Use of the Autoregressive Integrated Moving Average (ARIMA) Model to Forecast Near-Term Regional Temperature and Precipitation

YUCHUAN LAI AND DAVID A. DZOMBAK

Department of Civil and Environmental Engineering, Carnegie Mellon University, Pittsburgh, Pennsylvania

(Manuscript received 26 July 2019, in final form 25 December 2019)

ABSTRACT

A data-driven approach for obtaining near-term (2–20 years) regional temperature and precipitation projections utilizing local historical observations was established in this study to facilitate civil and environmental engineering applications. Given the unique characteristics of temporal correlation and skewness exhibited in individual time series of temperature and precipitation variables, a statistical time series forecasting technique was developed based on the autoregressive integrated moving average (ARIMA) model. Annual projections obtained from the ARIMA model—depending on individual series—can be interpreted as an integration of the most recent observations and the long-term historical trend. In addition to annual temperature and precipitation forecasts, methods of estimating confidence intervals for different return periods and simulating future daily temperature and precipitation were developed to extend the applicability for use in engineering. Quantitative comparisons of annual temperature and precipitation forecasts developed from the ARIMA model and other common statistical techniques such as a linear trend method were performed. Results suggested that while the ARIMA model cannot outperform all other techniques for all evaluated climate indices, the ARIMA model in general provides more accurate projections—especially interval forecasts—and is more reliable than other common statistical techniques. With the use of the ARIMA-based statistical forecasting model, interpretable and reliable near-term, location-specific temperature and precipitation forecasts can be obtained for consideration of changing climate in civil and environmental engineering applications.

1. Introduction

Future regional climate information is highly valuable for risk assessment and decision-making (Hallegatte 2009, e.g., infrastructure adaptation; Olsen 2015), and such regional climate projections can be acquired from the General Circulation Models (GCMs) with the application of downscaling methods. As GCM projections are often too coarse for accurate regional climate change assessment, statistical or dynamical downscaling methods (Hewitson and Crane 1996; Trzaska and Schnarr 2014) are often used to transform global projections into finer resolution. Some examples of statistical and dynamical downscaling techniques include bias correction and spatial disaggregation (BCSD; Wood et al. 2004), bias correction with constructed analogs (BCCA; Maurer et al. 2010), multivariate adapted constructed analogs (MACA; Abatzoglou and Brown 2012), localized constructed analogs (LOCA; Pierce et al. 2014), the North American Regional Climate Change Assessment Program (NARCCAP; Mearns et al. 2009), and Coordinated Regional Downscaling Experiment (CORDEX; Giorgi and Gutowski 2015). USGS Center for Integrated Data Analytics (CIDA 2019; https://cida.usgs.gov/thredds/catalog.html) and National Center for Atmospheric Research Climate Data Gateway (NCAR 2019; https://www.earthsystemgrid.org/) are some examples of the online repositories for accessing the statistical and dynamical downscaled GCM datasets.

Downscaled GCM projections are highly insightful and have been widely used in previous studies for regional climate change assessment [e.g., Chapra et al. (2017) with algal blooms, Cook et al. (2017) with stormwater design, Houser et al. (2017) with estimation of economic damages, Larsen et al. (2016) with public infrastructure in Alaska, and Van Vliet et al. (2016) with

Supplemental information related to this paper is available at the Journals Online website: https://doi.org/10.1175/WAF-D-19-0158.s1.

Corresponding author: Yuchuan Lai, ylai1@andrew.cmu.edu

DOI: 10.1175/WAF-D-19-0158.1

power generation systems]. However, the use of such projections in practical applications is still challenging. As discussed in previous studies (Olsen 2015; ESTCP 2017), practitioners are uncertain about the use of downscaled GCM projections because of the multiple tiers of uncertainties from future greenhouse gas emissions, climate models, and different downscaling methods. In addition, when utilizing downscaled GCM projections various gaps still exist between the provided climate information and the climatic design values used in engineering practice (Hyman et al. 2014).

Research is in progress to address the need for more credible climate information in near-term time frames such as 2–20 years (e.g., SERDP 2015); such work remains challenging (Stouffer et al. 2017). Acknowledging the uncertainty in the near-term projections, phase 5 of the Coupled Model Intercomparison Project (CMIP5) has initiated ''near-term experiments'' (Meehl et al. 2009) or decadal climate prediction project (Eyring et al. 2016) with 10–30 years as the projection time to improve near-term climate model forecasting capabilities. These experiments are currently in active research toward CMIP6 with some significant challenges (Stouffer et al. 2017). For the convenience of discussion and analyses, near-term projections (or forecasts) were defined as 2–20 years in projection time for this study.

Instead of using GCMs, an alternative is explored and developed in this study to obtain near-term regional temperature and precipitation forecasts by assessing and extrapolating the historical climate record of a particular location. Krakauer and Fekete (2014), for example, showed that a linear extrapolation of historical trend can provide useful forecasts for projection time less than 25 years. Moreover, engineers are ''comfortable''—as described in Hyman et al. (2014)—with the use of historical data during designs and an integration of historical trend with future projections can be ''advantageous.'' Additionally, increasing computational power and the development of statistical analysis tools such as the computer language R in recent years have greatly simplified the processes for data analyses.

Climate forecasts are routinely produced and provided by agencies such as the ''long-range climate prediction'' (O'Lenic et al. 2008) and the North American multimodel ensemble (Kirtman et al. 2014) from the Climate Prediction Center (CPC) of the NOAA National Weather Service (NWS). With the use of various dynamical climate models or statistical techniques (O'Lenic et al. 2008; Kirtman et al. 2014), these climate forecasts by NWS include three-month seasonal projections up to 12.5 months in advance (CPC 2019a; https://www.cpc.ncep.noaa.gov/products/predictions/90day/) or a monthly mean up to 7 months in advance (CPC 2019b; https://www.cpc.ncep.noaa.gov/products/NMME/). However, because some practical applications such as infrastructure designs need to consider durations of multiple years or decades, the method presented in this study focuses on providing improved projections in the 2–20-yr time frame. Additionally, while year-to-year short-term climate variations are important and addressed in NWS long-range seasonal forecasts, projections of these year-to-year climate variations are not the focus of the presented method. Instead, the developed statistical method is focused on providing probabilistic projections—average levels and thresholds of potential near-term future values—of regional temperature and precipitation for engineering applications under a changing climate.

Moreover, practical applications often require derived climate data (or climatic design values) such as in different return periods (Auld et al. 2010) and use of environmental and infrastructure process models often require the input of daily temperature and precipitation data (Verdin et al. 2018). Therefore, to further increase of applicability in engineering practice, another goal of the work was to facilitate the estimation of different return periods and generation of daily temperature and precipitation with obtained near-term forecasts.

Therefore, the objective of this study was to develop an alternative to GCMs and downscaling for obtaining near-term, location-specific temperature and precipitation projections for engineering applications with the use of statistical techniques and regional historical observations. The work involved identification of an efficient, interpretable, and reliable statistical forecasting method, development of techniques for conversion of annual probabilistic forecasts into desired formats used in engineering applications, and quantitative assessment of the forecasting accuracy of the statistical forecasting method selected. To accommodate the unique characteristics of temporal resolution and skewness from individual time series of temperature and precipitation variables, the autoregressive integrated moving average (ARIMA) model was selected as the forecasting model.

2. Selection of a statistical time series forecasting model

a. Location-specific historical climate data

High-quality long-term historical regional climate data are needed to apply analyses and develop forecasts. A long-term, high-quality, and location-specific historical daily temperature and precipitation dataset constructed by Lai and Dzombak (2019) was used in this study. The daily climate data compiled by Lai and Dzombak (2019) include long-term daily temperature

and precipitation data for 93 U.S. cities with both temperature and precipitation records starting earlier than 1900. These climate data were acquired from the Applied Climate Information System (ACIS) database developed by the NOAA Northeast Regional Climate Center (DeGaetano et al. 2015), part of the greater Global Historical Climatological Daily (GHCN-D) dataset, and originally collected from the NWS Cooperative Observer Network (COOP; Leeper et al. 2015). As the construction of daily time series for individual cities requires the combination of records from multiple local weather stations, forecasting of some time series (especially daily T_{min}-related extreme-occurrence indices) in some cities should be performed with caution. See section A in the online supplemental material for a more detailed discussion of this topic.

b. *Temporal correlation and skewed distribution*

To facilitate the use in engineering applications, an efficient and adaptable statistical forecasting method with high-interpretability was preferred in this study. Unique characteristics exhibited in the time series of climate variables, as discussed in our previous related work (Lai and Dzombak 2019), impede the use of a simple, direct approach such as a linear trend method. The unique characteristics for time series of climate variables include temporal correlation and nonnormal (skewed) distribution.

The climate record for a particular location can exhibit great variations from an average level with a few years to decades in cycle due to natural variability. Time series of climate variables, especially extreme indices, can also exhibit skewed distributions. Examples of the decade-long subtrends include the "global cooling period" in 1970s (Peterson et al. 2008) and the high precipitation amounts observed from the 1890s through the 1900s in the United States (Kunkel et al. 2003). From a time series analysis perspective, the decade-long subtrends indicate strong temporal correlation (Montgomery et al. 2016), also referred to as autocorrelation. In addition, temperature or precipitation extreme series can exhibit skewed distributions and consequently extreme value theory and various types of distributions are often utilized to perform statistical analyses (AghaKouchak et al. 2013). These characteristics of temporal correlation and skewed distribution thus increase the challenges of utilizing some common statistical techniques like a linear trend method.

Considering the potentially inherent temporal correlation and skewed distribution (Franzke 2012; Lai and Dzombak 2019) in individual time series of climate variables, an efficient, adaptable, broadly applicable, and statistically appropriate time series forecasting model—the ARIMA model—was identified, utilized, and assessed in this study for near-term regional climate forecasts.

c. *Background of the selected statistical techniques for time series forecasting*

1) THE ARIMA MODEL

The ARIMA model, developed by Box and Jenkins (1970) and also known as the Box–Jenkins method, refers to a class of time-domain models that are commonly used to fit and forecast time series exhibiting temporal correlation (Wilks 2006; Montgomery et al. 2016). Among various applications, the ARIMA model has been utilized in previous studies for forecasting time series of climate or climate-related variables [e.g., monthly rainfall and temperature by Kaushik and Singh (2008), monthly rainfall by Mahsin et al. (2012), and annual runoff by Wang et al. (2015)].

This class of ARIMA models has a general term ARIMA(p, d, q) that includes three components (or terms): autoregressive (AR), integrated (or differencing), and moving average (MA) terms with the corresponding order of p, d, and q. The magnitude of temporal correlation exhibited in the time series will determine the AR and MA terms while the differencing term can transform a nonstationary series to be stationary (Montgomery et al. 2016). The ARIMA (p, d, q) model for a time series is given as (Hyndman et al. 2018)

$$y'_t = c + \phi_1 y'_{t-1} + \cdots + \phi_p y'_{t-p} + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t, \quad (1)$$

where c is a constant (also known as the drift term when $d = 1$), $\phi_1 y'_{t-1} + \cdots + \phi_p y'_{t-p}$ is the AR term with ϕ_1 to ϕ_p as coefficients at p order, $\theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}$ is the MA term with θ_1 to θ_q as coefficients at q order, ε_t is an error term for random background noise at time t, and y'_t is the differencing series. For a first-order differencing,

$$y'_t = y_t - y_{t-1} \quad \text{(first-order differencing)}, \quad (2)$$

where y_t is the observation at time t.

Once the orders are determined and the coefficients are estimated for a particular time series by fitting the model with historical data, point forecasts and interval forecasts can be obtained (Hyndman and Athanasopoulos 2018). Point forecasts, or "best" forecasts determined through minimization of sum of squared errors (Montgomery et al. 2016), can be made using Eq. (1), while interval forecasts—known as forecast density or prediction intervals—can be obtained using point forecasts and estimated forecast errors. More detailed descriptions of

the ARIMA model fitting and forecasting procedures are provided in supplemental material section B.

The ARIMA model can be further developed to consider seasonal variations, temporal correlation in variance, or different external covariates. An integration of seasonal components with the ARIMA model yields a seasonal ARIMA (SARIMA) model (Montgomery et al. 2016), and the SARIMA model was utilized to provide future daily temperature and precipitation simulations as further discussed. In contrast to the ARIMA model that considers temporal correlation in mean, the general autoregressive conditional heteroskedasticity (GARCH) model (Taylor and Buizza 2004) can be used to describe temporal correlation in variance (or volatility). Although an ARIMA-GARCH model can potentially be used to describe volatility in the time series of climate variables (e.g., in Yusof and Kane 2013), such a model was not implemented in this study as the annual temperature and precipitation series evaluated for the U.S. cities do not exhibit substantial changes in variance (Lai and Dzombak 2019). Further discussions of the SARIMA and GARCH models are provided in supplemental material section B. In addition, the ARIMA model can be integrated with external covariates, for example, natural variability indices like El Niño–Southern Oscillation (ENSO; Hartmann et al. 2013) and interdecadal Pacific oscillation (Meehl et al. 2012) as well as greenhouse gas forcing. However, due to the scope of this work—to provide near-term climate forecasts solely based on regional historical observations—and the underlying complexity caused by various external covariates, an ARIMA forecasting model with inclusions of external covariates was not developed or utilized.

2) Combination of Box–Cox transformation

To further apply the ARIMA model for near-term climate forecasting, especially for time series exhibiting strong skewed distribution, the ARIMA model was utilized with potential for combination with Box–Cox transformation (Box and Cox 1964). Box–Cox transformation is a commonly used technique to normalize time series and improve the appropriateness of fitting data with statistical models that incorporate the assumption of normal distribution for residuals (Hyndman and Athanasopoulos 2018). In this study, Box–Cox transformation was utilized to improve normality prior to ARIMA model fitting if needed. Note that when the ARIMA model is discussed in the following text, if not specified, Box–Cox transformation of the data has been performed when needed. Box–Cox transformation of a time series is accomplished by using (Box and Cox 1964):

$$y_i^{(\lambda)} = \begin{cases} \dfrac{y_i^{\lambda} - 1}{\lambda}, & \text{if} \quad \lambda \neq 0 \\ \log(y_i), & \text{if} \quad \lambda = 0 \end{cases}, \tag{3}$$

where λ is the transformation coefficient and can be estimated by using the maximum likelihood method (e.g., with the log-likelihood function). Note that λ has to be larger than 0 when applying the Box–Cox transformation. If time series contain nonpositive values, a constant can be added to allow the transformation (Hyndman and Athanasopoulos 2018).

A reverse transformation for historical fitting, point forecasts, and interval forecasts is needed if the time series is modified with Box–Cox transformation. In the transformed space, both historical fitting and forecasts with the ARIMA model follow a normal distribution assumption. After the reverse transformation, the point forecasts will yield median projections (i.e., the median of the forecast density) and interval forecasts will follow a nonnormal distribution modified by the transformation.

3. Methodology of applying the ARIMA model to obtain near-term forecasts for engineering applications

a. Near-term projections of annual average temperature and total precipitation

A three-step process was used in this study to apply the ARIMA model to obtain near-term regional temperature and precipitation projections (further details provided in supplemental material section B):

(i) Shapiro–Wilk and Mann–Kendall tests were applied for individual time series to determine the significance of normality and trend. Box–Cox transformation and a nonzero drift term were integrated accordingly.
(ii) The time series was fitted with an ARIMA model with three orders selected using the auto.arima tool (Hyndman et al. 2018) in R with the default auto.arima method (the order of d for differencing was limited to be equal or less than 1).
(iii) Point forecasts and interval forecasts were obtained with the utilization of the "forecast" package (Hyndman et al. 2018) in R.

Examples of using the three steps to forecast annual average temperature and total precipitation are presented in Fig. 1. The 20-yr ARIMA forecasts starting from the years 1999 and 2019 were obtained for annual average temperature in Phoenix and annual total precipitation in Chicago, based on the historical observations

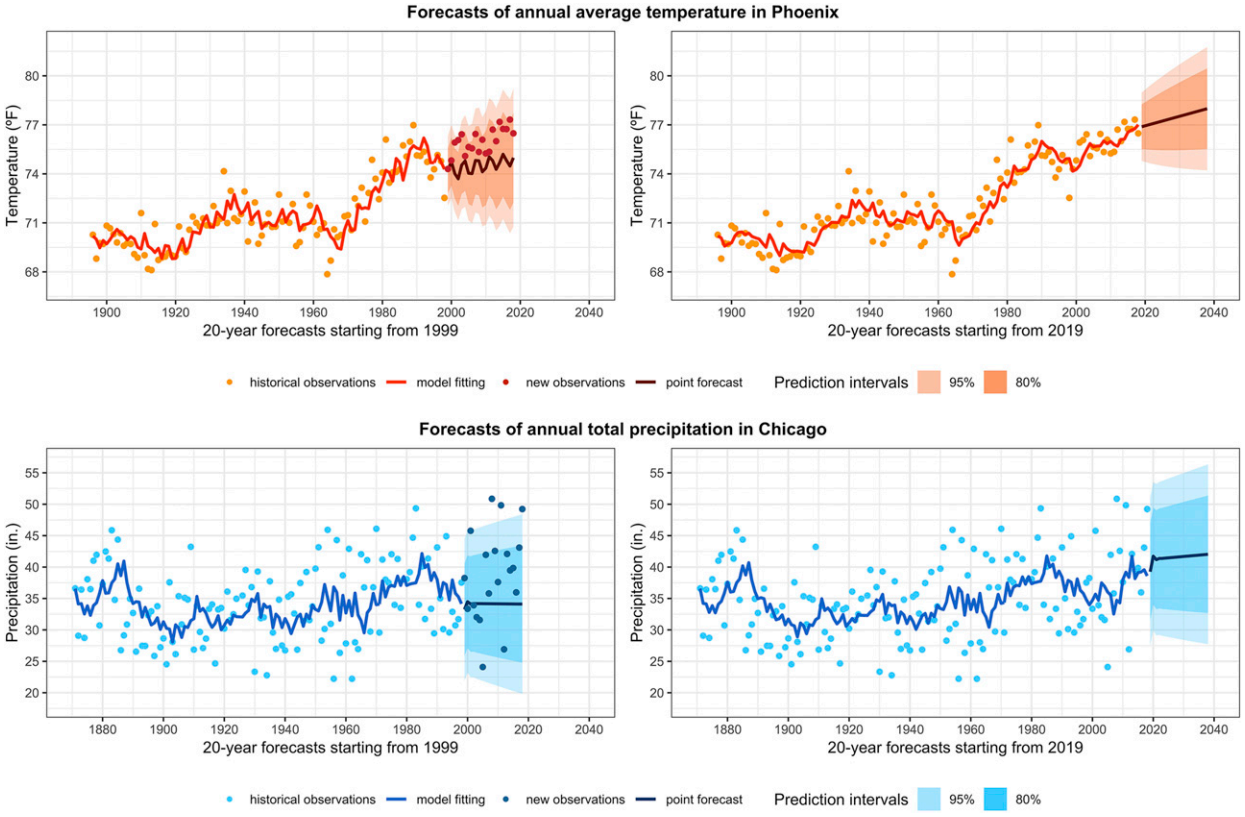

FIG. 1. The 20-yr ARIMA projections (starting from 1999 and 2019) of (top) annual average temperature in Phoenix and (bottom) total precipitation in Chicago. The lighter red (or blue) points show the historical temperature (or precipitation) observations, and darker points show new observations (for validation). The model fitting and point forecasts are presented as lighter and darker lines, while the 80% and 95% prediction intervals shown as the shaded area. Historical records start from 1896 for Phoenix and 1871 for Chicago.

prior to 1999 or 2019. Note that, to increase the flexibility, the orders of the ARIMA models were not fixed when fitting on the different lengths of the historical observations (prior to 1999 or 2019). By using the auto.arima tool, the orders of the ARIMA models can be different with the two starting years (further discussion is provided in supplemental material section B). For example, the annual average temperature forecasts in Phoenix starting from 1999 exhibited greater year-to-year fluctuations than the forecasts starting from 2019 because a different ARIMA model was used (the ARIMA [2, 1, 3] model for the forecasts beginning in 1999, compared to the ARIMA [0, 1, 1] model for forecasts initiating in 2019). In general, the ARIMA historical fitting—as exhibited in Fig. 1—fluctuates with the underlying subtrends (with ups and downs), allowing for the distribution of historical observations along the historical-fitting line. ARIMA point forecasts are substantially influenced by recent observations and the long-term historical trend. Prediction intervals are determined by the magnitude of residuals produced from the historical fitting, and at the same time increase with longer period of forecast, indicating increasing uncertainty in the more distant future.

b. *Near-term projections of annual temperature and precipitation extremes*

The ARIMA-based forecasting model was used to obtain near-term projections of annual temperature and precipitation extremes (i.e., climatic design values that are often required for engineering applications). The same three-step forecasting process was used during the forecast of the extremes, while a further integration of the bootstrap technique was implemented to provide additional information for the ARIMA forecasts.

It is worth noting that modeling of temperature and precipitation extremes is traditionally applied with extreme value theory such as the general extreme value (GEV) distribution and the general Pareto (GP) distribution (Wilks 1993), with several supplemental techniques to consider nonstationarity (AghaKouchak et al. 2013; Cheng et al. 2014). However, use of extreme value theory requires a specified parameterized distribution and is less ideal for estimation of precipitation amounts

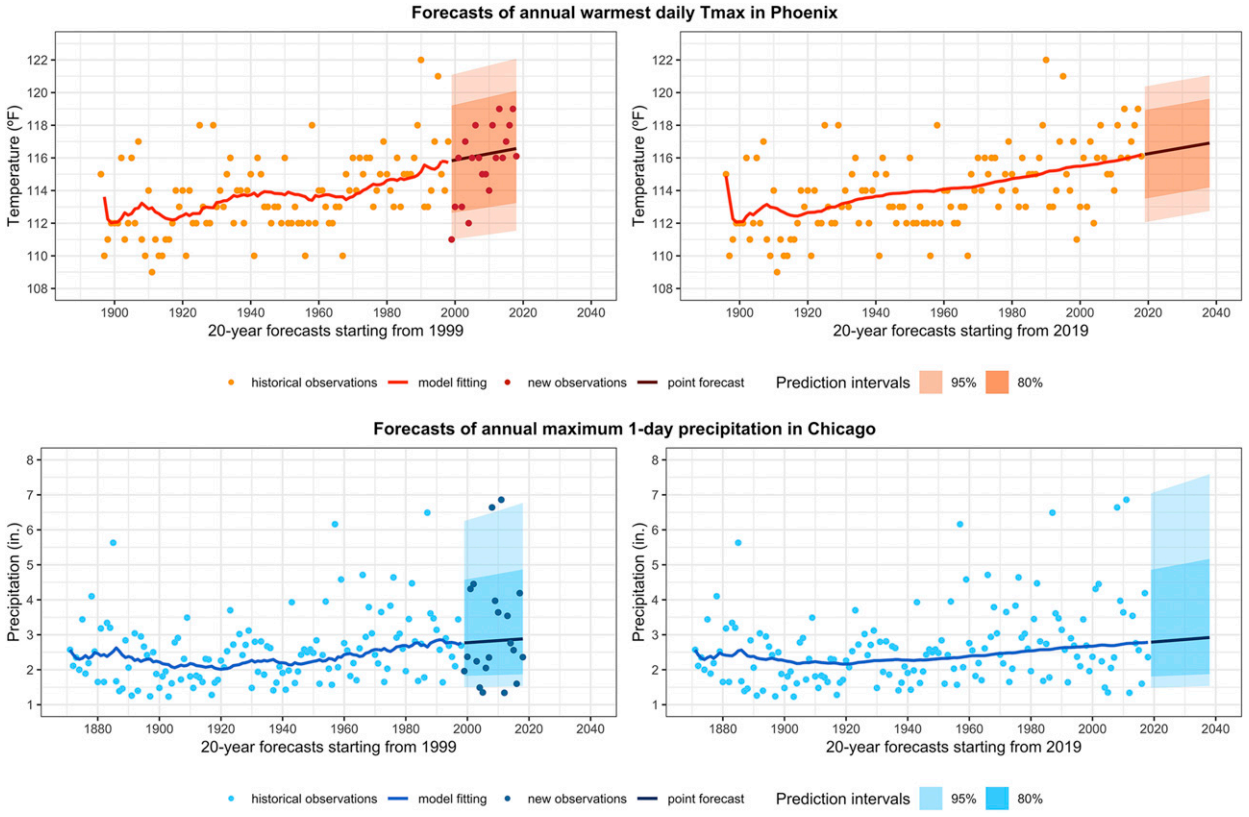

FIG. 2. The 20-yr ARIMA projections (starting from 1999 and 2019) of (top) annual warmest daily T_{max} in Phoenix and (bottom) annual maximum 1-day precipitation in Chicago. The lighter red (or blue) points show the historical temperature (or precipitation) observations and darker points are new observations (for validation). The model fitting and point forecasts are presented as lighter and darker lines, while the 80% and 95% prediction intervals are shown as the shaded area. Historical records start from 1896 for Phoenix and 1871 for Chicago.

with low return frequencies (DeGaetano and Zarrow 2011). With use of the ARIMA model and Box–Cox transformation to transform the time series of extremes when needed and avoid specifying a particular distribution, an efficient and adaptable approaching for fitting and forecasting temperature and precipitation extremes was utilized in this study.

Examples of applying the ARIMA model to provide near-term projections of annual temperature and precipitation extremes—annual warmest daily T_{max} in Phoenix and annual maximum 1-day precipitation in Chicago—are presented in Fig. 2. Similar to Fig. 1, 20-yr ARIMA forecasts were obtained starting from 1999 and 2019 based on historical observations prior to 1999 or 2019. While the two extreme series do not exhibit strong temporal correlation, right-skewed distribution can be observed for the annual maximum 1-day precipitation in Chicago. With use of Box–Cox transformation to modify the time series, the prediction intervals for the maximum 1-day precipitation in Chicago consequently have a higher upper-uncertainty bound.

The estimation of prediction intervals (especially the upper-uncertainty bounds) is important for practical application as different uncertainty bounds are often used to estimate climatic design values in different return periods, with a potential further estimation of confidence intervals for these different return periods (Bonnin et al. 2006; DeGaetano and Zarrow 2011; Cheng et al. 2014). The return period such as a 5-, 10-, or 50-yr event corresponds to a value that has an annual probability of exceedance (e.g., a 5-yr event has 20% of probability of exceedance annually), while the confidence intervals correspond to an estimation of uncertainty for that value. While different return periods can be directly calculated from the probability of exceedance with corresponding percentiles from interval forecasts, construction of confidence intervals for these percentiles requires estimation of additional uncertainty in forecasts, such as in Bonnin et al. (2006), DeGaetano and Zarrow (2011), and Cheng et al. (2014).

To facilitate an estimation of confidence intervals for different return periods, the ARIMA model can be

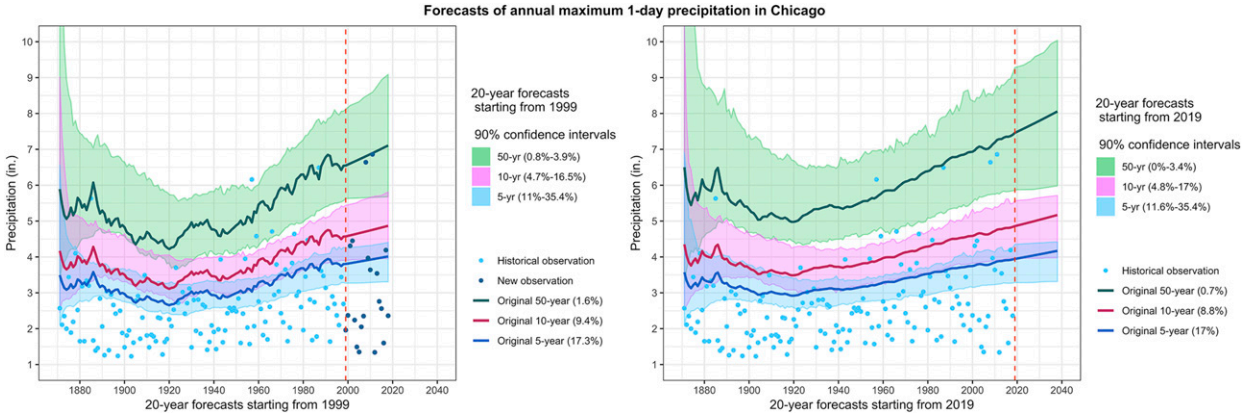

FIG. 3. The 20-yr forecasts (starting from 1999 and 2019) of 1-day precipitation amount in Chicago with 2-, 10-, and 50-yr return periods using an integration of the ARIMA model and bootstrap technique. The lighter blue points show the historical annual maximum 1-day precipitation observations, and darker points are new observations (for validation). The 90% confidence intervals for corresponding return periods are shown as shaded area. Historical records start from 1871. Percentages inside the parentheses (in the legend) indicate the percentages of historical observations exceeding the return period estimation during the historical fitting period (1871–1998 or 1871–2018).

further integrated with techniques like a bootstrap method to forecast annual extremes. In this study, the bootstrap method proposed by Pascual et al. (2004, 2005) was used to build on the ARIMA model and to estimate confidence intervals for different return periods. Detailed procedures of the described bootstrap technique are provided in supplemental material section C. The general approach is to utilize the same ARIMA model with same p, d, and q orders to fit the bootstrapped series, estimate the new coefficients for the three ARIMA terms, and calculate the new return periods.

An example of applying the ARIMA and bootstrap to estimate precipitation extremes with different return periods and associated 90% confidence intervals is provided in Fig. 3 for Chicago. With the application of the bootstrap technique, as shown in Fig. 3, the uncertainty bounds for both the historical and projected future precipitation extremes can be obtained and compared. Additionally, the confidence intervals can be used to evaluate the original ARIMA forecasts. For example, the original precipitation amount for the 5-yr return period in the 20-yr forecasts starting from 2019 (the right graph) showed a higher value than the median of the 90% confidence intervals, indicating that the original ARIMA forecast for the 5-yr event was potentially overestimated.

c. *Simulation of near-term daily temperature and precipitation*

In addition to annual extremes, practical applications of climate information often require daily temperature and precipitation as inputs for various process models (Verdin et al. 2018). An ARIMA-based simulation model was developed to facilitate the use of climate information in such engineering process models. These process models (or procedures in some cases) utilize daily temperature and precipitation to estimate the interactions with climate-related variables and to perform climate-related analyses such as for algal blooms (Chapra et al. 2017), pavement (FHWA 2016), water utilities (Vogel et al. 2016), and wildfire impact on culverts (FHWA 2017). An ARIMA-based statistical technique—employing an approach similar to that used in the ARIMA annual forecasts—was developed to provide future daily temperature and precipitation simulations (instead of providing point or interval forecasts in daily resolution).

It is worth noting that "weather generator" tools are widely available and can be used to provide daily temperature and precipitation simulations (e.g., Kilsby et al. 2007; Chen et al. 2010), with potential integration of GCM or downscaled GCM projections to provide simulations in future conditions (Maraun et al. 2010). Use of weather generators are valuable and informative for the use of process models and may potentially be required by engineering applications [e.g., the roof design as in Hosseini et al. (2018)]. However, weather generators and GCMs involve a substantial number of parameters (Srikanthan and McMahon 2001; Verdin et al. 2018) and can also underestimate interannual variability or frequencies of extremes (Maraun et al. 2010).

An alternative to use of GCMs or weather generators for obtaining future daily temperature and precipitation simulations was developed based on ARIMA forecasting. Detailed procedures for applying

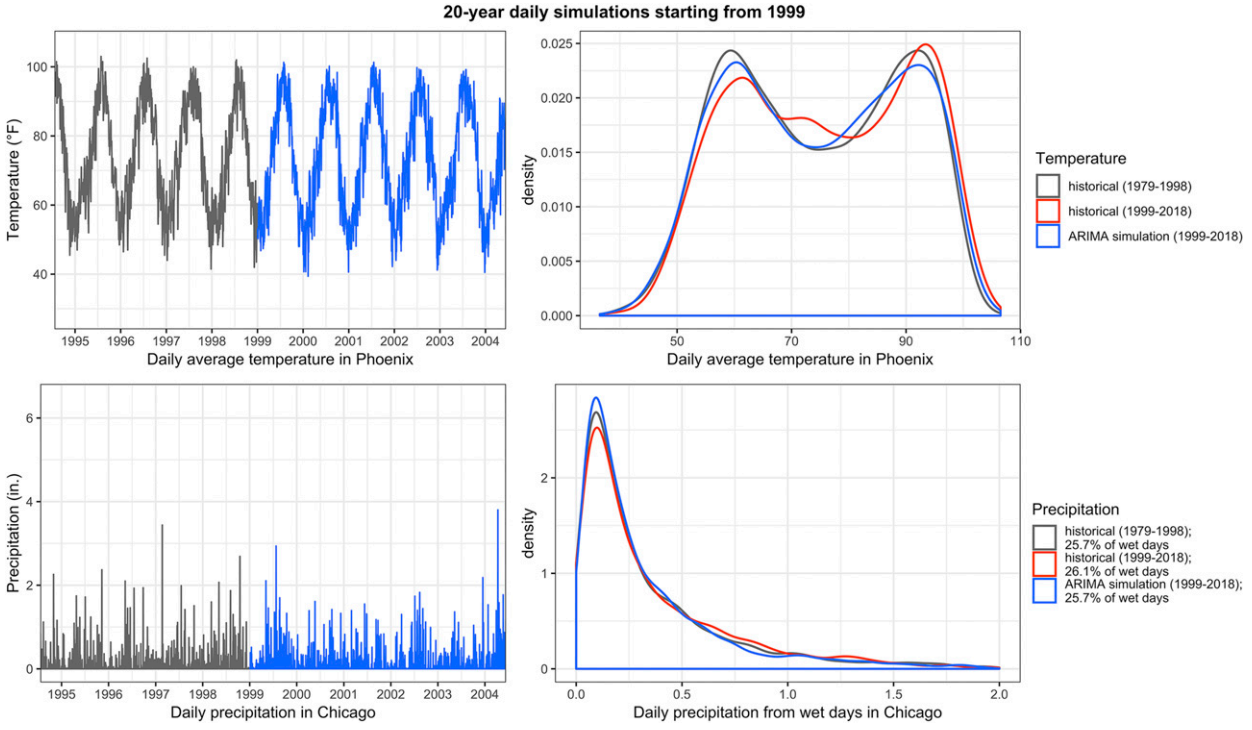

FIG. 4. Historical observations and the ARIMA simulations for (top) daily average temperature in Phoenix and (bottom) daily precipitation in Chicago. (left) Daily temperature and precipitation from historical observations and the daily simulations between 1995 and 2005. (right) Probability density functions of two 20-yr periods of historical observations (1979–98 and 1999–2018) and 20-yr ARIMA simulations (1999–2018). PDFs of daily precipitation from wet days (with precipitation amount more than 0.1 mm) are shown in the bottom right graph, while percentages of wet days for the three 20-yr periods can be found in the legend.

the ARIMA daily simulation are provided in supplemental material section D. The general procedure was to integrate several statistical techniques including decomposition, SARIMA, and bootstrap to simulate future daily events based on historical observations.

Examples of daily temperature simulations in Phoenix and daily precipitation simulations in Chicago obtained using the ARIMA-based simulation model are presented in Fig. 4. Observed daily temperature/precipitation records prior to 1999 were used to provide the ARIMA daily simulations for 1999–2018, as exhibited in Fig. 4. Probability density functions (PDFs) were further estimated for two 20-yr historical observations and the ARIMA simulations. Obtained ARIMA daily simulations, as exhibited in Fig. 4, show frequencies and magnitudes of daily temperature and precipitation events comparable to historical observations for the recent past 20 years (1979–98) and new observations for the next 20 years (1999–2018). Note that only one set of ARIMA 20-yr simulation results is presented in Fig. 4. If a greater number of simulation sets were included, the simulation results on aggregated annual scales would exhibit an envelope of potential future changes or uncertainties produced from the ARIMA forecasting model.

4. Evaluations of near-term regional temperature and precipitation projections

a. Comparisons of annual forecasts between the ARIMA and other statistical techniques

Several types of common statistical forecasting techniques were identified and utilized to assess the ARIMA forecasting accuracy with the location-specific annual average temperature, total precipitation, and extremes. The methods for comparison included a ‘‘baseline-all’’ utilizing all past historical observations, a ‘‘baseline-30’’ utilizing the recent 30 years of historical observations, a linear trend, the hinge-trend (Wilks and Livezey 2013), and a nonstationary GEV method. Before a more comprehensive comparison and quantitative evaluation between the ARIMA and other statistical techniques is presented, examples of the comparisons are presented in this section for data from Phoenix and Chicago.

Similar to the ARIMA forecasting, the compared statistical forecasting techniques are based on location-specific historical observations to obtain estimates of

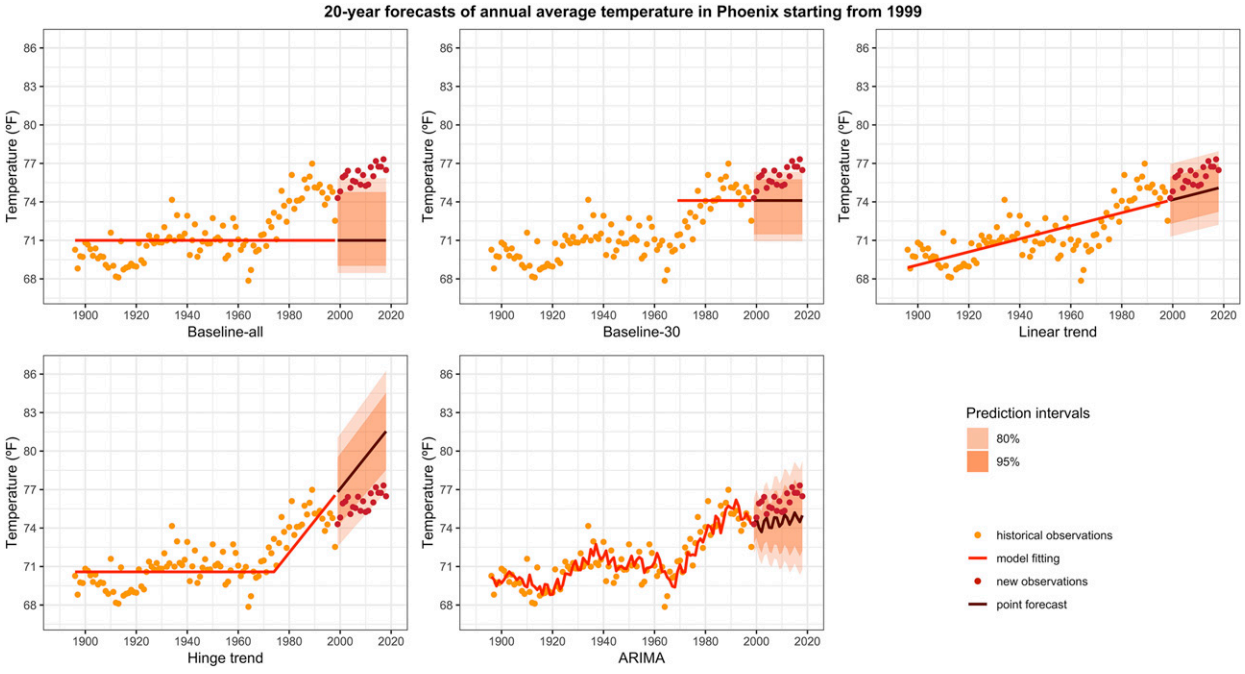

FIG. 5. The 20-yr projections (starting from 1999) of annual average temperature in Phoenix with baseline-all, baseline-30, linear trend, hinge-trend, and the ARIMA forecasting techniques. The lighter red points show the historical temperature observations and darker points are new observations (for validation). The model fitting and point forecasts are presented as lighter and darker lines, while the 80% and 95% prediction intervals are shown as the shaded area in the graphs. Historical records start from 1896.

future conditions. A "baseline-all" method assumes the future conditions remain as the same empirical distribution from all past historical observations. Point and interval forecasts for the baseline-all method were thus estimated as the median and corresponding intervals of the estimated empirical distribution from all historical observations. A "baseline-30" method, on the other hand, assumes the future conditions belong to the same empirical distribution from the recent 30 years of observations. Therefore, similar procedures from the baseline-all method were used in the baseline-30 method, with the exception that only the recent 30 years of observations were used to estimate the empirical distributions. In the linear trend method, point forecasts were obtained by extrapolating the historical trendline and interval forecasts were estimated from the sum of residuals from the linear fit. Instead of fitting a linear trend over all past observations, the hinge-trend method (Wilks and Livezey 2013)—an additional method to compare the forecasts of annual average temperature—assumes regional climate was stationary before 1975 and exhibits a linear trend after 1975. Finally, a nonstationary GEV method—assuming a linear trend in the location parameter for GEV distribution—was used to obtain and compare the forecasts in annual extremes.

A comparison of 20-yr projections (starting from 1999) of annual average temperature in Phoenix between the ARIMA and other forecasting techniques is provided in Fig. 5. Actual observations for the period 1999–2018 are presented in Fig. 5 for evaluation. The average temperature level for the 20-yr forecasts, as exhibited in Fig. 5, was better projected by the baseline-30, linear trend, and ARIMA forecasting methods, while both the linear trend and ARIMA point forecasts exhibited increasing trend similar to the observed trend for the forecast period 1999–2018. Interval forecasts from the ARIMA model, however, provided a slightly better estimation of future uncertainty compared to other techniques in this case.

A similar comparison of 20-yr projections of annual maximum 1-day precipitation in Chicago is provided in Fig. 6. While the point forecasts were similar among different methods in this case, the interval forecasts were notably different, as exhibited in Fig. 6. Because of right skewness of the maximum 1-day precipitation series in Chicago, all methods except the linear trend provided higher upper bounds for interval forecasts. Noticeably, with the integration of Box–Cox transformation, the heavy-tailed distribution exhibited in the precipitation observations, especially the high values observed in recent years, was better represented by the ARIMA prediction intervals. As the GEV is not ideal for estimation of precipitation extremes with greater return periods (DeGaetano and Zarrow 2011), the ARIMA

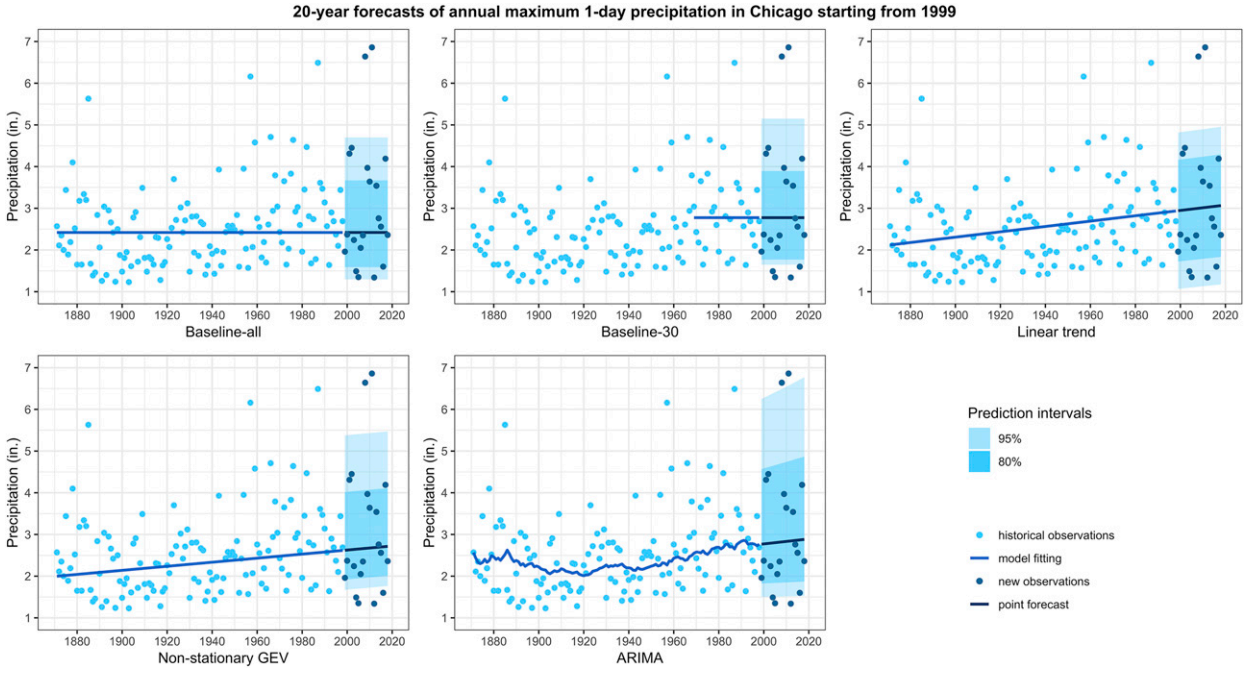

FIG. 6. The 20-yr projections (starting from 1999) of annual maximum 1-day precipitation in Chicago with baseline-all, baseline-30, linear trend, nonstationary GEV, and the ARIMA forecasting techniques. The lighter blue points show the historical precipitation observations and darker points are new observations (for validation). The model fitting and point forecasts are presented as lighter and darker lines, while the 80% and 95% prediction intervals are shown as the shaded area in the graphs. Historical records start from 1871.

model can potentially provide more accurate projections for precipitation extremes with greater return periods.

In addition, a further comparison of historical fitting between the ARIMA and linear trend methods was performed based on the forecasts in Figs. 5 and 6 and the results are presented in Fig. 7. Residuals from the ARIMA historical fitting of annual average temperature in Phoenix and annual maximum 1-day precipitation in Chicago are observably more evenly distributed than the linear trend method. Temporally, residuals from the ARIMA model fitting are more evenly distributed along the fitting line than other techniques as well, as exhibited in Figs. 5 and 6. Therefore, the temporally independent and normally distributed residuals produced from the ARIMA fitting are more statistically appropriate to facilitate the bootstrap technique used in daily temperature and precipitation simulations and estimation of confidence intervals for extremes with different return periods. As the estimation of these confidence intervals requires a consideration of additional model parameter uncertainty and is consequently more complex to be evaluated, a quantitative evaluation of confidence intervals for different return periods was not carried out in this study.

Comparisons between the ARIMA forecasting and downscaled GCM projections can be utilized to further assess the performance of the ARIMA model, as we have done in related work that will be published separately. Such comparisons can also provide insights into the effects of using downscaled GCM projections.

b. *Quantitative evaluation metrics for near-term regional climate forecasts*

A set of quantitative metrics were identified and utilized to evaluate the performance of near-term projections and compare the performance of the ARIMA model with other forecasting techniques. The evaluation of near-term projections followed a scheme of cross validation [i.e., separating historical observations into training or calibration period (for fitting data) and testing period (for evaluation)], similar to the results in Figs. 1–6 where historical observations were separated into observations used for forecasting and observations for evaluation. Such a process of forecast evaluation is also known as ''verification'' in meteorology studies and the quantitative metrics employed are often referred as forecast ''scores'' (Jolliffe and Stephenson 2003).

Different types of forecast scores were utilized to evaluate different aspects of forecasts, including mean absolute errors (MAE) and root-mean-square errors (RMSE) for evaluation of point forecasts and interval scores (SI) for evaluation of interval forecasts

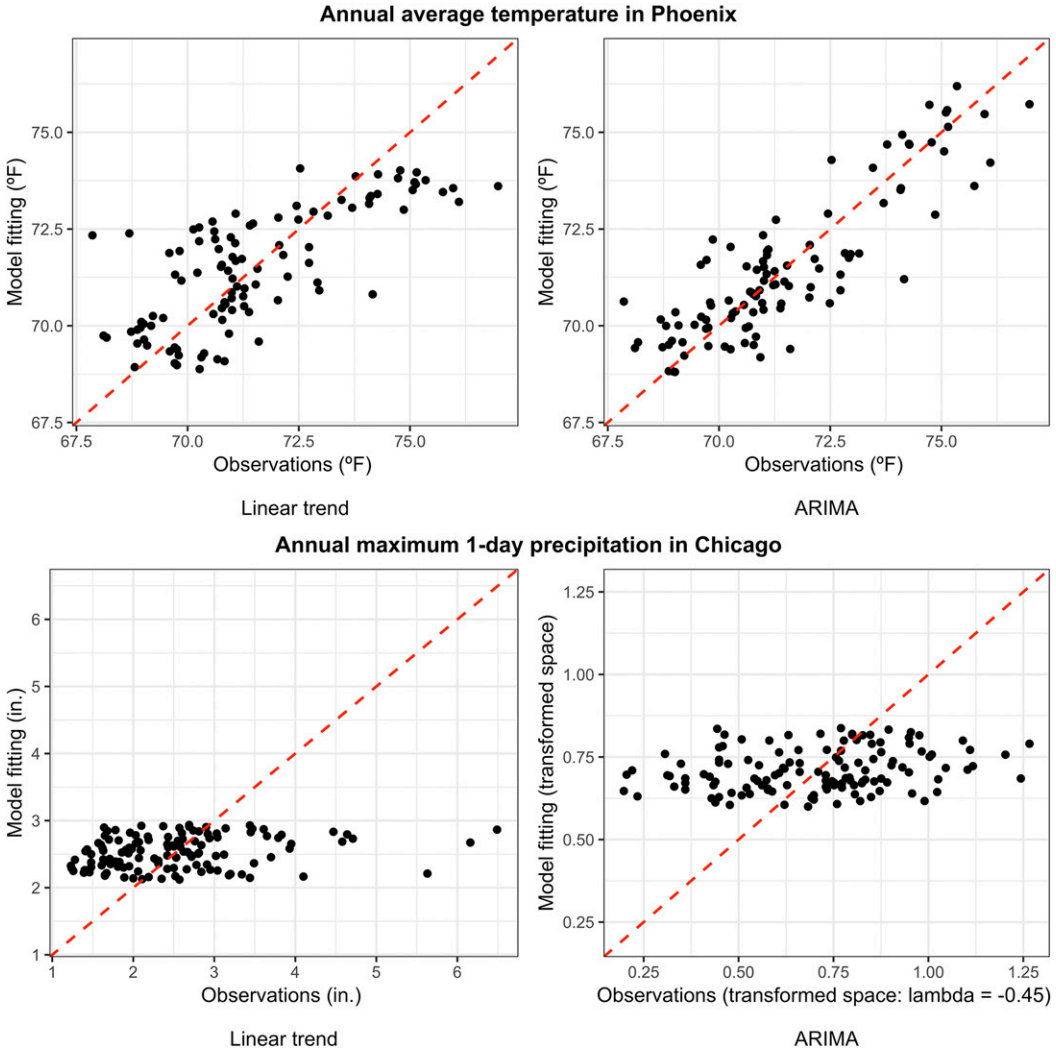

FIG. 7. The historical fitting from the linear trend and ARIMA model with respect to the historical observations of (top) annual average temperature in Phoenix and (bottom) annual maximum 1-day precipitation in Chicago. Model fitting was applied on climate records prior to 1999.

(Gneiting and Raftery 2007). Because interval forecasts are crucial in forecasting annual temperature and precipitation extremes as previously mentioned, intervals scores for both 80% prediction intervals (SI80) and 95% prediction intervals (SI95) were therefore used to assess different uncertainty levels. Further details of the calculation procedures for the forecast scores are provided in supplemental material section E. It is important to note that the forecast scores like RMSE are negatively correlated to the forecasting accuracy (i.e., the greater the forecast score, the less accurate the forecast).

c. *Results of the quantitative assessment on annual forecasts*

A systematic quantitative cross-validation process was carried out for annual average temperature, total precipitation, and their extremes at 93 U.S. cities. Specifically, the cross-validation process for individual time series of temperature and precipitation variables was applied with sets of moving forecasts—moving 5- and 20-yr projections (with 5 or 20 years of projection time) applied for each year starting from 1975 using the ARIMA and other previously discussed forecasting techniques. Each 5- or 20-yr forecast was compared with actual observations for that 5- or 20-yr period and different forecast scores were then calculated. A great quantity of forecast scores was then obtained with respect to different types of scores (MAE, RMSE, etc.), different climate variables, different forecasting techniques, different starting year for the forecasts, different lengths of forecasts (5 or 20 years), and for different cities.

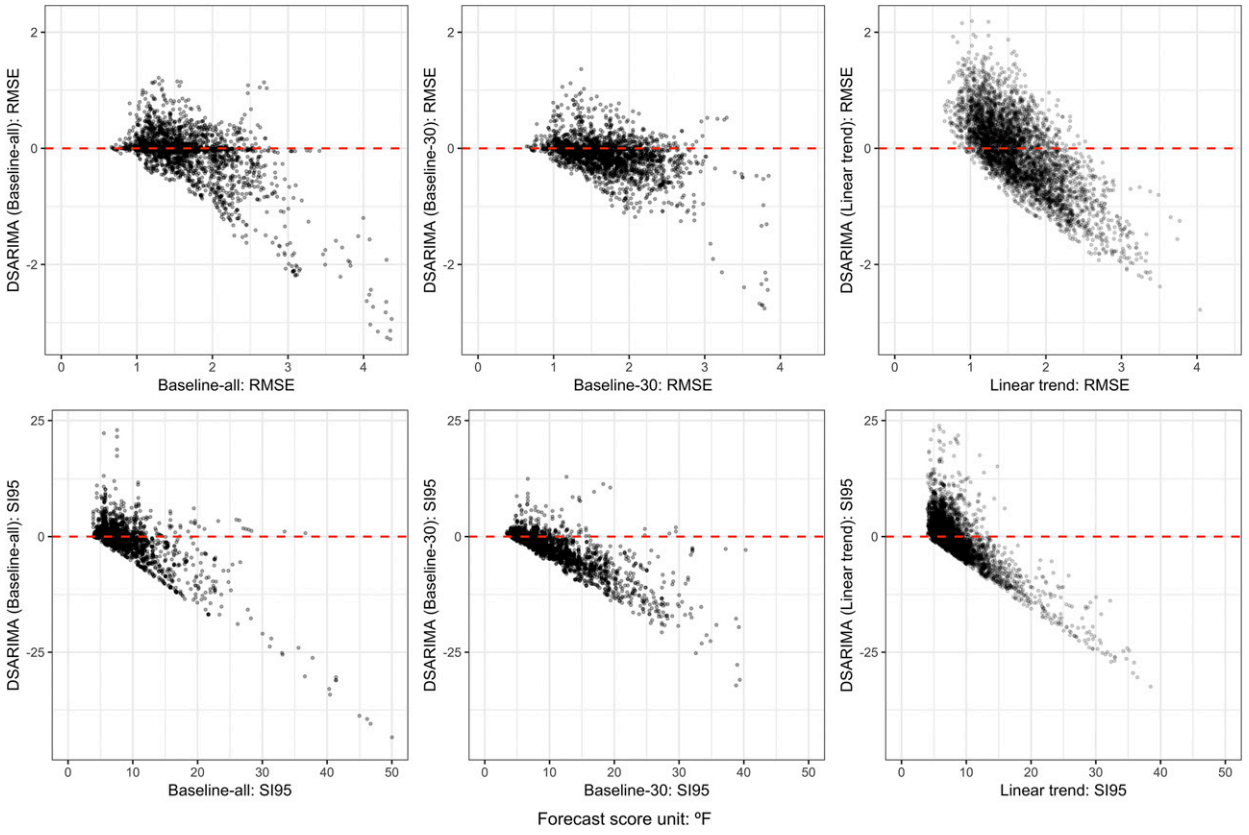

FIG. 8. Difference in forecast scores (RMSE and SI95; °F) for annual average temperature between the ARIMA and other three statistical techniques—baseline-all, baseline-30, and linear trend—with respect to the forecast scores of the other three methods. Forecast scores were produced from the moving 20-yr forecasts of annual average temperature in 93 U.S. cities starting each year from 1975.

To present the comparison results between the ARIMA and other forecasting techniques, the forecast score of the ARIMA model minus the forecast score of other forecasting techniques was calculated for each individual forecast and consequently produced the differences in the forecast scores (DS) between the ARIMA model and other methods (DSARIMA). If the DSARIMA values are negative (positive), the ARIMA model provided more (less) accurate projections.

DSARIMA values of RMSE and SI95 for 20-yr moving forecasts of annual average temperature and maximum 1-day precipitation across all different starting years and the 93 cities are presented in Figs. 8 and 9, as an example of the results. As DSARIMA values indicate the relative accuracy between the ARIMA and the compared forecasting method, Figs. 8 and 9 provide an assessment of the improvement by using the ARIMA model over the other techniques. The x axes in these figures exhibit the distributions of forecast scores from the compared forecasting method and y axes indicate the distributions of DSARIMA values (i.e., improvement from using the ARIMA model). For example, with respect to the 20-yr forecasts in annual average temperature, the ARIMA model provided slightly more accurate point forecasts and much improved interval forecasts than the baseline-all and baseline-30 methods. Additionally, the forecast scores of annual average temperature forecasts from the ARIMA and the linear trend methods are negatively correlated (i.e., one method can provide more accurate forecasts if the other one is insufficient). However, in the case of annual average temperature, more outliers of high forecast scores are observed for the linear trend method. Similar to Figs. 8 and 9, DSARIMA values of RMSE and SI95 for other climate variables are provided in supplemental material section F.

While the results for different forecasting techniques and for different types of forecast scores vary in Figs. 8 and 9, the DSARIMA values with respect to SI95 (i.e., for interval forecasts), demonstrate greater improvement from using the ARIMA model, while the DSARIMA values with respect to RMSE—for point forecasts—show moderate improvement or occasionally worse performance from the ARIMA model. As uncertainty bounds are often of greater interest for engineering applications,

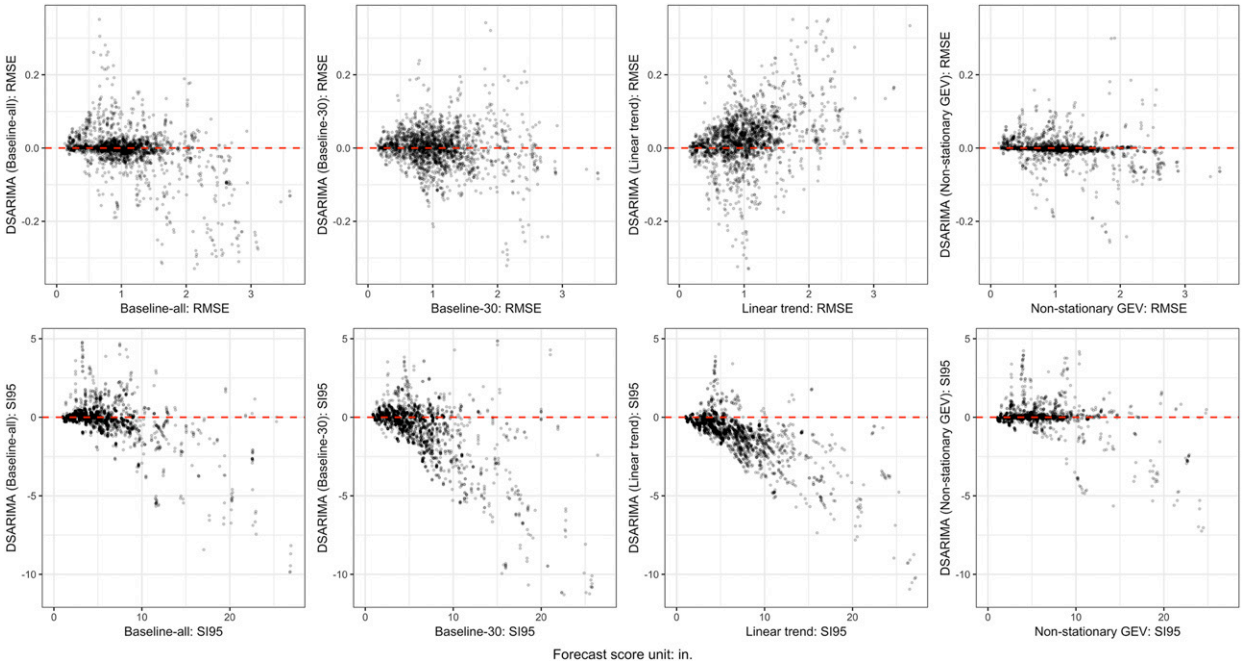

FIG. 9. Difference in forecast scores (RMSE and SI95; in.) for annual maximum 1-day precipitation projections between the ARIMA and other four statistical techniques—baseline-all, baseline-30, linear trend, and nonstationary GEV—with respect to the forecast scores of the other four methods. Forecast scores were produced from the moving 20-yr forecasts of annual maximum 1-day precipitation in 93 U.S. cities starting each year from 1975.

the ARIMA model, with better accuracy in interval forecasts, is potentially more favorable in providing near-term forecasts. Additionally, as exhibited in Figs. 7 and 8, when the compared forecasting techniques provide less accurate projections (with higher forecast scores), the DSARIMA value often further decreases below 0, especially for SI95. Therefore, the ARIMA forecasting model can provide substantially improved forecasts when the other forecasting methods cannot provide sufficiently accurate results.

Results for the DSARIMA were further aggregated to quantify the percentages of negative DSARIMA values (i.e., percentages of cases where the ARIMA model provided more accurate projections). Empirical distributions of DSARIMA values for different forecast score types and different climate variables (e.g., the individual graphs from Figs. 8 and 9) were utilized to estimate the percentages of negative values and the results are presented in Fig. 10. Percentages of negative DS values for other methods (e.g., scores of the linear trend minus scores of the baseline-all) were also calculated and results are provided in supplemental material section F.

As exhibited in Fig. 10, while the ARIMA forecasting model cannot outperform all other forecasting techniques considering different forecast scores, climate indices, and lengths of forecasts, the ARIMA model in general provide more accurate projections (especially in interval forecasts with 20-yr projection time) than other forecasting techniques (with more red cells than blue cells). Because of the unique characteristics of different forecasting techniques, a particular forecasting method can have unique advantages in providing near-term projections (e.g., the baseline-all and baseline-30 methods provide more accurate estimation of 5-yr 95% interval forecasts).

It is important to note that the finding from Figs. 8 and 9 (and other figures in supplemental material section F)—that the ARIMA model can substantially improve the forecasts when other forecasting methods are insufficient—was not well represented in the overall assessment in Fig. 10. In other words, although other common statistical forecasting methods can occasionally outperform the ARIMA model in forecasting particular climate variables, the ARIMA model can be more reliable (especially for interval forecasts) in reducing the number of cases where other forecasting models may generate large errors. For an engineering application that is limited in time and resources, a more reliable forecasting technique—the ARIMA forecasting method—represents a good option.

d. Evaluation of near-term daily temperature and precipitation simulations

A further evaluation of the ARIMA daily temperature and precipitation simulations was conducted and compared with the simulations provided from one weather

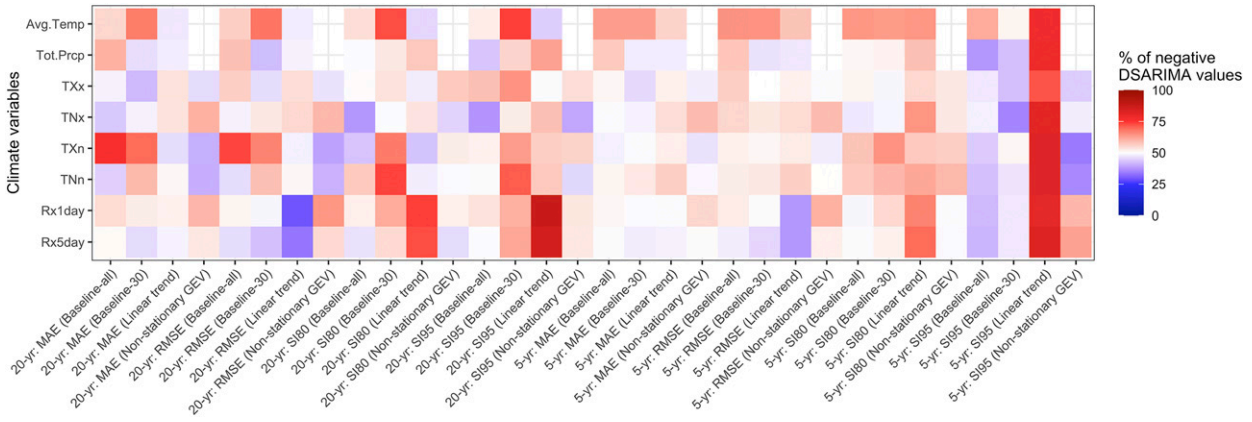

FIG. 10. Percentages of negative DSARIMA values (i.e., percentage of cases where the ARIMA model provided more accurate projections) for the 5- and 20-yr moving forecasts (starting from 1975) of eight annual climate indices with four forecast scores (MAE, RMSE, SI80, and SI95). The eight annual climate indices are: average temperature (Avg.Temp), total precipitation (Tot.Prcp), warmest daily T_{max} (TXx) and daily T_{min} (TNx), coldest daily T_{max} (TXn) and T_{min} (TNn), maximum 1-day precipitation (Rx1day), and maximum consecutive 5-day precipitation (Rx5day). Cells are marked as red (blue) if the percentages of negative DSARIMA values are greater (less) than 50%.

generator. The simulations of daily average temperatures in Phoenix and daily precipitation in Chicago from the ARIMA model and a weather generator were compared with the historical observations, the same example datasets used for other comparisons in this study.

PDFs of daily simulations and aggregated annual indices from the daily simulations are presented in Fig. 11 for temperature in Phoenix and precipitation in Chicago. Three sets of daily simulations were obtained from the ARIMA simulation model and Version 6 of the Long Ashton Research Station Weather Generator (LARS-WG) (Semenov and Barrow 2002) for the period 1999–2018, based on the historical observations in the two cities prior to 1999.

The resulting PDFs suggested that the ARIMA simulation model can provide daily temperature and precipitation simulations comparable to historical observations and simulations obtained from the LARS-WG, while the aggregated annual temperature from the daily simulations indicate that the ARIMA simulations seem to be more accurate in representing the observations from the recent and forecasting periods than the LARS-WG. As the LARS-WG estimates an average temperature level from all the past observations (prior to 1999) and no underlying trend was assumed, the average temperature level was consequently similar to a baseline-all estimation. While only three sets of simulations are provided in Fig. 11, the aggregated annual temperatures from the ARIMA simulations exhibit a similar average level to the ARIMA annual forecasts in Figs. 1 and 3. If more sets of simulations were provided, annual indices aggregated from the daily simulations can exhibit an envelope of different realizations and would be similar to the interval forecasts in Figs. 1 and 3. In addition, as exhibited in the aggregated annual temperatures, the LARS-WG simulations provided less realistic interannual variations than the ARIMA simulations.

Although further analyses may be needed to evaluate the usefulness and applicability of the ARIMA daily simulations in particular engineering applications, daily simulations obtained from the ARIMA model in this study exhibited good performance in representing the interannual variations and in integrating near-term climate forecasts. The ARIMA model provides an alternative to GCMs (with downscaling) and weather generators for considering near-term regional climate change in using engineering process models.

5. Summary and conclusions

A data-driven statistical approach was developed to obtain near-term (2–20 years), location-specific temperature and precipitation forecasts based on local historical observations for engineering applications. As individual time series of temperature and precipitation variables have unique characteristics of temporal correlation and skewed distribution, an ARIMA-based time series forecasting model was selected and utilized. To address a specific need for many engineering applications, the ARIMA-based forecasting model was extended to provide estimates of confidence intervals for temperature and precipitation extremes in different return periods, and to provide future daily temperature and precipitation simulations. (The developed ARIMA-based forecasting techniques are available as an open

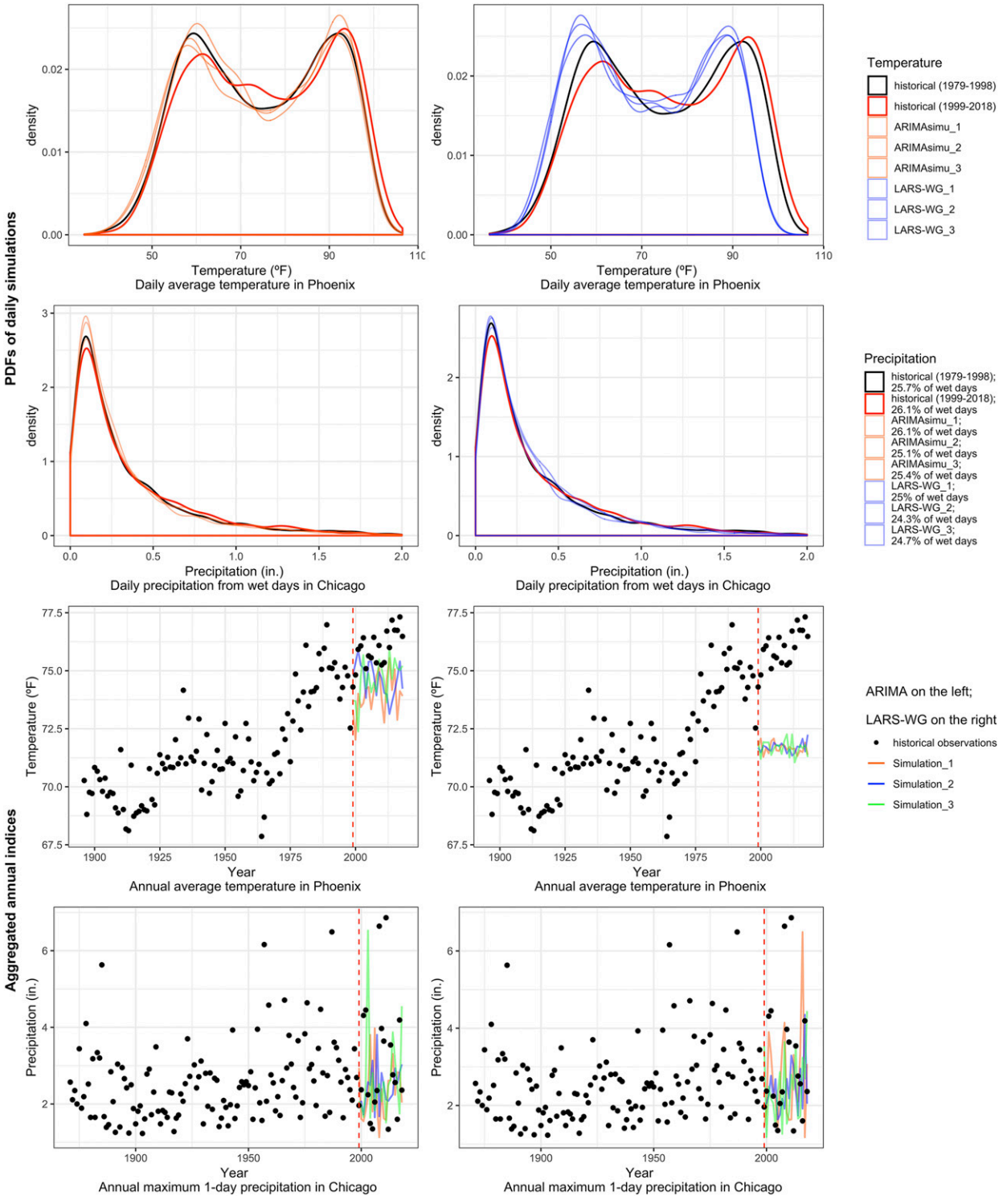

FIG. 11. Comparisons of the (left) ARIMA and (right) LARS-WG simulations of daily average temperature in Phoenix and daily precipitation in Chicago. Both the ARIMA and LARS-WG were fitted with the historical daily observations prior to 1999 to provide simulations for 1999–2018. Three separate sets of simulations are provided. PDFs of two 20-yr periods of historical observations (1979–98 and 1999–2018), 20-yr ARIMA simulations (1999–2018), and 20-yr LARS-WG simulations (1999–2018) are presented in the top two rows. Annual average temperature in Phoenix and maximum 1-day precipitation in Chicago—aggregated from the daily simulations—are presented in the bottom two rows.

source R package at: https://github.com/yuchuan-lai/scifi.) A quantitative assessment of the ARIMA model annual forecasting accuracy was performed by comparing the ARIMA model results with those of other common statistical techniques such as a linear trend method.

Evaluations of examples of the ARIMA projections suggested that the ARIMA point forecasts are mainly determined by the average of recent observations and historical long-term trend, while the interval forecasts increase with a longer projection time, corresponding to expanding uncertainty in the more distant future. In addition, with integration of the Box–Cox transformation technique for the time series, the ARIMA model can be used to obtain forecasts with climate variables exhibiting skewed distributions—especially annual extreme series—and consequently the ARIMA interval forecasts provide higher upper-uncertainty bounds.

Quantitative comparisons between the ARIMA model results and those of other statistical forecasting techniques were carried out for annual average temperature, total precipitation, and six extreme indices for 93 U.S. cities by obtaining moving 5- and 20-yr forecasts and using the cross-validation technique. Results suggested that the ARIMA model generally provides more accurate projections, especially with interval forecasts, than other forecasting techniques, although the ARIMA model did not outperform all other techniques for all evaluated climate indices. Because the ARIMA model generally increases the forecasting accuracy compared to other forecasting methods (especially in providing interval forecasts), the ARIMA model is a reliable approach for obtaining near-term regional temperature and precipitation forecasts if not necessarily a more accurate forecasting technique for all cases considering various locations, climate variables, and evaluation metrics. The ARIMA model will be useful for many engineering applications where general forecasting reliability and quantified future uncertainty bounds, especially in annual extremes, are highly valued.

The ARIMA annual forecasting model was further developed to convert the annual forecasts into desired formats commonly used in engineering applications, including estimation of confidence intervals for different return periods when forecasting extremes, and simulations of daily temperature and precipitation. The estimation of confidence intervals can help with assessments of historical and future changes for different return periods; the confidence intervals can be used directly as climatic design values. The ARIMA simulation model can provide daily temperature and precipitation simulations with PDFs comparable to historical observations and results from weather generators, while carrying forward the long-term historical annual trend and providing more realistic interannual variations than a common weather generator.

The ARIMA-based forecasting model is an efficient, interpretable, and reliable method for obtaining near-term (2–20 years) regional temperature and precipitation forecasts for use in various engineering applications. Given the current challenges and uncertainties of utilizing GCM projections and downscaling techniques in engineering applications, the ARIMA-based statistical time series forecasting model—together with techniques to estimate confidence intervals for return periods and to simulate future daily temperature and precipitation—provides an alternative approach for obtaining near-term regional temperature and precipitation information critical for civil and environmental engineering applications.

Acknowledgments. The research was supported by a Carnegie Mellon College of Engineering Dean's Fellowship to Yuchuan Lai, and by the Hamerschlag Chair of Professor Dzombak.

REFERENCES

Abatzoglou, J. T., and T. J. Brown, 2012: A comparison of statistical downscaling methods suited for wildfire applications. *Int. J. Climatol.*, **32**, 772–780, https://doi.org/10.1002/joc.2312.

AghaKouchak, A., D. Easterling, K. Hsu, S. Schubert, and S. Sorooshian, 2013: *Extremes in a Changing Climate: Detection, Analysis and Uncertainty*. Springer, 423 pp.

Auld, H., J. Waller, S. Eng, J. Klaassen, R. Morris, S. Fernandez, V. Cheng, and D. MacIver, 2010: The changing climate and national building codes and standards. *Ninth Symp. on the Urban Environment*, Keystone, CO, Amer. Meteor. Soc., 5.6, https://ams.confex.com/ams/19Ag19BLT9Urban/techprogram/paper_174517.htm.

Bonnin, G. M., D. Martin, B. Lin, T. Parzybok, M. Yekta, and D. Riley, 2006: *Precipitation-Frequency Atlas of the United States*. NOAA Atlas 14, Vol. 3, 65 pp.

Box, G. E. P., and D. R. Cox, 1964: An analysis of transformations. *J. Roy. Stat. Soc.*, **26B**, 211–243, https://doi.org/10.1111/j.2517-6161.1964.tb00553.x.

——, and G. Jenkins, 1970: *Time Series Analysis: Forecasting and Control*. Holdan-Day, 709 pp.

Chapra, S. C., and Coauthors, 2017: Climate change impacts on harmful algal blooms in U.S. freshwaters: A screening-level assessment. *Environ. Sci. Technol.*, **51**, 8933–8943, https://doi.org/10.1021/acs.est.7b01498.

Chen, J., F. P. Brissette, and R. Leconte, 2010: A daily stochastic weather generator for preserving low-frequency of climate variability. *J. Hydrol.*, **388**, 480–490, https://doi.org/10.1016/j.jhydrol.2010.05.032.

Cheng, L., A. AghaKouchak, E. Gilleland, and R. W. Katz, 2014: Non-stationary extreme value analysis in a changing climate. *Climatic Change*, **127**, 353–369, https://doi.org/10.1007/s10584-014-1254-5.

CIDA, 2019: CIDA USGS THREDDS catalog. Accessed 23 November 2019, https://cida.usgs.gov/thredds/catalog.html.

Cook, L. M., C. J. Anderson, and C. Samaras, 2017: Framework for incorporating downscaled climate output into existing engineering methods: Application to precipitation frequency curves. *J. Infrastruct. Syst.*, **23**, 04017027, https://doi.org/10.1061/(ASCE)IS.1943-555X.0000382.

CPC, 2019a: Three months outlooks—Official forecasts. Accessed 24 November 2019, https://www.cpc.ncep.noaa.gov/products/predictions/90day/.

——, 2019b: The North American multi-model ensemble. Accessed 24 November 2019, https://www.cpc.ncep.noaa.gov/products/NMME/.

DeGaetano, A., and D. Zarrow, 2011: Extreme precipitation in New York & New England: An interactive web tool for extreme precipitation analysis. Technical documentation and user manual, 93 pp., http://www.precip.net.

——, W. Noon, and K. L. Eggleston, 2015: Efficient access to climate products using ACIS web services. *Bull. Amer. Meteor. Soc.*, **96**, 173–180, https://doi.org/10.1175/BAMS-D-13-00032.1.

ESTCP, 2017: Nonstationary weather patterns and extreme events: Informing design and planning for long-lived infrastructure. ESTCP Tech. Rep., ESTCP Project RC-201591, 66 pp.

Eyring, V., S. Bony, G. A. Meehl, C. A. Senior, B. Stevens, R. J. Stouffer, and K. E. Taylor, 2016: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geosci. Model Dev.*, **9**, 1937–1958, https://doi.org/10.5194/gmd-9-1937-2016.

FHWA, 2016: Temperature and precipitation impacts to pavements on expansive soils: Proposed State Highway 170 in North Texas. TEACR Eng. Assess., FHWA-HEP-17-018, 77 pp.

——, 2017: Wildfire and precipitation impacts to a culvert: US 34 at Canyon Cove Lane, Colorado. TEACR Eng. Assess., FHWA-HEP-18-021, 77 pp.

Franzke, C., 2012: Nonlinear trends, long-range dependence, and climate noise properties of surface temperature. *J. Climate*, **25**, 4172–4183, https://doi.org/10.1175/JCLI-D-11-00293.1.

Giorgi, F., and W. J. Gutowski, 2015: Regional dynamical downscaling and the CORDEX initiative. *Annu. Rev. Environ. Resour.*, **40**, 467–490, https://doi.org/10.1146/annurev-environ-102014-021217.

Gneiting, T., and A. E. Raftery, 2007: Strictly proper scoring rules, prediction, and estimation. *J. Amer. Stat. Assoc.*, **102**, 359–378, https://doi.org/10.1198/016214506000001437.

Hallegatte, S., 2009: Strategies to adapt to an uncertain climate change. *Global Environ. Change*, **19**, 240–247, https://doi.org/10.1016/j.gloenvcha.2008.12.003.

Hartmann, D. L., and Coauthors, 2013: Observations: Atmosphere and surface. *Climate Change 2013: The Physical Science Basis*, T. F. Stocker et al., Eds., Cambridge University Press, 159–254.

Hewitson, B., and R. Crane, 1996: Climate downscaling: Techniques and application. *Climate Res.*, **7**, 85–95, https://doi.org/10.3354/cr007085.

Hosseini, M., F. Tardy, and B. Lee, 2018: Cooling and heating energy performance of a building with a variety of roof designs; the effects of future weather data in a cold climate. *J. Build. Eng.*, **17**, 107–114, https://doi.org/10.1016/j.jobe.2018.02.001.

Houser, T., and Coauthors, 2017: Estimating economic damage from climate change in the United States. *Science*, **356**, 1362–1369, https://doi.org/10.1126/science.aal4369.

Hyman, R., R. Kafalenos, B. Beucler, and M. Culp, 2014: Assessment of key gaps in the integration of climate change considerations into transportation engineering. Accessed 7 October 2018, https://trid.trb.org/view/1395309.

Hyndman, R. J., and G. Athanasopoulos, 2018: *Forecasting: Principles and Practice*. OTexts, 382 pp.

——, and Coauthors, 2018: forecast: Forecasting functions for time series and linear models, version 8.3. R Package, accessed 19 March 2019, https://researchportal.bath.ac.uk/en/publications/forecast-forecasting-functions-for-time-series-and-linear-models.

Jolliffe, I. T., and D. B. Stephenson, 2003: *Forecast Verification: A Practitioner's Guide in Atmospheric Science*. Wiley, 254 pp.

Kaushik, I., and S. Singh, 2008: Seasonal ARIMA model for forecasting of monthly rainfall and temperature. *J. Environ. Res. Dev.*, **3**, 506–514.

Kilsby, C. G., and Coauthors, 2007: A daily weather generator for use in climate change studies. *Environ. Modell. Software*, **22**, 1705–1719, https://doi.org/10.1016/j.envsoft.2007.02.005.

Kirtman, B. P., and Coauthors, 2014: The North American Multimodel Ensemble: Phase-1 seasonal-to-interannual prediction; phase-2 toward developing intraseasonal prediction. *Bull. Amer. Meteor. Soc.*, **95**, 585–601, https://doi.org/10.1175/BAMS-D-12-00050.1.

Krakauer, N. Y., and B. M. Fekete, 2014: Are climate model simulations useful for forecasting precipitation trends? Hindcast and synthetic-data experiments. *Environ. Res. Lett.*, **9**, 024009, https://doi.org/10.1088/1748-9326/9/2/024009.

Kunkel, K. E., D. R. Easterling, K. Redmond, and K. Hubbard, 2003: Temporal variations of extreme precipitation events in the United States: 1895-2000. *Geophys. Res. Lett.*, **30**, 1900, https://doi.org/10.1029/2003GL018052.

Lai, Y., and D. A. Dzombak, 2019: Use of historical data to assess regional climate change. *J. Climate*, **32**, 4299–4320, https://doi.org/10.1175/JCLI-D-18-0630.1.

Larsen, P., and Coauthors, 2016: Climate change damages to Alaska public infrastructure and the economics of proactive adaptation. *Proc. Natl. Acad. Sci. USA*, **114**, E122–E131, https://doi.org/10.1073/pnas.1611056113.

Leeper, R. D., J. Rennie, and M. A. Palecki, 2015: Observational perspectives from U.S. Climate Reference Network (USCRN) and Cooperative Observer Program (COOP) Network: Temperature and precipitation comparison. *J. Atmos. Oceanic Technol.*, **32**, 703–721, https://doi.org/10.1175/JTECH-D-14-00172.1.

Mahsin, M., Y. Akhter, and M. Begum, 2012: Modeling rainfall in Dhaka division of Bangladesh using time series. *J. Math. Model. Appl.*, **1**, 67–73.

Maraun, D., S. Brienen, H. W. Rust, T. Sauter, M. Themeßl, V. K. C. Venema, and K. P. Chun, 2010: Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user. *Rev. Geophys.*, **48**, RG3003, https://doi.org/10.1029/2009RG000314.

Maurer, E. P., H. G. Hidalgo, T. Das, M. D. Dettinger, and D. R. Cayan, 2010: The utility of daily large-scale climate data in the assessment of climate change impacts on daily streamflow in California. *Hydrol. Earth Syst. Sci.*, **14**, 1125–1138, https://doi.org/10.5194/hess-14-1125-2010.

Mearns, L. O., W. J. Gutowski, R. Jones, L.-Y. Leung, S. McGinnis, A. M. B. Nunes, and Y. Qian, 2009: A regional climate change assessment program. *Eos, Trans. Amer. Geophys. Union*, **90**, 311, https://doi.org/10.1029/2009EO360002.

Meehl, G. A., and Coauthors, 2009: Decadal prediction: Can it be skillful? *Bull. Amer. Meteor. Soc.*, **90**, 1467–1486, https://doi.org/10.1175/2009BAMS2778.1.

——, J. M. Arblaster, and G. Branstator, 2012: Mechanisms contributing to the warming hole and the consequent U.S. east–west differential of heat extremes. *J. Climate*, **25**, 6394–6408, https://doi.org/10.1175/JCLI-D-11-00655.1.

Montgomery, D. C., C. L. Jennings, and M. Kulahci, 2016: *Introduction to Time Series Analysis and Forecasting*. JohnWiley & Sons, 627 pp.

NCAR, 2019: Climate data at the National Center for Atmospheric Research. Accessed 23 November 2019, https://www.earthsystemgrid.org/.

Olsen, J. R., 2015: *Adapting Infrastructure and Civil Engineering Practice to a Changing Climate*. American Society of Civil Engineers, 93 pp.

O'Lenic, E. A., D. A. Unger, M. S. Halpert, and K. S. Pelman, 2008: Developments in operational long-range climate prediction at CPC. *Wea. Forecasting*, **23**, 496–515, https://doi.org/10.1175/2007WAF2007042.1.

Pascual, L., J. Romo, and R. Esther, 2004: Bootstrap predictive inference for ARIMA processes. *J. Time Ser. Anal.*, **25**, 449–465, https://doi.org/10.1111/j.1467-9892.2004.01713.x.

——, ——, and E. Ruiz, 2005: Bootstrap prediction intervals for power-transformed time series. *Int. J. Forecasting*, **21**, 219–235, https://doi.org/10.1016/j.ijforecast.2004.09.006.

Peterson, T. C., W. M. Connolley, and J. Fleck, 2008: The myth of the 1970s global cooling scientific consensus. *Bull. Amer. Meteor. Soc.*, **89**, 1325–1337, https://doi.org/10.1175/2008BAMS2370.1.

Pierce, D. W., D. R. Cayan, and B. L. Thrasher, 2014: Statistical downscaling using Localized Constructed Analogs (LOCA). *J. Hydrometeor.*, **15**, 2558–2585, https://doi.org/10.1175/JHM-D-14-0082.1.

Semenov, M. A., and E. M. Barrow, 2002: LARS-WG: A Stochastic Weather Generator for Use in Climate Impact Studies, User Manual. Rothamstead Research, 27 pp.

SERDP, 2015: Providing useful climate information at moderate time scales: Proof of concept. SERDP, 3 pp., https://www.serdp-estcp.org/Funding-Opportunities/SERDP-Solicitations/Past-Solicitation-Pages/FY-2017-Files/SEED-SONs-FY17/RCSEED-17-01.

Srikanthan, R., and T. A. McMahon, 2001: Stochastic generation of annual, monthly and daily climate data: A review. *Hydrol. Earth Syst. Sci.*, **5**, 653–670, https://doi.org/10.5194/hess-5-653-2001.

Stouffer, R. J., V. Eyring, G. A. Meehl, S. Bony, C. Senior, B. Stevens, and K. E. Taylor, 2017: CMIP5 scientific gaps and recommendations for CMIP6. *Bull. Amer. Meteor. Soc.*, **98**, 95–105, https://doi.org/10.1175/BAMS-D-15-00013.1.

Taylor, J. W., and R. Buizza, 2004: A comparison of temperature density forecasts from GARCH and atmospheric models. *J. Forecasting*, **23**, 337–355, https://doi.org/10.1002/for.917.

Trzaska, S., and E. Schnarr, 2014: *A Review of Downscaling Methods for Climate Change projections: African and Latin American Resilience to Climate Change (ARCC)*. Tetra Tech ARD, 42 pp.

Van Vliet, M. T. H., D. Wiberg, S. Leduc, and K. Riahi, 2016: Power-generation system vulnerability and adaptation to changes in climate and water resources. *Nat. Climate Change*, **6**, 375–380, https://doi.org/10.1038/nclimate2903.

Verdin, A., B. Rajagopalan, W. Kleiber, G. Podestá, and F. Bert, 2018: A conditional stochastic weather generator for seasonal to multi-decadal simulations. *J. Hydrol.*, **556**, 835–846, https://doi.org/10.1016/j.jhydrol.2015.12.036.

Vogel, J., E. McNie, and D. Behar, 2016: Co-producing actionable science for water utilities. *Climate Serv.*, **2–3**, 30–40, https://doi.org/10.1016/j.cliser.2016.06.003.

Wang, W.-C., K.-W. Chau, D.-M. Xu, and X.-Y. Chen,2015: Improving forecasting accuracy of annual runoff time series using ARIMA based on EEMD decomposition. *Water Resour. Manage.*, **29**, 2655–2675, https://doi.org/10.1007/s11269-015-0962-6.

Wilks, D. S., 1993: Comparison of three-parameter probability distributions for representing annual extreme and partial duration precipitation series. *Water Resour. Res.*, **29**, 3543–3549, https://doi.org/10.1029/93WR01710.

——, 2006: *Statistical Methods in the Atmospheric Sciences*. 2nd ed. International Geophysics Series, Vol. 100, Academic Press, 648 pp.

——, and R. E. Livezey, 2013: Performance of alternative ''normals'' for tracking climate changes, using homogenized and nonhomogenized seasonal U.S. surface temperatures. *J. Appl. Meteor. Climatol.*, **52**, 1677–1687, https://doi.org/10.1175/JAMC-D-13-026.1.

Wood, A. W., L. R. Leung, V. Sridhar, and D. P. Lettenmaier, 2004: Hydrologic implications of dynamical and statistical approaches to downscaling climate model outputs. *Climatic Change*, **62**, 189–216, https://doi.org/10.1023/B:CLIM.0000013685.99609.9e.

Yusof, F., and I. L. Kane, 2013: Volatility modeling of rainfall time series. *Theor. Appl. Climatol.*, **113**, 247–258, https://doi.org/10.1007/s00704-012-0778-8.