Finally, the BCM V2, CNRM-CM3, and GIS-AOM models showed superiority in detecting and
estimating the trend components in the southern parts of the region. However, not surprisingly, the GCMs for the southern parts of the region were able to represent the trend component of the historical time series more accurately, compared to the central and northern parts of the region. Perhaps these accurate estimations were able to pave the way for these GCMs’ superior overall performance in the southern parts.
Evaluation of the stochastic component

As the last step, the stochastic component of the time series from GCMs was analyzed. The deseasonalized and detrended time series were used to construct an ARMA \((p,q)\) model as a representative of the stochastic component of the time series. As an example, the ARMA \((p,q)\) model parameters for the southern parts of the region in January are summarized in Table 2. To ensure the fair representativeness of the probabilistic term of the stochastic component for each modeled time series, 20 ensembles were generated and the average of these ensembles was analyzed using the \(PB\). Figures 10–12 show the \(PB\) values of the stochastic component for the northern, central, and southern parts of the region, respectively. As shown in Figure 10, although no model was able to accurately simulate the stochastic component of the historical temperature time series in the southern parts of the region, the ECHO-G and INGV-SXG models with acceptable simulations for eight and seven months, respectively, outperformed the other GCMs. Figure 11, on the other hand, illustrates that although none of the tested GCMs were able to accurately simulate the stochastic component of the historical temperature time series in the central parts of the region in all months, CSIR-MK3.5 and INGV-SXG were able to correctly simulate the stochastic component of the time series for 11 months (all except November) and 10 months (all except January and December), respectively. And finally, Figure 12

Table 2 | Time series modeling results for the southern parts of the region in January

<table>
<thead>
<tr>
<th>Model parameters</th>
<th>Observation ARMA (1,1)</th>
<th>BCM V2 ARMA (3,1)</th>
<th>CGCM3T47 ARMA (3,2)</th>
<th>CNRM-CM3 ARMA (1,1)</th>
<th>CSIR-MK3.5 ARMA (1,2)</th>
<th>ECHO-G ARMA (2,2)</th>
<th>GIS-AOM ARMA (2,2)</th>
<th>INGV-SXG ARMA (1,1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\phi_1)</td>
<td>1.00</td>
<td>-0.31</td>
<td>-0.55</td>
<td>1.01</td>
<td>1.02</td>
<td>0.19</td>
<td>0.38</td>
<td>1.00</td>
</tr>
<tr>
<td>(\phi_2)</td>
<td>-</td>
<td>0.87</td>
<td>0.72</td>
<td>-</td>
<td>-</td>
<td>0.84</td>
<td>0.63</td>
<td>-</td>
</tr>
<tr>
<td>(\phi_3)</td>
<td>-</td>
<td>0.33</td>
<td>0.98</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>(\theta_1)</td>
<td>-1.00</td>
<td>1.00</td>
<td>0.59</td>
<td>-0.93</td>
<td>-0.52</td>
<td>-0.02</td>
<td>-0.28</td>
<td>-1.00</td>
</tr>
<tr>
<td>(\theta_2)</td>
<td>-</td>
<td>-</td>
<td>-0.94</td>
<td>-</td>
<td>-0.48</td>
<td>-0.90</td>
<td>-0.72</td>
<td>-</td>
</tr>
</tbody>
</table>
demonstrates that, unlike the northern and central parts of the region, GCMs were able to correctly simulate the observed behavior of the stochastic components of the temperature time series in the southern parts. CSIR-MK3.5 and INGV-SXG with acceptable simulations of the stochastic component for 12 and 11 months (all except January) stand out as the superior models. Concerning the simulations of the historical time series of air temperatures, the overall performances of GCMs in the southern and central parts were better than those in the northern parts. Lastly, one should bear in mind that this study mainly aimed to investigate the variability in the simulations of different

Figure 10 | PB (percent bias) values for the stochastic components in the northern parts of the region.

Figure 11 | PB (percent bias) values for the stochastic components in the central parts of the region.
GCMs by comparing against the historical time series of air temperatures. In addition, while conducting similar simulations and analyses for other climatic variables are required to make a conclusive statement about the overall performance of these models, one should also note that producing an accurate simulation with regard to the historical data, does not necessarily guarantee a satisfactory performance of a GCM when climate change projections are concerned.

**CONCLUDING REMARKS**

GCMs, as tools for long-term meteorological modeling, paved the way for most climate change studies. Nowadays, using GCMs has become a common practice in such studies. Even though there has been a significant improvement in GCMs' overall performance and simulation capabilities, their performances in regional-scale studies might not necessarily converge to a solitary simulated outcome. Although the aforementioned notions could justify a comprehensive study on the variability in GCMs' simulations in each regional-scale study, these assessments for regions like the Middle East should have priority over others, due to their verified vulnerability to climate change phenomena, and their unique and diverse climatic conditions. Note that, however, such comparison results, under no circumstances, provide evidence indicating that one model is more accurate than another. This study aimed to evaluate the performance of a few available GCMs in the Karkheh River basin, using the non-index-based time series analysis. To do so, each GCM-simulated temperature time series was broken down into three essential components, and each component was then compared to its counterpart of the observed temperature time series. The results demonstrated that the simulation results from the GCMs significantly differed from one another. The CSIR-MK3.5 and INGV-SXG models outperformed others in representing an accurate estimation of the climatic behavior of the southern plains of the Karkheh River. However, for the topographically challenging northern parts of the Karkheh River basin with numerous mountains, the GCMs failed to provide a realistic simulation of the climatic behavior. Even though the improvement of GCMs reduced the inaccurate simulations for most months, there were still some time periods, mostly in cold seasons, in which the accuracy of the GCMs' simulations of temperature could decline dramatically. Further in-depth studies that involve monitoring and analyzing a wider range of climatic variables are required to reveal

---

**Figure 12** | PB (percent bias) values for the stochastic components in the southern parts of the region.
the overall performance of GCMs in such regional-scale studies.

REFERENCES


Knutti, R. 2008 Why are climate models reproducing the observed global surface warming so well? Geophysical Research Letters 35 (18), L18704.


Min, S. K. & Hense, A. 2006 A Bayesian approach to climate model evaluation and multi-model averaging with an application to global mean surface temperatures from IPCC.


First received 22 May 2017; accepted in revised form 31 March 2018. Available online 23 April 2018