CMIP5 Models’ Ability to Capture Observed Trends under the Influence of Shifts and Persistence: An In-Depth Study on the Colorado River Basin

KAZI ALI TAMADDUN
Department of Civil and Environmental Engineering and Construction, University of Nevada, Las Vegas, Nevada

AJAY KALRA
Department of Civil and Environmental Engineering, Southern Illinois University, Carbondale, Illinois

SANJIV KUMAR
School of Forestry and Wildlife Sciences, Auburn University, Auburn, Alabama

SAJJAD AHMAD
Department of Civil and Environmental Engineering and Construction, University of Nevada, Las Vegas, Nevada

(Manuscript received 20 September 2018, in final form 19 May 2019)

ABSTRACT
This study evaluated the ability of phase 5 of the Coupled Model Intercomparison Project (CMIP5) to capture observed trends under the influence of shifts and persistence in their data distributions. A total of 41 temperature and 25 precipitation CMIP5 simulation models across 22 grid cells (2.5° × 2.5° squares) within the Colorado River basin were analyzed and compared with the Climate Research Unit Time Series (CRU-TS) observed datasets over a study period of 104 years (from 1901 to 2004). Both the modeled simulations and observations were tested for shifts, and the time series before and after the shifts were analyzed separately for trend detection and quantification. Effects of several types of persistence were accounted for prior to both the trend and shift detection tests. The mean significant shift points (SPs) of the CMIP5 temperature models across the grid cells were found to be within a narrower range (between 1957 and 1968) relative to the CRU-TS observed SPs (between 1924 and 1985). Precipitation time series, especially the CRU-TS dataset, had a lack of significant SPs, which led to an inconsistency between the models and observations since the number of grid cells with a significant SP was not comparable. The CMIP5 temperature trends, under the influence of shifts and persistence, were able to match the observed trends very satisfactorily (within the same order of magnitude and consistent direction). Unlike the temperature models, the CMIP5 precipitation models detected SPs that were earlier than the observed SPs found in the CRU-TS data. The direction (as well as the magnitude) of trends, before and after significant shifts, was found to be inconsistent between the modeled simulations and observed precipitation data. Shifts, based on their direction, were found either to strengthen or to neutralize the preexisting trends in both the model simulations and the observations. The results also suggest that the temperature and precipitation data distributions were sensitive to different types of persistence—such sensitivity was found to be consistent between the modeled and observed datasets. The study detected certain biases in the CMIP5 models in detecting the SPs (tendency of detecting shifts earlier for precipitation and later for temperature than the observed shifts) and also in quantifying the trends (overestimating the trend slopes)—such insights may be helpful in evaluating the efficacy of the simulation models in capturing observed trends under uncertainties and natural variabilities.

Supplemental information related to this paper is available at the Journals Online website: https://doi.org/10.1175/JAMC-D-18-0251.s1.

Corresponding author: Sajjad Ahmad, sajjad.ahmad@unlv.edu

DOI: 10.1175/JAMC-D-18-0251.1
© 2019 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the AMS Copyright Policy (www.ametsoc.org/PUBSReuseLicenses).
1. Introduction

Understanding of trends in various hydroclimatological data, under the influence of climate change, to yield practical solutions to the existing and imminent threats is a major topic of research in the scientific community, especially among climate data modelers (Asrar et al. 2012; Arnell 2011). The Fifth Assessment Report (AR5) of the Intergovernmental Panel on Climate Change (IPCC) (IPCC 2014) provided a synthesis of results for policymakers based on simulation results provided by phase five of the Coupled Model Intercomparison Project (CMIP5) (Taylor et al. 2012). Many studies, on both the continental and regional scales, have focused on historical as well as projected temperature and precipitation trends in the multimodel CMIP5 data to examine extreme climate conditions (Wuebbles et al. 2014; Lynch et al. 2016). A study by Kumar et al. (2013) evaluated the effects of persistence, that is, clustering behavior in the hydroclimatological observations, in the long-term trends on a global scale and compared the results with multiple observed datasets. The study suggested that the CMIP5 models are capable of capturing the effects of persistence in temperature better than precipitation. Presence of persistence, for example, positive serial correlation, has been considered as one of the major reasons of uncertainty in hydroclimatological data since it reduces the effective sample size, which leads to an underestimation of variance, and overestimates the probability of a significant trend (Hamed and Rao 1998; Yue et al. 2002; Koutsoyiannis 2003). Another important feature of non-stationary hydroclimatic data is the presence of shifts (abrupt changes) in the data distributions (McCabe and Wolock 2002; Villarini et al. 2009; Tamaddun et al. 2019a,b). Villarini et al. (2009) suggest that a trend is “likely to continue in the future,” whereas a change point (shift point) implies “shift from one regime to another, and the status is likely to remain the same until a new regime shift occurs.” Ignoring the effect of shift points (SPs) in trend analyses may lead to ambiguous results (Villarini et al. 2009). Figure 1 shows an example of how ignoring an SP in the data distribution may lead to erroneous detection of trends (a similar example was shown by Villarini et al. (2009) to illustrate that performing a trend test before considering the SP may result in a detection of erroneous significant trend). However, if the presence of an SP was considered, the two subseries—before and after the SP—individually do not suggest statistically significant trends (note the horizontal slopes in Fig. 1). More details on the importance of detecting SPs in association with trends can be found in the works of Villarini et al. (2009) and Mallakpour and Villarini (2016). Studies such as Kalra et al. (2008), Miller and Piechota (2008), and Sagarika et al. (2014) have also used trend and shift detection tests—however, these studies considered trends and shifts to be independent of each other, at least when applying the tests. A recent study by Serinaldi and Kilsby (2016) suggests that the shift detection tests are also affected by the presence of trends and persistence, especially by the short-term persistence (STP). Hence, detrending the data (if required) and removal of persistence, before the trend and/or shift detection tests, becomes imperative to produce results with higher reliability. In cases of trend detection tests with persistence in data, the trend (or the slope of the trend to be specific) is removed from the time series (known as detrending) before applying the correction for persistence (autocorrelation) (Kumar et al. 2009). Once the persistence is removed, the trend is added back to the time series for further analyses. A similar procedure can be applied for shift detection tests as well (Serinaldi and Kilsby 2016).
In this study, simulated temperature and precipitation time series of CMIP5 (referred to as CMIP5 temperature and precipitation models in the later sections of the manuscript) for the upper and lower Colorado River basins (UCRB and LCRB, respectively) were analyzed to determine the presence of significant trends and shifts with respect to persistence. A total of 41 temperature and 25 precipitation CMIP5 models were analyzed over a study period of 104 years, from 1901 to 2004. For each model, a significant SP was detected (if present) using the Pettitt’s test (Pettitt 1979), followed by the analyses of trends before and after the significant SP using the Mann–Kendall (MK) test (Mann 1945; Kendall 1975). Both the Pettitt’s and the MK tests are nonparametric in nature, which makes them appropriate for climate data distributions that are often subject to nonstationarity and nonnormality (Kumar et al. 2009). Villarini et al. (2009) compared multiple parametric and nonparametric tests (using datasets with predefined historical shift points) to detect both shifts and trends and concluded that the Pettitt’s and the MK test were the most reliable ones in detecting shifts and trends, respectively. For the shift detection test, trend-free prewhitening (TFPW), based on Yue et al. (2002), was applied as suggested by Serinaldi and Kilsby (2016), to remove the effect of STP from the models. Several modified MK tests were applied, based on Kumar et al. (2009), to determine the effects of both the long-term persistence (LTP) and STP on the identified trends. To evaluate the ability of CMIP5 models to capture the observed trends, under the influence of SPs and persistence, data from the Climate Research Unit Time Series version 3.10 (CRU-TS) were analyzed. Both the CMIP5 and CRU-TS datasets were analyzed at identical spatial and temporal resolutions.

From a literature review of recent studies (Villarini et al. 2009; Venema et al. 2012; Williams et al. 2012), this study hypothesized that trend analyses are better explained if SP detection tests are included prior to the trend tests as was earlier explained with the example in Fig. 1. Hence, the first objective of the study was to detect SPs in both the modeled and observed datasets and to evaluate the presence of trends independently before and after the SPs. Studies also suggest the importance of the removal of persistence preceding both the trend and shift detection tests. Therefore, as the second objective, the study evaluated the effects of persistence on the detected trends and shifts. As the final objective, the efficacy of CMIP5 models to match the observed trends, considering both the significant SPs and persistence, was evaluated across the Colorado River basin (CRB).

2. Study area and data

Data from 41 temperature and 25 precipitation CMIP5 simulation models (https://cmip.llnl.gov/cmip5) were obtained (in terms of anomalies calculated with respect to 1961–90 mean to be consistent with the observed data) from 1901 to 2004 for the UCRB and LCRB. The basin boundaries were obtained from the USGS hydrologic unit map (https://water.usgs.gov/GIS/huc.html). For each model, only the first ensemble member was considered in the analyses. The first ensemble member was deemed to be representative of the model as a whole based on the work of Kumar et al. (2013) who found the least variability between members of the same model and most variability between different models. The obtained data were regridded [as suggested by Kumar et al. (2013) using an area-average preserving method] to a common spatial resolution of $2.5^\circ \times 2.5^\circ$, which resulted in a total of 22 grid cells (boxes/squares) for the entire CRB (Fig. 2). Grid cells 1–10, 14, and 15 had the majority of their portions in the LCRB, and grid cells 11–13 and 16–22 had the majority of their portions in the UCRB (Fig. 2). The observed CRU-TS temperature and precipitation time series data (in terms of anomalies with respect to 1961–90 mean) (https://crudata.uea.ac.uk/cru/data/hrg/) were updated from 0.5° × 0.5° to $2.5^\circ \times 2.5^\circ$ using an area-average preserving method as well, as suggested by Kumar et al. (2013) and Mitchell and Jones (2005), to match the modeled data resolution. For both datasets, mean annual time series data were produced by aggregating the monthly temperature and precipitation data. Table S1 in the online supplemental material lists the names of all of the CMIP5 models analyzed in the study, and model acronym expansions may be found online (https://www.ametsoc.org/Pubs/AcronymList). Figure S1 of the supplementary material shows the temperature and precipitation anomaly data obtained from the CMIP5 model simulations (as multimodel averages) and CRU-TS observations across the study period for each of the grid cells.

3. Method

For the trend analyses, four different versions of the MK test (Mann 1945; Kendall 1975), adopted from Kumar et al. (2009), were applied. Similar modified versions of the MK test were found in the literature relating to determination and quantification of trends (Tamaddun et al. 2016). The first version, referred to as MK1 in the following sections, is the original form of the MK test (Lettenmaier et al. 1994). The second version, referred to as MK2, is a modified MK test with TFPW, as
explained by Yue et al. (2002), and considers the effect of STP or lag-1 autocorrelation in the data distribution. The third version, referred to as MK3, is a modified MK test that takes into account all of the significant autocorrelation structures present in a time series (Hamed and Rao 1998). The fourth version, referred to as MK4, is a modified MK test that considers the effect of LTP, also known as the Hurst phenomenon (Hamed 2008) in the hydrology literature (Koutsoyiannis 2003).

The MK test, at its core, is based on the null hypothesis that the time series has no trend. The test determines the direction of a trend, which can be either positive, negative, or absent, based on the first MK test statistic (obtained from a signum function). Later, another test statistic is determined (from the first test statistic), which gives the significance level of rejecting the null hypothesis. The modified versions of the MK test, as mentioned earlier, check for the presence of autocorrelation (Kumar et al. 2009) and based on the type of autocorrelation, either modify the time series (MK2 applies prewhitening) before determining the significance of the MK test, or adjust the formula to determine the test statistic (MK3 and MK4). Hence, a time series under MK1 may suggest a trend at a certain significance level; however, the same time series may not be statistically significant under MK2, MK3, or MK4, or vice versa. Besides the MK test, the Theil–Sen approach (TSA) (Theil 1950; Sen 1968) was adopted to quantify

FIG. 2. Map showing the upper and lower Colorado River basins across the selected grid cells. Each grid cell had 41 temperature and 25 precipitation values (from different simulation models) at each year of the study period (i.e., from 1901 to 2004). For example, this map shows the variation in temperature anomaly across the grid cells during 1901 in model 1.
the magnitude of the trends. TSA is nonparametric in nature, hence it is considered appropriate to be applied in conjunction with the MK test. TSA determines the median slope of all the possible pairs in a time series. Thus, the test is robust against possible outliers.

Pettitt's test (Pettitt 1979), with and without TFPW (Yue et al. 2002), was used to determine the presence of significant SPs in the data distributions, as was suggested by Serinaldi and Kilsby (2016). Pettitt’s test (Pettitt 1979), at its core, detects the shift (using a signum function similar to the MK test) in a time series by testing the anomaly between the mean of two independent samples from the same time series. The detected shift, which can be either positive, negative, or absent, is then tested for a probability estimate based on the significance level used. The TFPW process removes the trend (detrending) from a time series by subtracting the slope of the trend (Yue et al. 2002) and then removes the serial correlation from the detrended time series. Hence, with TFPW, the Pettitt’s test becomes more reliable in cases of time series with underlying trends. Depending on the presence of a significant SP, each of the time series was divided at the SP, and all the MK tests were applied before and after the SP separately. This approach of detection of an SP and then applying the MK tests prior to and after the SP has been labeled as “shift trend” in the following sections. A confidence level of 90% (significance level of 10% or $p \leq 0.10$) was used for the significance of all the tests.

4. Results

a. Shift trends in temperature data

The majority of the CMIP5 temperature models at each of the grid cells suggested a positive shift [online supplementary material: column four (±) of Table S2] approximately between 1940 and 1980 (according to the SPs one standard deviation above and below the mean SP) (Fig. 3a). The overall range of the SPs was found to be slightly wider (1916 to 1993). The mean SPs of all the CMIP5 models had a significantly narrower range, that is, between 1960 and 1970, except for grid cells 11, 13, and 14, which experienced a slightly earlier shift. The modal SPs across the CMIP5 models (Table S2) suggest that almost all the grid cells had shift points...
between 1960 and 1980 (with a few exceptions of earlier SPs). A few models at each grid cell showed negative shifts, but the numbers were very few relative to the positive shifts (Table S2). The effect of STP on the SPs was found to vary across the grid cells—for some grid cells, the mean SP was shifted by more than a decade (e.g., grid cells 3 and 13–15) (Fig. 3a).

Comparison between the trends before and after the shifts showed that the total number of models at each grid cell with positive trends increased by approximately a factor of three after the shifts (last row of Table S2). On the contrary, negative trends were hardly observed after positive shifts, even though there were models with negative trends before the shifts. Both MK3 and MK4 had a smaller number of models with trends than did MK1, whereas MK2 had a higher number of models with trends than did MK1 (both before and after the SPs) (Table S2). The LCRB had a higher number of models with positive trends than did the UCRB, both before and after the SPs. On the contrary, the UCRB had a higher number of models with negative trends than did the LCRB before the shifts, even though there are more grid cells in the LCRB than the UCRB. There was only one model with a negative trend after the SP in the UCRB (under MK1 and MK2), while the negative trends in the LCRB were all neutralized after the SPs (Table S2). The effect of positive shifts in the detected trends was also evident from the TSA trend slopes evaluated before and after the shifts. Figures 4a and 4b show that the mean trend slope across the grid cells increased by a factor of 10 after the shift. The median TSA slopes, both before and after SPs, were all found to be positive across the models at each grid cell.

The CRU-TS temperature data showed a higher variation among the SPs across the grid cells as compared with the mean SPs of the modeled data. Almost all the grid cells had a positive SP between 1930 and 1980 (Fig. 3b). The effect of STP was found to be higher in the CRU-TS data since the majority of the grid cells did not show significant SPs after applying TFPW. Grid cells without a significant SP were infilled by the average of the significant SPs found across the basin (named as the basin-mean SP) in Fig. 3b. The effect of shifts in the detected trends was found to be less in the CRU-TS temperature data than it was in the modeled temperature data—especially in MK3 and MK4—as the total number of grid cells with significant trends were similar before and after the significant shifts (online supplementary material: Table S3). In MK1 and MK2, the number of grid cells with positive trends was close to double after the positive shifts. Except for two grid cells in the LCRB, the direction of the trends in all the other grid cells was positive, both before and after the positive shifts (Table S3). The UCRB had a higher number of grid cells with positive trends relative to the LCRB, which was the opposite of what was found among the CMIP5 models, even though there are more grid cells in the LCRB than the UCRB. The mean TSA trend slopes across the grid cells showed a significant rise, an increase by a factor of 4, after the positive shifts (Figs. 4c,d). The rate of increase in the slope values, from before the shifts to after the shifts, was found to be higher in the central grid cells compared to the eastern and western grid cells.

A basinwide analysis revealed a noticeable increase in the basin-mean trend slope (obtained by averaging the slopes of all the grid cells within the CRB) after a significant shift—in both the modeled and observed datasets (Fig. 4e). Although the modeled mean shift (1962) (obtained from averaging all the SPs within the CRB) occurred a decade later than the observed mean shift (1952), the increase in slope after the shift was found to be very comparable (within the same order and consistent direction) with the observed data. The modeled mean slope changed from +0.0023 to +0.0232 (increased by a factor of 10), while the observed mean slope changed from +0.0038 to +0.0157 (increased by a factor of 4) (Fig. 4).

b. Shift trends in precipitation data

In the CMIP5 precipitation models, it was difficult to make a conclusion about the general direction of the shifts since the number of models with positive and negative shifts at many of the grid cells was comparable (equal in some instances) [online supplementary material: column four (±) of Table S4]. Besides, only a small proportion of the precipitation models showed a significant shift as compared with the temperature models (last row of Table S4). However, for the entire CRB, a higher tendency toward negative shifts was observed from the analyses [last column of column four (±) of Table S4]. The SPs at each grid cell across the majority of the models were found between 1920 and 1980 (according to the SPs one standard deviation above and below the mean SP as well as in the overall range) (Fig. 3c). The mean SPs among the models at each grid cell were found within a narrower range (i.e., between 1926 and 1962). There was no significant mode among the SPs across the models (Table S4). The effect of STP was found to be higher among the CMIP5 precipitation models, as TFPW shifted the mean SPs by more than a decade for many of the grid cells, for example, cells 3–5, cell 11, and cells 14–15 (Fig. 3c).

A neutralizing tendency, in terms of the direction of the trends, was observed between the trends before and
after the shifts. The number of models with positive trends after the shifts was reduced to close to one-half relative to before the shifts (last row of Table S4). Models with negative trends were hardly present after the shifts, except for one model under the MK3 in the UCRB. MK1, MK2, and MK4 had comparable results—both before and after the shifts (Table S4). MK3 had the highest number of positive and negative trends in models. The LCRB had a higher number of models with positive trends both before and after the shifts than the UCRB (Table S4). The LCRB also had the highest number of models with negative trends before the shifts under MK3, but all the trends were neutralized after the shifts. From the TSA trend slopes, it was difficult to determine the general direction of the precipitation trends across the grid cells (Figs. 5a,b). Before the shifts, the mean TSA slope was found to be slightly above zero, with the eastern and western grid cells having relatively...
positive slopes and the central grid cells having relatively negative slopes. After the shifts, the mean TSA slope was found to increase, but that increase can be attributed to the absence of significant SPs in many of the modeled data distributions—which overestimated the overall mean trend across the CRB. The TSA slopes after the shifts showed a higher rate of increase among the central grid cells than among the eastern and western grid cells.

The effect of STP was not conclusive as the majority of the grid cells (online supplementary material: Table S5)—both with and without TFPW—did not show significant SPs in the CRU-TS precipitation data (Fig. 3d). The results of the different versions of the MK tests were quite similar and showed only negative trends both before and after the shifts. In fact, MK1, MK2, and MK4 had exactly the same results (Table S5). The majority of the shifts were found to be positive, which neutralized the preexisting negative trends and resulted in the absence of trends after the shifts. The LCRB had a higher number of grid cells with negative trends than did the UCRB—both before and after the shifts (there was only one grid cell with negative shift in the LCRB after the shift) (Table S5). The mean TSA slope across the grid cells before the shifts was close to zero (slightly negative) (Fig. 5c). Many of the grid cells did not have
a significant shift, which produced inconsistency in the estimation of trend slopes. As the grid cells without significant shifts had no trend after the shift (such grid cells had a trend for the entire study period represented as a trend before the shift in Fig. 5c), those cells did not contribute in calculating the mean slope after the shift—hence, those cells fell on the no trend line in Fig. 5d (shown by open symbols). Therefore, the mean slope was found to decrease after the shift since the negative slopes were found to be of higher magnitude relative to the positive slopes among the grid cells with significant shifts (Fig. 5d). A higher magnitude of negative slopes among a handful of grid cells (all the grid cells did not have a significant shift) may have created a negative bias in the mean slope after the shift. The eastern grid cells, both before and after the shifts, showed fewer slopes than did the western grid cells. Many of the grid cells that did not show a significant SP had positive trends across the entire study period.

The basinwide analysis did not produce a satisfactory comparison as the basin-mean slopes (obtained by averaging the slopes of all the grid cells within the CRB) were found to be inconsistent between the modeled and observed datasets. The modeled data showed an increase (from 0.0091 to 0.0195 mm yr\(^{-1}\)) in the basin-mean slope (obtained from averaging all the SPs within the CRB) while the observed data suggested a decrease (from \(-0.0039\) to \(-0.0109\) mm yr\(^{-1}\)) (Fig. 5e). Unlike the temperature models, precipitation models experienced the mean shift earlier (1951) than the observed shift (1958) (Fig. 5e). The distribution of the observed data revealed that the grid cells with decreasing trends had a higher influence on the basinwide mean.

5. Discussion

The ability of CMIP5 simulation models to match the observed trends under the influence of shifts and persistence was evaluated in this study. The Pettitt’s test, preceded by TFPW (Yue et al. 2002) was used to detect the significant SPs, as was suggested by Serinaldi and Kilsby (2016). In comparison with the CMIP5 temperature models, the precipitation models experienced a higher influence of STP (Figs. 3a,c) as the mean SP of some of the grid cells moved by more than a decade after the TFPW. Such movement in SPs can be attributed to the absence of significant SPs in some of the modeled data distributions after applying TFPW—which can produce a bias in the mean SP. In the observed data distributions, the effect of STP was higher as the majority of the grid cells did not show any significant SP after TFPW (Figs. 3b,d). However, for such grid cells having SPs both before and after TFPW, the locations of the SPs were found to be comparable (e.g., grid cells 1, 8, 15, 19, and 22 in Fig. 3b and grid cells 6, 7, 14, and 16 in Fig. 3d). The results suggest that the effect of STP was higher in both the modeled and observed precipitation data relative to the temperature data. Other prewhitening techniques, for example, AR(1)-based prewhitening and fGn-based prewhitening, can be evaluated in future studies to verify the obtained results. However, the results may not vary significantly as was suggested by Serinaldi and Kilsby (2016).

The CMIP5 temperature models had a narrower range of mean SPs (between 1957 and 1968) relative to the observed temperature data (between 1924 and 1985) (Fig. 3 and Table S3). Furthermore, the temperature models showed the tendency of detecting significant SPs at a later date relative to the observed data. The variation of SPs among the adjacent grid cells was much higher in the CMIP5 precipitation models than in the temperature models. The results of the modeled and observed precipitation datasets were not highly comparable as many of the grid cells in the observed data did not have a significant SP. Contrary to the temperature models, the precipitation models detected the SPs earlier than the observed SPs. Such a tendency of detecting shifts earlier than the observations was also evident from the basinwide shift of temperature and precipitation (Figs. 4e and 5e). This suggests an overall bias (a tendency of detecting shifts earlier or later than the observed shifts) in the modeled data distributions.

The analyses suggest that the SPs across the grid cells (also the basin-mean SP) were located within a range of a few decades for both the temperature and precipitation models. Previous studies have looked into the correspondence between the phases of large-scale climate variabilities originating from the SST fluctuations (i.e., El Niño–Southern Oscillation, Pacific decadal oscillation, and Atlantic multidecadal oscillation) and the station-based hydroclimatic data of the CRB (Tamaddun et al. 2017; Kalra et al. 2017; Nowak et al. 2012). The obtained results were found to be fairly consistent with the shifts observed in the previous studies. The range of results among the models and their inherent uncertainty due to natural climate variability was discussed by Deser et al. (2012a,b), which may explain the comparable SPs across the models. Deser et al. (2012a) also mention the importance of large-scale fields on the downscaled local climate models. Future studies may consider the findings of the study in intermodel and intervariable analyses of the models as a response to large-scale climate variabilities.

The effect of persistence on trends was found to be consistent between the variables in the modeled and
observed datasets. Both of the temperature datasets (modeled and observed) were more greatly influenced by STP, relative to the other types of autocorrelations, as TFPW (MK2) increased the number of significant models when compared with MK1, whereas the numbers of significant models were fewer in MK3 and MK4 (Tables S4 and S5). The precipitation data distributions were found to be highly influenced by the significant autocorrelation structures, with a higher number of significant models in MK3 than in the other MK tests (Tables S4 and S5). Studies suggest that the effect of persistence can have different associations among correlated variables even within a particular region (Kumar et al. 2009). The consistency between the modeled and observed datasets observed in the persistence of a variable can be compared with other hydrologic basins to evaluate its uniformity across a larger study area in future studies.

The effect of shifts on the detected trends was found to be much higher in temperature than in the precipitation datasets. The positive shifts in the temperature datasets increased the trend slopes significantly after the shifts. The slopes—both before and after the shifts—were found to be fairly consistent between the modeled and observed datasets. However, the rate of increase, from before the shift to after the shift, was higher in the CMIP5 models. Before the shifts, the slopes varied from the eastern to the western grid cells, but they were found to be consistent (all positive) after the shifts, which suggests a basinwide shift toward a positive trend in both the modeled and observed datasets. Though CMIP5 temperature models overestimated the magnitude of the trend, both before and after the shifts, the models satisfactorily captured the observed trend slopes in terms of direction (which was not the case with the precipitation models).

In the precipitation datasets, the absence of significant SPs may have led to a bias in the estimation of trends since the effective number of grid cells (while calculating the mean slope) was inconsistent between the modeled and observed distributions as well as between before the shift cases and after the shift cases (Fig. 5). Some of the CMIP5 models did not have a significant SP, which overestimated the variance of the trend slopes and may have produced biased results. A similar occurrence of bias was found in the observed data as many of the grid cells did not have a significant SP. Though the trend slopes between the modeled and observed datasets were inconsistent (and hence incomparable), both of the datasets showed a lack of SPs in their distributions. This suggests that the precipitation and temperature trends in the CRB did not experience identical changes over the study period. The presence of shifts, in strengthening or neutralizing an existing trend, was much higher in the temperature trends than in the precipitation trends. Although Kumar et al. (2013) did not evaluate the presence of SPs, they concluded that “globally, precipitation trends are distributed (spatially) at about zero in both the models and the observations.” They also found that modeled temperature trends correspond well with the observed trends on a global scale, which was also found in the current study for the CRB. The influence of modeled and observed datasets varied across the UCRB and LCRB without any noticeable consistency within the variables. However, before the shifts, especially in the temperature data, the eastern and western grid cells showed higher slopes, while after the shifts, the central grid cells showed higher slopes. The observed variation from east to west can be a scope of research for future studies.

6. Conclusions

This study evaluated the effects of SPs and persistence on the trend analyses of the CMIP5 temperature and precipitation models for the entire Colorado River basin across 104 years. The obtained results were compared with the observed CRU-TS datasets. The results showed that the trends in temperature models were relatively consistent with the observed trends—both in the direction and magnitude. For both of the precipitation datasets, the absence of significant SPs in many of the grid cells led to incomparable and inconsistent results. The variation in SPs among the adjacent grid cells was higher in the modeled precipitation and observed temperature data. Both of the temperature datasets showed a higher influence of STP in their detected trends (MK2), whereas both precipitation datasets were highly influenced by the significant autocorrelation structures of the time series (MK3).

The major contributions of this study are the following:

1) A shift-trend approach was adopted in which the direction and magnitude of the trends were evaluated independently before and after significant SPs for both the modeled and observed datasets. This allowed the current study to evaluate the effect of shifts in strengthening or neutralizing the preexisting trends.

2) The effect of persistence was thoroughly examined not only for the trend tests but also for the detection of SPs. The trend analyses revealed that the effects of persistence varied between the variables but were consistent between the modeled and observed datasets. The SPs were found to be influenced by
persistence in both of the datasets with higher influence on the observed datasets.

3) The inconsistencies detected between the CMIP5 and observed datasets—under the influence of SPs and persistence—can be helpful in improving the models. The observed biases can be helpful to practitioners who depend highly on data-driven modeling.

Future studies may look into the uncertainties related to large-scale climate variabilities, natural variability, and the modeled distributions. The inclusion of multiple ensemble members can also be another scope of research. Comparing the results with adjacent hydrologic basins may provide important insights with regard to other possible biases in the CMIP5 models.

Acknowledgments. The authors thank phase 5 of the Coupled Model Intercomparison Project (CMIP5) and Climate Research Unit Time Series (CRU-TS) (version 3.1) for the datasets used in this study. The authors also thank the U.S. Geological Survey (USGS) for their hydrologic unit map. The authors are grateful to the reviewers whose comments improved the overall quality of this paper.

REFERENCES


——, ——, M. Bernaderr, and S. Ahmad, 2017: Multi-scale correlation between the western U.S. snow water equivalent and


