# Urbanization Effects on Estimates of Global Trends in Mean and Extreme Air Temperature

PANFENG ZHANG,<sup>a</sup> Guoyu Ren,<sup>a,b</sup> Yun Qin,<sup>a</sup> Yaqian Zhai,<sup>c</sup> Tianlin Zhai,<sup>d</sup> Suonam Kealdrup Tysa,<sup>a</sup> Xiaoying Xue,<sup>a</sup> Guowei Yang,<sup>a</sup> and Xiubao Sun<sup>e</sup>

<sup>a</sup> Department of Atmospheric Science, School of Environmental Studies, China University of Geosciences, Wuhan, China <sup>b</sup> Laboratory for Climate Studies, National Climate Center, China Meteorological Administration, Beijing, China

 $^{\circ}$  School of Geography and Information Engineering, China University of Geosciences, Wuhan, China

<sup>d</sup> School of Resource and Environmental Sciences, Wuhan University, Wuhan, China

<sup>e</sup> South China Sea Institute of Oceanology, Chinese Academy of Sciences, Guangzhou, China

(Manuscript received 28 May 2020, in final form 2 December 2020)

ABSTRACT: Identifying and separating the signal of urbanization effects in current temperature data series is essential for accurately detecting, attributing, and projecting mean and extreme temperature change on varied spatial scales. This paper proposes a new method based on machine learning to classify the observational stations into rural stations and urban stations. Based on the classification of rural and urban stations, the global and regional land annual mean and extreme temperature indices series over 1951–2018 for all stations and rural stations were calculated, and the urbanization effects and the urbanization contribution of global land annual mean and extreme temperature indices series are quantitatively evaluated using the difference series between all stations and the rural stations. The results showed that the global land annual mean time series for mean temperature and most extreme temperature indices series generally occurred after the mid-1980s, and there were significant differences of the magnitudes of urbanization effects among different regions. The urbanization effects of annual mean and extreme temperature indices series in East Asia is generally the strongest, which is consistent with the rapidly urbanization process in the region over the past decades, but it is generally small in Europe during the recent decades.

KEYWORDS: Climate change; Climate variability; Surface temperature; Algorithms; Trends; Machine learning

#### 1. Introduction

Climate change has a significant impact on natural and human systems (Handmer et al. 2012). Mean and extreme temperature indices (temperature indices hereafter) are important indicators in monitoring and detecting climate change. However, many regions in the world have experienced a rapid urbanization process in the past decades to century, which has led to the increasing of urban areas and the strengthening of the urban heat island (UHI) effect in urban areas or suburbs. Many of the land observational stations are located in the urban areas or the suburbs because of the expansion of the cities. However, urban areas accounted for only 0.39% of global land areas based on a global land-use/land-cover (LULC) product in the year 2018 (Hollmann et al. 2013). Therefore, the instrument surface air temperature data may have a systematic bias caused by urbanization around the observational sites in many regions over the past decades to century, and yet they are used to analyze regional- to global-scale changes in climate (Peterson 2003; Kalnay and Cai 2003; Zhou et al. 2004; Ren et al. 2008; Parker 2010; Hansen et al. 2001, 1999, 2010; Tysa et al. 2019). Identifying and separating the signal of urbanization effects in the current

temperature data series is essential for accurately detecting, attributing, and projecting mean and extreme temperature change on varied spatial scales (Ren and Zhou 2014).

In the past decades, many global and regional efforts have been involved in the study of urbanization effects on the mean temperature series, and most of them have shown that the urbanization effects indisputably exist in the currently observed temperature series (Karl et al. 1988; Peterson et al. 1999; Kalnay and Cai 2003; Zhou et al. 2004; Ren et al. 2008; Fujibe 2009; Yan et al. 2010; Hu et al. 2010; Ren and Ren 2011; Yang et al. 2011; Das et al. 2011; Tysa et al. 2019; Wen et al. 2019). The question is to what extent the urbanization signal exists in the global and regional mean temperature series. At the global scale, the IPCC Fifth Assessment Report argued that the uncorrected urbanization influences contribute no more than 10% to the centennial global land averaged temperature trends (Hartmann et al. 2013); at the regional scale, most research effort was focused on China, with the estimated urbanization effects somehow different because of the distinct methods, datasets, study areas, and study periods used (Ren et al. 2008; Ren and Ren 2011; Hu et al. 2010; Yang et al. 2011; Yan et al. 2010), in spite of the fact that most studies applying sophisticated methods to determine reference stations showed a large and significant urbanization contribution to the estimated warming trends (Tysa et al. 2019; Wen et al. 2019).

Meanwhile, there are only a few single station or regional studies of urbanization effects on extreme temperature changes, mostly conducted for China's mainland (e.g., L. Zhang et al.

Supplemental information related to this paper is available at the Journals Online website: https://doi.org/10.1175/JCLI-D-20-0389.s1.

Corresponding author: Guoyu Ren, guoyoo@cma.gov.cn

2011; Zhou and Ren 2011; Li et al. 2014; Ren and Zhou 2014; Bian et al. 2015; Yang et al. 2017; Sun et al. 2019; Zhao et al. 2019). The studies showed that the urbanization effect on the extreme temperature change could not be neglected. Ren and Zhou (2014) made a comprehensive investigation of the urbanization effect in the country-averaged extreme temperature indices series for China's mainland, and found a significant urbanization contribution (from 5.7% to 58.1%) to the estimates of their trends in the last decades. Yang et al. (2017) showed that the urbanization impact on extreme hot events change in the urban agglomeration of east China is even comparable to the impact from greenhouse gases.

How to select the representative rural stations as reference sites is a key to the studies of urbanization effect on trends of surface air temperature series (Ren et al. 2008). There are several methods to select rural or reference stations. They include 1) methods based on the population in residential sites where the station is located (Karl et al. 1988; Peterson et al. 1999; Ren et al. 2008); 2) methods based on artificial nighttime illumination as measured by satellites or land surface temperature data around the stations (Peterson et al. 1999; Hansen et al. 2001; Peterson 2003; Hansen et al. 2010; Ren and Ren 2011); 3) methods developed using LULC data derived from satellite remote sensing (Gallo et al. 1996; Kalnay and Cai 2003; He et al. 2007; Ge et al. 2007; Wang and Ge 2012; Patra et al. 2018; Tysa et al. 2019); and 4) a comprehensive procedure that considers the numbers and distance of relocation, station density, population around the city or towns, percentage of urban areas, and the straight distance between the stations and the center of the cities or towns (Ren and Zhou 2014; Ren et al. 2015). These methods have been confirmed to be effective and applicable in regions like China's mainland and the United States.

Despite its importance, few works, if any, have been conducted so far to investigate the urbanization effect on trend estimates of extreme temperature change on a spatial scale larger than subcontinental outside China's mainland, mainly due to the difficulty of obtaining a reliable dataset of rural stations. Most of the abovementioned procedures for determining rural stations are not workable on global land scale due to the lack of detailed metadata, the incompatibility of the data, and the time-consuming nature of the task. It is thus not clear to what extent urbanization has exerted an impact on the currently estimated linear trends of temperature indices series on global land or a region large enough for a robust detection of extreme climate change, such as those reported in many previous works including Alexander et al. (2006), Donat et al. (2013a,b), Zhang et al. (2019), Dunn et al. (2020), Klein Tank and Können (2003), Peterson et al. (2008), Trewin (2001), and Vincent and Mekis (2006).

The main objective of this paper is to apply a new machine learning method called "isolation forest" (Liu et al. 2008) to select rural stations as reference stations, and to assess the urbanization effect on surface air mean and extreme temperature indices change on global land scale. We describe the data and method in section 2. The results are presented in section 3, and a brief discussion of the results is provided in section 4. The conclusions are summarized in section 5.

#### 2. Data and methods

#### a. Data sources

Three types of datasets were used in this study. They are the daily maximum and minimum surface air temperature datasets, the locations (longitude and latitude) of the U.S. Climate Reference Network (USCRN) dataset (Diamond et al. 2013), and the global LULC dataset (Hollmann et al. 2013). The daily surface air temperature datasets are used to calculate temperature indices, the locations of USCRN are used as a training dataset in machine learning, and the LULC dataset is used to estimate the percentage of urban areas around the stations at different buffer radii.

## 1) DAILY SURFACE AIR TEMPERATURE DATASETS

Several international or national daily surface air temperature datasets were collected and then integrated to form a new global land daily surface air temperature dataset in this study. The sources of datasets are 1) the Global Historical Climatology Network-Daily (GHCND) dataset (Menne et al. 2012a,b); 2) the homogenized datasets of European Climate Assessment and Dataset (ECA&D) (Squintu et al. 2019) and for Australia (Trewin et al. 2020), Canada (Vincent et al. 2012), and China's mainland (Cao et al. 2016); and 3) the three national datasets (for South Korea, Russia, and Vietnam) that were exchanged with China Meteorological Administration (Xu et al. 2014; Zhang et al. 2019). Details of the dataset sources are presented in Table 1.

It should be noted that although the uncorrected urbanization effects at most stations manifest as gradual changes, at some stations they may manifest as step changes, and the series of these stations would be identified as inhomogeneous and be adjusted by the homogenization procedure (Menne et al. 2009; Trewin 2013). It is possible that some part of the urbanization effect signal in the homogenized datasets used in this study may have already been removed by homogenization.

In forming the new dataset, the GHCND (Menne et al. 2012a,b) was used as a benchmark to integrate other datasets. The datasets for the ECA&D, Australia, Canada, and China's mainland in the GHCND (Menne et al. 2012a,b) were directly replaced by the corresponding homogenized datasets of ECA&D (Squintu et al. 2019), Australia (Trewin et al. 2020), Canada (Vincent et al. 2012), and China's mainland (Cao et al. 2016). When the exchanged (South Korea, Russia, and Vietnam) datasets (Xu et al. 2014; Zhang et al. 2019) were integrated, there was a need to check whether there were duplicated stations between GHCND and the exchanged stations. The stations are regarded as duplicated stations if the differences of latitude and longitude between the stations of two sources are less than 0.01°. In this case, the stations with longer series are retained. When integrating datasets, only the stations with at least 10 years of data in the reference period (1961-90) were retained, and a total of 14004 stations were retained in the end.

Although the quality control of many original datasets had been conducted, after these datasets were integrated into a new global dataset a renewed quality control was conducted. The main purpose of the quality control is to identify possible erroneous data. It mainly includes an internal consistency check, an outlier check, a local climatological extreme value check,

TABLE 1. Summary of dataset sources.

Source	Name/nation	Homogenized	No. of stations	Reference
Global	GHCND	No	28 334	Menne et al. (2012a,b)
Regional	ECA&D	Yes	2495	Squintu et al. (2019)
National	China's mainland	Yes	2419	Cao et al. (2016)
	Australia	Yes	112	Trewin et al. (2020)
	Canada	Yes	338	Vincent et al. (2012)
	Russia	No	88	Xu et al. (2014); Zhang et al. (2019)
	South Korea	No	70	Xu et al. (2014); Zhang et al. (2019)
	Vietnam	No	22	Xu et al. (2014); Zhang et al. (2019)

and a check on whether the record exceeds the world record. The details of the quality control procedure are same as described in the study of Zhang et al. (2019). After quality control, 13 956 stations with at least 10 years of non-missing values during 1961–90 were selected. If the missing values reach more than 15 days in a year, the annual value of this year is regarded as a missing value.

After that, the inhomogeneity test for the datasets that had not been homogenized was performed using the RHtests V4 software (Wang 2008a,b; Wang and Feng 2013) without a reference series. The use of a reference series is preferable if it is available; however, it is always difficult to find enough homogeneous reference series in combination with metadata, and it is also a laborintensive task, which prevents the method (reference series based RHtests V4) from being applied to global-scale datasets (Wang and Feng 2013). For global datasets, the metadata are not always easily available; even if the metadata are available, they are often incomplete (Trewin 2013). Meanwhile, if the inhomogeneity test was carried out using reference series, such a method may also have difficulty in handing network-wide changes, such as a national change in instrument type or observing time (e.g., the instrument automation of the China's observational stations, which mainly occurred during 2003-05).

The monthly mean temperature series was used to detect the changepoints because the daily temperature series has higher noise and thus it is more difficult to detect inhomogeneities (Wang and Feng 2013). If the breakpoints were detected at the 99.99% confidence level, the station data series were excluded in the subsequent analysis. The 99.99% confidence level was adopted so that only the most significant inhomogeneous data series would be identified. Similarly, the inhomogeneity test conducted in this research may also detect the step changes caused by urbanization for certain stations, and these series affected by urbanization effect could be excluded in the above procedure analysis. A total of 1451 stations were identified as inhomogeneous and were discarded. In the end, 12 505 stations were retained for use in this study.

#### 2) USCRN DATASET

The U.S. National Oceanic and Atmospheric Administration (NOAA)/National Climatic Data Center (NCDC) developed the USCRN dataset (Diamond et al. 2013), which has 234 stations across the United States. The series length of USCRN stations is no more than 20 years, and it is not enough to assess any long-term climate change. Therefore, only the location

information (longitude and latitude) of the USCRN stations was used to fit the model for selecting rural stations from the global land temperature stations in the abovementioned dataset.

## 3) LULC DATASET

The European Space Agency (ESA) Climate Change Initiative (CCI) Land Cover project developed a satellite-based global LULC product at 300-m spatial resolution from 1992 to 2018 (Hollmann et al. 2013), which makes it possible to select rural stations free from urbanization effects on the global scale. This dataset contains 38 LULC typologies including urban areas, cropland, tree cover, grassland, bare areas, water bodies, and so on; therefore, it can be determined whether the observational settings of the stations are affected by urbanization based on the percentage of urban areas around the stations at the different spatial scales (Tysa et al. 2019). We only used the global LULC product in the latest year of 2018 corresponding the study period of 1951–2018.

### b. Selection of rural reference stations

The USCRN data are free subject to LULC change, and can be used in the future to assess climate change precisely (Diamond et al. 2013). In particular, the locations of USCRN stations were carefully selected with a strict criterion and were placed in rural environments that are away from urban areas, and thus they can be regarded as typical rural stations that are not affected by urbanization. The rural environment was expected to be free of the impact of urbanization for at least 50 years. Our goal was to select some rural stations from the global land temperature stations as reference stations, and the current LULC around the selected rural stations should be similar to that of the USCRN stations.

The problem of dividing global land stations into rural stations and urban stations can be regarded as a binary classification problem in machine learning. The traditional binary classification paradigm of machine learning, including logistic regression, random forest, and support vector machine methods, aimed to classify an unknown dataset into two categories based on a known training dataset comprising the positive and negative instances (Khan and Madden 2010). However, only the information of positive instances of the training dataset [i.e., the rural stations (USCRN)] was available in this study; the information of negative instances (i.e., the stations that were affected by urbanization) was not available. The urban stations may be affected by urbanization to varying degrees (e.g., slightly, somewhat, or highly), and so they are not likely to conform to a



FIG. 1. The percentage of urban areas around the (a) USCRN stations and (b) all global land stations at 1-12-km (1-km increment) buffer radii, respectively. The terms r1-r12 on the x-axis labels refer to buffer radii of 1-12 km (1-km increment), respectively. The figure was originally presented as a scatterplot; the x axis of each of the 12 panels represents the ordinal of the points, and there are so many points that they obscure each other. To avoid overlapping points, the data (points) are binned into hexagons for display. The color bar shows the station counts in a hexagon.

statistically representative profile; therefore the urban stations can be regarded as anomalies or outliers. In this case, we needed a method to decide whether a new instance (from global land stations) belongs to the same distribution as existing instances (i.e., USCRN stations).

Since only one information category was available, the problem of selecting rural stations can be regarded as the problem of oneclass classification (OCC) in machine learning (Khan and Madden 2010), and the OCC can be applied to the anomaly detection theme (Chandola et al. 2009; Khan and Madden 2014). The anomaly detection included novelty detection and outlier detection; in the context of novelty detection, the training dataset was not contaminated by anomalies or novelties; and in the outlier detection context, the training dataset was partly contaminated by anomalies or outliers (Pedregosa et al. 2011; Scikit-learn developers 2019). Figure 1a shows the percentage of urban areas around the 234 USCRN stations at each of the 1–12-km (1-km increment) buffer radii, as proposed in Tysa et al. (2019). As can be seen, the percentages of urban areas were very small for most stations, but relatively large for a few stations, which can be regarded as anomalies/outliers. The outliers indicate that the USCRN stations may be contaminated to some extent (i.e., a few USCRN stations had been actually affected by urbanization). Therefore, the method of outlier detection was adopted for this study.

One new efficient algorithm referred to as "isolation forest" (Liu et al. 2008) was employed for performing the outlier detection. This algorithm detects anomalies/outliers only based on the concept of isolation instead of employing the distance or density measure of the existing methods, and it tries to fit the regions of which the high-density regions of the training dataset, ignoring the low-density regions (anomalies/outliers) of the training instances (Liu et al. 2008; Scikit-learn developers 2019). The experiment showed that this algorithm outperforms most existing anomaly detection approaches in performance (Liu et al. 2012). The algorithm was implemented using an object "ensemble.IsolationForest" of the Python module "Scikit-learn"<sup>1</sup> (Pedregosa et al. 2011) in this study. The specific steps of selecting rural stations from the global land stations based on the LULC dataset and USCRN adopting the isolation forest algorithm (Liu et al. 2008) were summarized as follows:

Calculate the percentage of urban areas around the stations. It is well known that the percentage of urban areas around the station is a key indicator to decide whether or to what extent a station is affected by urbanization (Ren et al. 2015; Li et al. 2019; Tysa et al. 2019). However, what buffer radius should be used to calculate the percentage of urban areas remains a question.

Referring to Ren et al. (2015) and Tysa et al. (2019), the percentages of urban areas around the 243 USCRN stations and 12 505 global land stations at each of the 1-12-km (1-km increment) buffer radii were calculated respectively based on the ESA CCI LULC product in the year of 2018 (Hollmann et al. 2013) using the R package raster (Hijmans 2020). If the distance between the grid cell center and the station location was less than or equal to the buffer radius, then the grid cell was included in this buffer radius extent. The percentages of urban areas of USCRN stations are shown in Fig. 1a, and the percentages of urban areas of all the global land stations are shown in Fig. 1b. Thus, each station used 12 dimensions to depict how much the observational site was affected by urbanization. The percentages of urban areas around the USCRN stations at 1–12-km buffer radii were used as the training dataset (a 243  $\times$ 12 matrix) to fit the model with an unsupervised way, and then the model was applied to predict whether a new instance from the global land stations belongs to the most densely distribution of the training dataset (it is an inlier, i.e., a rural station), or



FIG. 2. (a) The outlier rates for each contamination parameter. Blue indicates the training dataset, which was sampled 70% from the USCRN dataset; red indicates the test dataset, which was the remaining 30% dataset of the USCRN. (b) The number of sorted global rural stations and urban stations based on the fitted model learned from the full USCRN dataset with a different contamination parameter.

the new instance should be considered as an outlier (i.e., an urban station).

2) Select an appropriate *contamination* parameter to fit the model. As shown in Fig. 1a, there were a few anomalies/ outliers in the USCRN dataset (i.e., the USCRN dataset was contaminated). Therefore, a key issue was to determine the appropriate contamination parameter in fitting the model by using the contaminated training dataset (USCRN).

First, we randomly sampled 70% USCRN dataset (a  $170 \times$ 12 matrix) as training dataset to fit the model, and the remainder 30% was used as test dataset (a  $70 \times 12$  matrix) to test the predict result. Second, the contamination parameter was set to 0-0.5 (0.05 increment), respectively. The experiment result is shown in Fig. 2a. As can be seen, the outlier rates (the rate of urban stations) of the training dataset and test dataset are similar, so the USCRN dataset can be used to fit the model. Then we used the full USCRN dataset as the training dataset to fit the model, and the contamination parameter was similarly set to 0-0.5. The new instances (global land stations) were sorted as inliers (rural stations) or outliers (urban stations) with the learned model from the training dataset. As the contamination value increased, the number of sorted rural stations decreased (see Fig. 2b). When the contamination parameter reached 0.3, the numbers of

<sup>&</sup>lt;sup>1</sup> The version of "Scikit-learn" used in this study is 0.23.0, which requires Python version 3.6 or higher.



FIG. 3. As in Fig. 1, but for the (a) global land rural stations and (b) global land urban stations at 1–12-km buffer radii, respectively.

classified global rural stations and urban stations were 3444 and 9061, respectively. The corresponding percentages of urban areas around the stations at 1–12-km buffer radii are shown in Fig. 3. Most of the percentages of urban areas around the classified global land rural stations were less than 3%, which is consistent with what we had known about the rural stations. Therefore, 0.3 was adopted as the final contamination parameter. Other parameters in the object "ensemble.IsolationForest" of the Python module "Scikitlearn" (Pedregosa et al. 2011) were determined using the default parameters settings. Obviously, simply defining rural stations as the observational sites that have less than a specified percentage (e.g., 5% or 3%) of urban areas within 12 km (or the other buffer radius) may be still feasible, but determining this percentage threshold is somewhat subjective and arbitrary. In addition, for a regionscale research, artificially selecting rural stations based on the site selection criteria of USCRN is also feasible, but for globalscale dataset, the artificial selection is a labor-intensive task and difficult to complete. The method applied in this research not only refers to the procedures and results of artificial selection in the USCRN, but also realizes the semi-automation,



FIG. 4. (a) Overlay map of LULC (spatial resolution of  $300 \text{ m} \times 300 \text{ m}$ ) and stations classification in Beijing. Also shown are (b) Thanghekou, (c) Xiayunling, and (d) Zhaitang, the three stations in the Beijing region. The light-green background represents "rural" areas, and the yellow background and grids are "urban" areas. Black dots and red dots indicate the stations classified as rural and urban stations, respectively. The original LULC product contains 38 LULC typologies, such as urban areas, cropland, tree cover, grassland, bare areas, and water bodies. All LULC typologies except urban areas were reclassified as rural areas here.

which can be further generalized and potentially applicable in the relative research.

Another problem is the resolution of each station's latitude and longitude (coordinates). The dataset used in this study comes from different sources with different spatial resolution. The station coordinates of China's mainland (Cao et al. 2016) are stored in whole degrees and minutes, which is equivalent to a resolution of  $0.016^{\circ}$  (1'/60') when converted to decimal degrees. The station coordinates of ECA (Squintu et al. 2019) are stored in degrees, minutes, and seconds, which is equivalent to a resolution of  $0.0003^{\circ}$  (1"/3600") when converted to decimal degrees. The resolution of the stations coordinates of Australia (Trewin et al. 2020) and Canada (Vincent et al. 2012) are 0.01°. Although the resolution of coordinates in the GHCND are rounded to 0.0001°, it is actually an integrated dataset, and its original data accuracy cannot be equal to 0.0001°. Coordinate accuracy within 0.01° will be accurate to about 1 km, which means that the percentage of urban areas within a 1-km buffer radius will be unreliable for some stations. However, considering that the coordinate resolution of some stations is higher than 0.01°, and the urban LULC within the 1-km buffer radius of the stations has an important influence on the temperature series, the percentage of urban areas within 1 km was still used for each station in this research.

Finally, 3444 stations with a percentage of urban areas similar to the USCRN stations were classified as rural stations, while the other 9061 stations affected by urbanization to varying degrees were classified as urban stations. Figure 4 presents examples of stations defined as urban or rural stations in Beijing. It shows that the stations located in or near urban areas can be correctly defined as urban stations, but the Thanghekou and Zhaitang stations, which are actually located in rural areas, have been determined as urban stations because of the strict criteria used and the insufficient resolution of the stations' coordinates. In this way, the last classified rural stations are likely to be real rural stations, but a few of the classified urban stations are actually also rural stations. This might not be a big problem for the averaged series of regional or



FIG. 5. The station distribution of each region for (a) all stations and (b) rural stations. Different colors refer to the different regions. The extent of East Asia is defined as  $3^{\circ}$ -50°N, 90°-150°E, as shown in the rectangular box. The numbers in parentheses in the legend indicate the number of all stations (left value) and rural stations (right value) for each region.

global scales, but for local scales (i.e., grid) the impact could be relatively large, because a few so-called urban stations and rural stations in the same grid are probably not much different.

The distribution maps of all stations and rural stations are shown in Fig. 5. As indicated in Fig. 5, the selected rural stations had a sufficient global spatial coverage to allow comparison with all of the global land stations. The annual station numbers of Australia, East Asia, Europe, North America, and global land for all stations and rural stations are shown in Fig. 6. Relatively fewer stations were classified as rural stations in East Asia, and there were larger numbers of rural stations in Europe and North America. The number of stations available in the early and late stages of the study period is relatively small, especially in East Asia, and the insufficient station coverage could increase the uncertainty of trend estimates.

#### c. Analysis methods

The annual average value of daily maximum temperature (Tmax), annual average value of daily minimum temperature (Tmin), and annual average value of daily mean temperature (Tmean; the daily mean temperature is the mean of daily maximum and minimum temperature), and some of the ETCCDI (Expert Team on Climate Change Detection and Indices) indices (X. Zhang et al. 2011), including cold nights (TN10p), cold days (TX10p), warm nights (TN90p), warm days (TX90p), yearly maximum value of daily maximum temperature (TXx), yearly maximum value of daily minimum temperature (TXx), yearly minimum value of daily minimum temperature (TXn), yearly minimum value of daily maximum temperature (TXn), yearly minimum value of daily maximum temperature (TNn), and diurnal temperature range (DTR), were used in this study. The ETCCDI indices were calculated based on the R language package "climdex.pcic" (Bronaugh 2020).

The stations used in this research are not evenly distributed across the global land, which would result in the global signal be dominated by the regions of higher station density if a simple arithmetic mean were used. Thinning the station network so that there are approximately evenly distributed stations in each region is an alternative method, used by Frich et al. (2002). However, this approach also discards a lot of useful climate data, and the choice of which stations to keep or discard is also somewhat subjective (Alexander et al. 2006). Therefore, we gridded the stations' temperature indices anomaly series (relative to 1961-90) into a regular latitude-longitude grid  $(5^{\circ} \times 5^{\circ})$  first, and then applying the grid area-weighted average method to obtain the global/regional average time series for all stations and rural stations, as Jones and Hulme (1996) proposed. The specific details are consistent with the study of Zhang et al. (2019). In this way, the impact caused by the uneven station distribution could be partly alleviated. It should be noted, however, that the values for each grid boxes were calculated by averaging all stations within the grid boxes, but only the grid boxes with at least one urban and one rural station were retained. Meanwhile, the grid difference series between all stations and rural stations for each grid were also calculated. When calculating the linear trends of the grid boxes, only the grid boxes with at least 66% of data series during 1951-2018 and the last year of the data series not earlier 2005 were used; when calculating the global average annual time series, only the grid boxes with at least 90% temporal completeness of data series during 1951-2018 were used.

A total of 229 grid boxes are available, which accounts for about 26% of the global land grid boxes (and about 32% of the global land area), and each grid box has at least 66% of data series during the research period and the last year of the data series is not earlier than 2005. Since 2014, however, the number of available global land grids has decreased significantly.

The urbanization effect ( $\Delta T_{\text{all-rural}}$ ) refers to the impact of urbanization on the estimated trends of temperature indices anomaly series for stations, grid boxes, and regional average (Ren et al. 2008; Ren and Zhou 2014); it is also defined here as the linear trend of temperature indices anomaly difference series between all stations and rural stations. Consider  $Y_{\text{all}}$  and  $Y_{\text{rural}}$  as the time series for all stations and rural stations, respectively. The temperature indices anomaly difference series  $Y_d$  can be expressed as

$$Y_d = Y_{\rm all} - Y_{\rm rural},\tag{1}$$

$$Y_d = a + b_d t_i + \omega_i, i = 1, 2, \dots, n,$$
 (2)

where *a* refers to the intercept,  $b_d$  refers to the linear trend of the anomaly difference series (i.e., the urbanization effect),  $\omega$  refers to the regression residual, and *t* refers to the ordinal time. The term  $b_d \times 10$  (i.e., the changing trend per decade) was actually adopted in this study.

Many temperature indices series do not follow a normal distribution (Zhang et al. 2019) and there may be outliers at the beginning and end of the series. Therefore, the nonparameter Theil–Sen trend estimator (Sen 1968) and Mann–Kendall test (Mann 1945; Kendall 1955), which are less affected by non-normal distribution and outliers in the series (Zhang et al. 2000), were adopted in this study. Meanwhile, lag-1 autocorrelation



FIG. 6. The station numbers for all stations and rural stations of Australia, East Asia, Europe, North America, and global land.

usually existed in most temperature index series (Zhang et al. 2019), which makes it easier to obtain a statistically significant trend (von Storch and Navarra 1999, 15–18). Therefore, an iterative prewhitening procedure described in appendix A of Wang and Swail (2001) was used to diminish the effect of lag-1 series auto-correlation when calculating the linear trends and testing their statistical significance. In addition, the correction term for ties (repetitive values in the time series) was included in calculating the variance *S* of the Mann–Kendall test (Qian et al. 2019).

The urbanization contribution  $(U_c)$  was defined as the percentage of the urbanization effect accounting for the overall trends of all stations (Ren et al. 2008; Ren and Zhou 2014), expressed as

$$U_{c} = |b_{d}/b_{all}| \times 100\%, \tag{3}$$

where  $b_{\rm all}$  refers to the linear trend of all stations (including rural stations and urban stations). The urbanization contribution was

calculated only when the urbanization effect was statistically significant at the 0.05 level.

In addition to the global land, regional averages of the temperature indices series and their urbanization effect and contribution for Australia, East Asia, Europe, and North America were also calculated for comparison. The four regions were chosen because of the relatively good data coverage of both all stations and rural stations. The divisions of the regions and the distributions of stations in each of them are shown in Fig. 5.

## 3. Results

Figure 7a shows the time series of anomalies of TN90p obtained using all stations and rural stations for global land, Australia, East Asia, Europe, and North America, respectively. As can be seen, the global and regional land averaged time series for TN90p indices over 1951–2018 all experience



FIG. 7. The global/regional land average annual anomalies (relative to the mean of 1961–90) over 1951–2018 for the warm nights (TN90p) index. (a) The all-stations series and rural-stations series for (top to bottom) global land, Australia, East Asia, Europe, and North America, respectively. (b) The difference series of annual mean anomalies between all stations and rural stations for (top to bottom) global land, Australia, East Asia, Europe, and North America, respectively. The global/regional average series were calculated using only those grid cells whose series completeness exceeds 90% during 1951–2018. The black straight lines in the bar chart of (b) are the fitted trend lines.

increasing trends regardless of whether the series are calculated from all stations or from rural stations only. At first glance, similar temporal variability and long trends are observed in series for all stations as well as rural stations, although the number of rural stations only accounts for 27.5% of the number of all stations and has less spatial coverage. However, a careful inspection of the global/regional time series anomalies of TN90p over 1951–2018 reveals that the increase in all stations was larger than for just for the rural stations. Since the magnitude of difference between the all-stations series and

rural-stations series is relatively small, the difference series between all stations and rural stations for global land and the four regions are computed, as shown in Fig. 7b.

A clear trend of the difference series (i.e., an urbanization effect) was observed in TN90p indices for global land, Australia, East Asia, and North America, which indicates that warm nights at urban stations occurred more frequently than in rural areas (Fig. 7b). The urbanization effect mainly occurred after the mid-1980s, which is consistent with the world urbanization process. Global land and North America, which have a relatively large number of stations, show a clear and stable (low fluctuation) urbanization signal since the mid-1980s. The most evident signal of urbanization effect was observed in East Asia, which may be related to the unprecedentedly rapid process of urbanization in East Asia (especially in China's mainland) in recent decades. The signals of the urbanization effect in Australia and East Asia are clear, but the temporal fluctuations of the difference series are larger than those in global land and North America, which may be related to the available station numbers being relatively small in these two regions, especially at the beginning of the study period. There are no statistically significant trends (at the 0.05 level) observed in Europe, which may be due to the stagnant growth of UHI effects around the observational stations in recent decades. The studies of Jones et al. (2008) and Jones and Lister (2009) showed that the UHI for Vienna and central London might have developed before the start of the twentieth century; therefore the stations in Vienna and central London did not show the urban-related warming trends compared to the nearby rural stations. However, these results for Vienna and London may not be fully extrapolated to other cities of Europe. In addition, some of the defined urban station locations of Europe might have been less affected by urbanization, which would cause the difference between defined rural stations and urban stations to be relatively small in this study.

The original anomaly series and difference series of TN10p, TX10p, TX90p, TXx, TNx, TXn, and TNn were also calculated, but they were presented in supporting material (see Figs. S1-S7 in the online supplemental material) due to the limited space of this paper. In the TN10p differences series, only Australia and East Asia experienced a statistically significant urbanization effect. No statistically significant urbanization effect was observed for TX10p in all regions. For TX90p, North America and global land experienced a slightly urbanization effect associated with urban warming, but Australia experienced an urbanization effect associated with urban cooling. On the whole, the warm temperature indices related to the daily minimum temperature (TN90p) experienced a stronger urbanization effect, while the cold temperature indices related to the daily maximum temperature did not experience any significant urbanization effect in any regions.

Figure 8a shows the Tmax anomalies time series of all stations and rural stations for global land, Australia, East Asia, Europe, and North America, respectively. Figure 8b shows the corresponding difference series between all stations and rural stations. A slightly weak urban warming signal was detected in global land and North America, and a strong urban cooling effect was detected in Australia. The Tmax difference series of Australia experienced a downward trend; that is, the ruralstations series is warming faster than the all-stations series, which is not consistent with other regions.

Figure 9 shows the Tmin time series anomalies of all stations and rural stations for each region. A strong urbanization effect can be observed in the East Asia, and a slightly weak urbanization effect can be observed in global land, Australia, and North America. The urbanization effect mainly occurred after the mid-1980s, and the temporal evolution characteristics of Tmin are generally similar to TN90p indices. The anomaly difference series in Australia and Europe shows large fluctuations, but no statistically significant trend is detected in Europe.

The DTR difference series witnessed an evident urbanization signal for Australia and East Asia (Fig. 10). However, the causes of the urbanization effect in these two regions are different. In Australia, the DTR anomalies time series showed a slightly gradual upward trend for all stations and rural stations, with that of all stations having a smaller positive trend than the rural station; however, the anomalies series of East Asia experienced a downward trend for both all stations and rural stations, with that of all stations having a larger negative trend than the rural stations. A positive and significant urbanization effect in the DTR difference series was observed in Europe. This was mainly caused by the more rapid increase in DTR of all stations than that of rural stations after the 1980s. A slightly weak urbanization effect can be found in global land.

The annual mean Tmean of all stations had a statistically significant urbanization effect in global land, East Asia, and North America (Fig. 11). The urbanization effect in East Asia is the strongest, which is consistent with other temperature indices. On global land, the magnitude of urbanization effect in the Tmean data series was between those of East Asia and North America.

The left panel of Fig. 12 shows the spatial distribution of urbanization effect for the TN10p, TX10p, TN90p, and TX90p indices, and the right panel of Fig. 12 shows the corresponding the urbanization effect and urbanization contribution of regional average series.

The urbanization effects in TN10p series of all stations are generally negative, although they have some insignificant positive values in some grid cells; the largest urbanization effects are mainly distributed in East Asia. The urbanization contribution of the TN10p series in global land and East Asia reach 7.2% and 13.7%, respectively. The urbanization contributions in other regions were not calculated because the urbanization effects are not statistically significant at the 0.05 level. Many statistically significant grid cells can be found in Europe and North America (Fig. 12a), but the overall urbanization effect of the regional average TN10p series in these two regions are not statistically significant, which may be due to the offset of the positive and negative value grid cells in different areas within these regions. For example, southern Europe mainly shows negative trends, while northern Europe mainly shows positive trends.

The urbanization effects of TX10p (Fig. 12b) for all regional average series are not significant statistically, so the



FIG. 8. As in Fig. 7, but for Tmax (maximum temperature).

corresponding urbanization contributions were not calculated. However, a few statistically significant grids are observed in Europe and the United States.

The urbanization effects of TN90p (Fig. 12c) are mostly positive values, and the largest urbanization effects are mainly distributed in East Asia and North America. Although the magnitude of the urbanization effect and urbanization contribution in Australia is largest, the results have great uncertainty because the grid coverage is insufficient in this region. Europe as a whole shows a negative trend. The urbanization contributions of global land, Australia, East Asia, and North America in the TN90p series reach 17.2%, 35.7%, 20%, 7.6%, and 21.7%, respectively.

For TX90p (Fig. 12d), northern Europe is mainly characterized by positive grid cells, and central Europe has more negative grid cells. Statistically significant negative values are observed in the North China Plain, indicating that the urbanization effect has reduced the frequency of warm days (TX90p), which may be related to the more serious air pollution in urban areas of the North China Plain in the recent three decades as





compared to earlier decades (Qian et al. 2003; Zhang et al. 2012). There are also a few statistically significant negative grids observed in arid region of western China, but this may be related to the enhanced local oasis effect (Ren and Zhou 2014). In these regions, there is relatively more vegetation and irrigation in urban areas than rural areas, the evapotranspiration is relatively strong, and the urban cold island effect may be formed in particular during daytime (Su and Hu 1988). Since the positive and negative valued grid cells in Europe and East Asia cancel each other out, the overall urbanization effect in these regions is not

statistically significant. The urbanization effects of global land, Australia, and North America are statistically significant, and the corresponding urbanization contributions are 11.2%, 23.8%, and 16.2%, respectively.

The left panel of Fig. 13 shows the urbanization effects for Tmax, Tmin, DTR, and Tmean indices, and the right panel of Fig. 13 shows the corresponding urbanization contributions of regional average series.

For Tmax (Fig. 13a), only a few grid cells show a statistically significant urbanization effect. A few negative values in grid





cells can be found in the North China Plain, which may also be related to the excessive anthropogenic emissions of aerosols in the region. Among the urbanization effects of the regional annual average series, only global land, Australia, and North America are statistically significant, and the corresponding urbanization contributions are 11.1%, 29.9%, and 13.8%, respectively. Meanwhile, the urbanization effects for North America and the global land are positive values, whereas that in Australia is a negative value (Fig. 8). However, the urban cooling effect of Tmax for Australia may not be caused by the increased aerosol because the aerosol pollution in Australia is less serious than the Northern Hemispheric average condition and has been declining in recent decades (Keywood et al. 2016); in addition, there are only a few available gridbox series for constructing the regional average time series of Australia, which may cause great uncertainty in the estimation of the urbanization effects in Australia.

For Tmin (Fig. 13b), East Asia and central Asia show the strongest urbanization effect. The urbanization effect of regional average series in global land, Australia, East Asia, and



FIG. 11. As in Fig. 7, but for Tmean (mean temperature).

North America are statistically significant, and the corresponding urbanization contributions are 13.9%, 14.7%, 21.7%, and 6.7%, respectively.

Statistically significant negative grid cells of DTR (Fig. 13c) are observed in East Asia, central Asia, and North America. However, the overall urbanization effect in North America is not statistically significant, which may be because the positive and negative values offset each other in the region. The negative urbanization effects in global land, Australia, and East Asia and the positive urbanization effect in Europe

are statistically significant; the corresponding urbanization contributions reach 20.8%, 100%, 47.0%, and 100%, respectively.

For Tmean (Fig. 13d), the urbanization effect in Europe is generally small and statistically insignificant. Many statistically significant grid cells are observed in Asia and North America. The urbanization effects in global land, East Asia, and North America as a whole are statistically significant, with the corresponding urbanization contributions for the period of 1951–2018 are 12.7%, 15%, and 9.1%, respectively. East Asia



FIG. 12. (left) The trends of grid differences series between all stations and rural stations and (right) the corresponding regional average urbanization effects and urbanization contributions of Australia, East Asia, Europe, North America, and global land for (a) TN10p, (b) TX10p, (c) TN90p, and (d) TX90p during 1951–2018. The trends of grid cells were calculated using only those grid cells whose difference series completeness exceeds 66% during 1951–2018 and the non-missing value of the last year of the series does not occur earlier than 2005, and each grid cell contains at least one rural station and one urban station. The black points in the grid cells of the trend distribution map indicate that the trends are statistically significant at the 5% level. The error bars on the bar charts refer to the 95% confidence interval for the urbanization effects. The blue percentage numbers above or below the bar charts refer to the urbanization contributions. The TN10p and TX10p are cold threshold indices, the positive trends of grid difference series between all stations and rural stations indicate urban varming effects. Therefore, the color bar of these two indices were inverted so that the warm colors always represent urban varming.



FIG. 13. As in Fig. 12, but for (a) Tmax, (b) Tmin, (c) DTR, and (d) Tmean.

generally registered the largest urbanization contribution in terms of regional-averaged annual mean temperature trend.

It should be noted that in data-sparse areas, an urbanization signal (i.e., the statistically significant trends of grid difference series between all stations and rural stations) may actually be detecting a spatial gradient in trends. There are a few outliers in the early research period in East Asia and Australia, and the outliers may be caused by the insufficient station coverage in these regions. But this does not seem to have a significant effect on the estimated trends because the Theil–Sen trend estimator, which is less affected by outliers in the series, was adopted in this study.

### 4. Discussion

### a. Representativeness of rural stations

In evaluating urbanization effects in the observational data series, an important issue would be how to select the "real" rural stations or reference stations. In this work, we applied machine learning method to select the global land rural stations. This might also be the first time that the machine learning method has been used to select rural reference stations for evaluating the impact of urbanization on changes in the surface climate elements of stations. The isolated forest algorithm (Liu et al. 2008) belongs to the category of anomaly detection, so it can only determine whether the detected object and the learned training dataset belong to the same distribution, and cannot be used to rank the urbanization influence on stations. For the purpose of this paper, however, this is enough, because we only care about the impact of urbanization and the relative contribution of urbanization to the full change trends.

However, when using the isolated forest algorithm to select rural stations, there are still two issues to be considered.

The first question is what buffer radius should be used to calculate the percentage of urban areas around the stations? Ren and Zhou (2014) and Ren et al. (2015) used a 2-km buffer radius, but comprehensively considered the distance of relocation and the population of the town or city where the stations belong, considering that the town or city population is actually equivalent to increasing the use of a larger buffer radius. If the buffer radius is too small, there may be a situation where the station has a small percentage of urban land within the buffer radius, which meets the standards of rural stations, but the station may still be affected by an urban heat air dome (Ren et al. 2015). Fan et al. (2017) used a simple energy balance model to suggest that the horizontal extent of the urban heat air dome can reach 1.5-3.5 times the city radius at the night and 2.0-3.3 times at the day. Therefore, the most ideal method is to select the rural stations according to the actual city radius. If the station falls outside 3.5 times the city radius, then it can be considered as meeting the standards of a rural station. However, there are two difficulties in doing so. One is that the metadata of the city radius of each city are still difficult to obtain automatically, and it is also difficult to automatically determine the location of the city center; another is that some cities are approximately square or round in shape, but some cities have a spatial layout along river valleys or coastal lines. There is no such thing as a city radius if it is distributed in a band shape.

In addition, if the rural stations are selected strictly according to location outside 3.5 times the city radius, then there may not be enough stations for use. According to the LULC dataset provided by the ESA (Hollmann et al. 2013), for an extreme example, the radius of urban areas in Beijing in 2018 is about 60 km. If the maximum extent of the urban heat air dome in the horizontal direction is 3.5 times the city radius, then within 210 km from the urban center observational stations may still be affected by the urban heat air dome. In this rare case, even if the finally selected rural stations whose percentage of urban areas around the stations at 1–12-km buffer radius are close to 0, then on a larger spatial scale, the "rural stations" may still be affected

by the urban heat air dome or heat air plume; that is, our estimated urbanization effect based on these rural reference stations may still be underestimated.

Tysa et al. (2019) calculated the correlation coefficient between the percentages of urban areas around the stations at 1-20 km (1-km increment) buffer radius and the annual mean temperature trends of period 1960-2015 for 2286 national stations of China's mainland, and found that when the buffer radius reaches 4 km, the correlation coefficient is the largest, and then the correlation gradually weakens with increase of the radius, but it more or less stabilizes when it is larger than 16 km, as shown in Fig. S1 of Tysa et al. (2019). Therefore, the percentages of urban areas within the buffer radius of 1-16 km were used to select the rural stations, and the reciprocal of the buffer radius is used as the weight when dividing the level of the stations affected by urbanization. According to Fig. S1 of Tysa et al. (2019), when the buffer radius reaches 12 km, the correlation coefficient has not changed evidently. When the machine learning method is used to fit the model, the higher the dimension of the feature matrix of the training dataset, the more likely it is to contain some extraneous redundant information, thereby reducing the performance of machine learning algorithm for prediction or classification. This is called the "curse of dimensionality." Therefore, we only used the percentage of urban areas within a buffer radius of 1-12 km around the stations to fit the model.

The second issue is how to select an appropriate contamination parameter when using the isolation forest algorithm (Liu et al. 2008) to fit the model.

Since the establishment of the USCRN, the LULC around a small number of stations has changed (Diamond et al. 2013), which leads to the training dataset (USCRN) not being pure, or being affected by urbanization to a small extent. Figure 1a shows that the percentages of urban areas around some USCRN stations have even exceeded 40%. Therefore, when using the isolated forest algorithm (Liu et al. 2008) to learn the training dataset, a key is to find an appropriate contamination parameter. However, this is not an easy problem to solve, because we are not sure what proportion of USCRN stations are contaminated (or are affected by urbanization). Determining this contamination parameter requires repeated experiments. Finally, when the contamination parameter is set to 0.3, the percentages of urban areas around the selected global land rural stations at 1-12-km buffer radius are generally less than 3%, as shown in Fig. 3a. Therefore, a value of 0.3 was adopted as the final contamination parameter.

The choice of the contamination parameter is somehow subjective and arbitrary. Obviously, if a larger contamination value is used, the representativeness of the selected rural stations will be increased, but the number of the available rural stations will be decreased (see Fig. 2b). In the case of application of a more representative rural station network, the estimated urbanization effects would be larger than those presented in this paper. Therefore, as always, the urbanization effects and contributions contained in the daily temperature dataset as estimated in this work are conservative.

Data inhomogeneities induced by relocations of stations and replacement of instrumentations will also affect the

TABLE 2. Comparison of urbanization effect  $(U_e)$  and urbanization contribution  $(U_c)$  for different studies. In the research of Zhou and Ren (2011) and Ren and Zhou (2014), the units of  $U_e$  for TN10p, TX10p, TN90p, and TX90p are days decade<sup>-1</sup> because the units of the indices in these studies were converted to days from percentages; in the research of this paper and Sun et al. (2019), the units of  $U_e$  for TN10p, TX10p, TN90p, and TX90p are % decade<sup>-1</sup>. The units of  $U_e$  for Tmax, Tmin, DTR, Tmean, TXx, TNx, TXn, and TNn are °C decade<sup>-1</sup>. Boldface values indicate that the  $U_e$  is significant at the 5% level. "N/A" (not available) indicates that the  $U_c$  was not calculated because the corresponding  $U_e$  is not significant at the 5% level. A dash (—) indicates that  $U_e$  and  $U_c$  were not calculated in the study.

		This paper (East Asia, 1951–2018)	Zhou and Ren (2011) (North China, 1961–2008)	Ren and Zhou (2014) (China's mainland, 1961–2008)	Sun et al. (2019) (East China, 1958–2012)
TN10p	$U_e$	-0.21	-3.76	-1.49	-0.52
-	$U_{c}$ (%)	13.7	43.7	17.6	30.0
TX10p	$U_e$	-0.02	-0.36	-0.39	-0.09
	$U_{c}$ (%)	N/A	10.5	12.5	N/A
TN90p	$U_e$	0.44	3.87	2.27	0.82
	$U_{c}$ (%)	20.0	48.0	26.4	40.0
TX90p	$U_e$	0	-0.08	0.56	0.01
	$U_{c}$ (%)	N/A	N/A	10.7	N/A
Tmax	$U_e$	0	—	0.02	—
	$U_{c}$ (%)	N/A	—	10.1	—
Tmin	$U_e$	0.07	_	0.07	—
	$U_{c}$ (%)	21.7	—	18.4	—
DTR	$U_e$	-0.05	—	-0.05	—
	$U_{c}\left(\% ight)$	47.0	_	31.8	—
Tmean	$U_e$	0.04	_	0.05	—
	$U_{c}\left(\% ight)$	15.0	_	15.4	—
TXX	$U_e$	0.01	-0.02	0.03	
	$U_{c}$ (%)	N/A	N/A	22.7	
TNX	$U_e$	0.05	0.16	0.07	
	$U_{c}$ (%)	23.7	61.5	26.1	
TXN	$U_e$	0.04	0.02	0.05	
	$U_{c}$ (%)	N/A	N/A	13.6	
TNN	$U_e$	0.08	0.31	0.12	
	$U_{c}$ (%)	N/A	50.8	19.8	

representativeness of selected rural stations. Knowing that a station is located in a rural area now does not mean it has always been there. If the selected rural station was moved from urban areas to the present rural areas, then the temperature series of the "rural station" would also include the signal of urbanization effect after the data were homogenized (Zhang et al. 2014). Therefore, the evaluation of urbanization effects would still contain biases, generally underestimating the magnitudes of the urbanization contribution in case of the movement of stations from urban to rural areas. However, the use of the homogenized dataset can also avoid biases in other circumstances such as an application of the carefully selected rural station data series as references in homogenization (Trewin 2013). Additionally, although we used some homogenized data and made a new check of the inhomogeneities, some small breakpoints may still exist in the current global land daily temperature data. The possible impact of the remaining inhomogeneities and of the processing procedure to homogenize the data on the selection of the rural stations and the estimates of urbanization effect and contribution for all stations needs to be further addressed in the future.

## b. Comparison with other studies

As mentioned in the introduction, at present there is no research of urbanization effect on the global scale about the ETCCDI indices, and only a few regional studies focusing on China's mainland have been conducted so far. These include the analyses of North China presented in Zhou and Ren (2011), China's mainland presented in Ren and Zhou (2014), and East China presented in Sun et al. (2019). A comparison of these studies is provided in Table 2.

The methods, study area, study period, and the dataset used for these studies are not exactly the same, thus weakening the comparability of these results. However, the comparison is still noteworthy. Most of the changes in the ETCCDI extreme indices, especially the nighttime extremes (TN10p and TN90p), show a relatively large impact of urbanization, but there are differences in the impact of urbanization among different studies, which may be due to the varied methods, study periods, study areas, and datasets used. The research of Zhou and Ren (2011) and Sun et al. (2019) generally showed the strongest urbanization effects, which may be related to eastern and northern China being the fastest-growing areas of urbanization in China's mainland.

It is interesting to note that the magnitude of the urbanization effect on global land averaged annual mean surface air temperature (Tmean) change over 1951–2018 in this study is 0.03°C decade<sup>-1</sup> and the corresponding urbanization contribution is 12.7%, which is larger than the results reported by the IPCC Fifth Assessment Report (Hartmann et al. 2013) and by other researchers (Peterson et al. 1999; Hansen et al. 1999, 2010; Parker 2006). It is also noteworthy that the daily temperature data coverage is relatively incomplete compared to that of the monthly mean temperature datasets, and the study period is also different, which may obviate a rigorous comparison with the previous studies. However, the new estimate of urbanization effect indicates a need to pay more attention to the systematic bias in the current analyses of global land and regional-scale surface air temperature change (Karl et al. 1988; Peterson et al. 1999; Ren et al. 2008; Hansen et al. 2010; Wang and Ge 2012; Jones et al. 2012; Tysa et al. 2019). This should be especially emphasized when the conservativeness of the estimate of urbanization effect due to the difficulty to obtain the real rural stations is considered.

#### 5. Conclusions

This paper proposes a method of machine learning (the "isolation forest" algorithm) to classify the observational stations into rural stations and urban stations. Based on the classification of rural and urban stations, the global/regional land annual mean temperature indices series for all stations and rural stations were calculated, and the urbanization effects and urbanization contribution of global land and regional annual mean temperature indices series for all stations are quantitatively evaluated. The main conclusions are summarized as follows:

- From the global land perspective, statistically significant urbanization effects were detected for most temperature indices series, especially for the warm temperature indices derived from daily minimum temperature, for the period 1951–2018.
- 2) During the period 1951–2018, the TN10p, TN90p, and TX90p index series experienced significant urbanization effects, reaching -0.08%, 0.25%, and 0.11% decade<sup>-1</sup>, respectively, and the urbanization contributions are 7.2%, 17.2%, and 11.2%, respectively; the urbanization effects of Tmax, Tmin, DTR, and Tmean reach 0.02°, 0.03°, -0.01°, and 0.03°C decade<sup>-1</sup>, respectively, and the corresponding urbanization contributions reach 11.1%, 13.9%, 20.8%, and 12.7%, respectively; the urbanization effects of TXx, TNx, and TXn are 0.02°, 0.03°, and 0.06°C decade<sup>-1</sup>, respectively, and the corresponding urbanization graph is a constrained by the urbanization effects of TXx, TNx, and TXn are 0.02°, 0.03°, and 0.06°C decade<sup>-1</sup>, respectively, and the corresponding urbanizations reach 16.5%, 14.9%, and 20.4%.
- 3) The urbanization effects on trends of global land and regional average annual mean temperature indices series generally occurred after the mid-1980s, which is most evident in the TN90p and Tmin series in East Asia, and to a lesser extent in North America and Australia.
- 4) There are significant differences on the urbanization effects for different regions. The urbanization effect in the temperature indices series of East Asia is the strongest, while Europe experienced the weakest urbanization effect for the temperature indices. Within each of the continents, there are also significant differences. A weaker urbanization effect signal was found in Europe, probably due to the cancellation of the positive and negative trends of urban heat island intensity in the south and north of the region.

Acknowledgments. The global land use/land cover dataset is provided by the ESA CCI Land Cover project (www.esacci.org). This study is supported by the National Key R&D Program of China (2018YFA0605603). We thank Dr. Chao Zhang for his help in drawing maps. We are also grateful to the three anonymous reviewers whose comments helped improve the original manuscript.

#### REFERENCES

- Alexander, L. V., and Coauthors, 2006: Global observed changes in daily climate extremes of temperature and precipitation. J. Geophys. Res., 111, D05109, https://doi.org/10.1029/ 2005JD006290.
- Bian, T., G. Ren, B. Zhang, L. Zhang, and Y. Yue, 2015: Urbanization effect on long-term trends of extreme temperature indices at Shijiazhuang station, North China. *Theor. Appl. Climatol.*, **119**, 407–418, https://doi.org/10.1007/s00704-014-1127-x.
- Bronaugh, D., 2020: climdex.pcic: PCIC implementation of Climdex routines, version 1.1-11. R package, https://CRAN.R-project. org/package=climdex.pcic.
- Cao, L., Y. Zhu, G. Tang, F. Yuan, and Z. Yan, 2016: Climatic warming in China according to a homogenized data set from 2419 stations. *Int. J. Climatol.*, **36**, 4384–4392, https://doi.org/ 10.1002/joc.4639.
- Chandola, V., A. Banerjee, and V. Kumar, 2009: Anomaly detection: A survey. ACM Comput. Surv., 41 (3), 1–58, https:// doi.org/10.1145/1541880.1541882.
- Das, L., J. D. Annan, J. C. Hargreaves, and S. Emori, 2011: Centennial scale warming over Japan: Are the rural stations really rural? *Atmos. Sci. Lett.*, **12**, 362–367, https://doi.org/ 10.1002/asl.350.
- Diamond, H. J., and Coauthors, 2013: U.S. Climate Reference Network after one decade of operations: Status and assessment. Bull. Amer. Meteor. Soc., 94, 485–498, https://doi.org/ 10.1175/BAMS-D-12-00170.1.
- Donat, M. G., and Coauthors, 2013a: Updated analyses of temperature and precipitation extreme indices since the beginning of the twentieth century: The HadEX2 dataset. J. Geophys. Res. Atmos., 118, 2098–2118, https://doi.org/10.1002/jgrd.50150.
- —, L. V. Alexander, H. Yang, I. Durre, R. Vose, and J. Caesar, 2013b: Global land-based datasets for monitoring climatic extremes. *Bull. Amer. Meteor. Soc.*, **94**, 997–1006, https:// doi.org/10.1175/BAMS-D-12-00109.1.
- Dunn, R. J. H., and Coauthors, 2020: Development of an updated global land in situ-based data set of temperature and precipitation extremes: HadEX3. J. Geophys. Res. Atmos., 125, e2019JD032263, https://doi.org/10.1029/2019JD032263.
- Fan, Y., Y. Li, A. Bejan, Y. Wang, and X. Yang, 2017: Horizontal extent of the urban heat dome flow. *Sci. Rep.*, 7, 11681, https:// doi.org/10.1038/s41598-017-09917-4.
- Frich, P., L. Alexander, P. Della-Marta, B. Gleason, M. Haylock, A. Klein Tank, and T. Peterson, 2002: Observed coherent changes in climatic extremes during the second half of the twentieth century. *Climate Res.*, **19**, 193–212, https://doi.org/ 10.3354/cr019193.
- Fujibe, F., 2009: Detection of urban warming in recent temperature trends in Japan. Int. J. Climatol., 29, 1811–1822, https://doi.org/ 10.1002/joc.1822.
- Gallo, K. P., D. R. Easterling, and T. C. Peterson, 1996: The influence of land use/land cover on climatological values of the diurnal temperature range. J. Climate, 9, 2941–2944, https://doi.org/ 10.1175/1520-0442(1996)009<2941:TIOLUC>2.0.CO;2.

- Ge, J., J. Qi, B. M. Lofgren, N. Moore, N. Torbick, and J. M. Olson, 2007: Impacts of land use/cover classification accuracy on regional climate simulations. J. Geophys. Res., 112, D05107, https://doi.org/10.1029/2006JD007404.
- Handmer, J. Y., and Coauthors, 2012: Changes in impacts of climate extremes: Human systems and ecosystems. *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation*, C. B. Field et al., Eds., Cambridge University Press, 231–290.
- Hansen, J., R. Ruedy, J. Glascoe, and M. Sato, 1999: GISS analysis of surface temperature change. J. Geophys. Res., 104, 30 997– 31 022, https://doi.org/10.1029/1999JD900835.
- —, —, M. Sato, M. Imhoff, W. Lawrence, D. Easterling, T. Peterson, and T. Karl, 2001: A closer look at United States and global surface temperature change. J. Geophys. Res., 106, 23 947–23 963, https://doi.org/10.1029/2001JD000354.
- —, —, —, and K. Lo, 2010: Global surface temperature change. *Rev. Geophys.*, 48, RG4004, https://doi.org/10.1029/ 2010RG000345.
- 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– 218.
- He, J. F., J. Y. Liu, D. F. Zhuang, W. Zhang, and M. L. Liu, 2007: Assessing the effect of land use/land cover change on the change of urban heat island intensity. *Theor. Appl. Climatol.*, 90, 217–226, https://doi.org/10.1007/s00704-006-0273-1.
- Hijmans, R. J., 2020: raster: Geographic Data Analysis and Modeling, version 3.1-5. R package, https://CRAN.R-project. org/package=raster.
- Hollmann, R., and Coauthors, 2013: The ESA climate change initiative: Satellite data records for essential climate variables. *Bull. Amer. Meteor. Soc.*, 94, 1541–1552, https://doi.org/ 10.1175/BAMS-D-11-00254.1.
- Hu, Y., W. Dong, and Y. He, 2010: Impact of land surface forcings on mean and extreme temperature in eastern China. J. Geophys. Res., 115, D19117, https://doi.org/10.1029/2009JD013368.
- Jones, P. D., and M. Hulme, 1996: Calculating regional climatic time series for temperature and precipitation: Methods and illustrations. Int. J. Climatol., 16, 361–377, https://doi.org/10.1002/(SICI) 1097-0088(199604)16:4<361::AID-JOC53>3.0.CO;2-F.
- —, and D. H. Lister, 2009: The urban heat island in Central London and urban-related warming trends in Central London since 1900. *Weather*, **64**, 323–327, https://doi.org/10.1002/wea.432.
- —, —, and Q. Li, 2008: Urbanization effects in large-scale temperature records, with an emphasis on China. J. Geophys. Res., 113, D16122, https://doi.org/10.1029/2008JD009916.
- —, —, T. J. Osborn, C. Harpham, M. Salmon, and C. P. Morice, 2012: Hemispheric and large-scale land-surface air temperature variations: An extensive revision and an update to 2010. J. Geophys. Res., 117, D05127, https://doi.org/10.1029/ 2011JD017139.
- Kalnay, E., and M. Cai, 2003: Impact of urbanization and land-use change on climate. *Nature*, **423**, 528–531, https://doi.org/ 10.1038/nature01675.
- Karl, T. R., H. F. Diaz, and G. Kukla, 1988: Urbanization: Its detection and effect in the United States climate record. *J. Climate*, 1, 1099–1123, https://doi.org/10.1175/1520-0442(1988)001<1099: UIDAEI>2.0.CO;2.
- Kendall, M. G., 1955: Rank Correlation Methods. Charles Griffin, 196 pp.
- Keywood, M. D., K. M. Emmerson, and M. F. Hibberd, 2016: Ambient air quality: Assessment Summaries. Australia state of

the environment 2016, Australian Government Department of the Environment and Energy, Canberra, accessed 1 December 2020, https://soe.environment.gov.au/theme/ambientair-quality/assessment-summaries.

- Khan, S. S., and M. G. Madden, 2010: A survey of recent trends in one class classification. *Artificial Intelligence and Cognitive Science*, L. Coyle and J. Freyne, Eds., Springer, 188–197.
- —, and —, 2014: One-class classification: Taxonomy of study and review of techniques. *Knowl. Eng. Rev.*, **29**, 345–374, https://doi.org/10.1017/S026988891300043X.
- Klein Tank, A. M. G., and G. P. Können, 2003: Trends in indices of daily temperature and precipitation extremes in Europe, 1946–99. J. Climate, 16, 3665–3680, https://doi.org/10.1175/ 1520-0442(2003)016<3665:TIIODT>2.0.CO;2.
- Li, Q., J. Huang, Z. Jiang, L. Zhou, P. Chu, and K. Hu, 2014: Detection of urbanization signals in extreme winter minimum temperature changes over Northern China. *Climatic Change*, **122**, 595–608, https://doi.org/10.1007/s10584-013-1013-z.
- Li, Y., L. Wang, H. Zhou, G. Zhao, F. Ling, X. Li, and J. Qiu, 2019: Urbanization effects on changes in the observed air temperatures during 1977–2014 in China. *Int. J. Climatol.*, **39**, 251– 265, https://doi.org/10.1002/joc.5802.
- Liu, F. T., K. M. Ting, and Z.-H. Zhou, 2008: Isolation Forest. 2008 Eighth IEEE Int. Conf. on Data Mining (ICDM), Pisa, Italy, IEEE, 413–422.
- —, —, and —, 2012: Isolation-based anomaly detection. ACM Trans. Knowl. Discov. Data, 6 (1), 1–39, https://doi.org/ 10.1145/2133360.2133363.
- Mann, H. B., 1945: Nonparametric tests against trend. *Econ. Soc.*, 13, 245–259, https://doi.org/10.2307/1907187.
- Menne, M. J., C. N. Williams Jr., and R. S. Vose, 2009: The U.S. Historical Climatology network monthly temperature data, version 2. *Bull. Amer. Meteor. Soc.*, **90**, 993–1008, https:// doi.org/10.1175/2008BAMS2613.1.
- —, and Coauthors, 2012a: Global Historical Climatology Network– Daily (GHCN-Daily), version 3.27, accessed 26 February 2020, https://doi.org/10.7289/V5D21VHZ.
- —, I. Durre, R. S. Vose, B. E. Gleason, and T. G. Houston, 2012b: An overview of the Global Historical Climatology Network–Daily database. J. Atmos. Oceanic Technol., 29, 897–910, https://doi.org/10.1175/JTECH-D-11-00103.1.
- Parker, D. E., 2006: A demonstration that large-scale warming is not urban. J. Climate, 19, 2882–2895, https://doi.org/10.1175/ JCLI3730.1.
- —, 2010: Urban heat island effects on estimates of observed climate change. *Wiley Interdiscip. Rev.: Climate Change*, 1, 123–133, https://doi.org/10.1002/wcc.21.
- Patra, S., S. Sahoo, P. Mishra, and S. C. Mahapatra, 2018: Impacts of urbanization on land use/cover changes and its probable implications on local climate and groundwater level. J. Urban Manage., 7, 70–84, https://doi.org/10.1016/j.jum.2018.04.006.
- Pedregosa, F., and Coauthors, 2011: Scikit-learn: Machine learning in Python. J. Mach. Learn. Res., 12, 2825–2830, http:// arxiv.org/abs/1201.0490.
- Peterson, T. C., 2003: Assessment of urban versus rural in situ surface temperatures in the contiguous United States: No difference found. J. Climate, 16, 2941–2959, https://doi.org/ 10.1175/1520-0442(2003)016<2941:AOUVRI>2.0.CO;2.
- —, K. P. Gallo, J. Lawrimore, T. W. Owen, A. Huang, and D. A. McKittrick, 1999: Global rural temperature trends. *Geophys. Res. Lett.*, **26**, 329–332, https://doi.org/10.1029/1998GL900322.
- —, X. Zhang, M. Brunet-India, and J. L. Vázquez-Aguirre, 2008: Changes in North American extremes derived from daily

weather data. J. Geophys. Res., 113, D07113, https://doi.org/ 10.1029/2007JD009453.

- Qian, C., X. Zhang, and Z. Li, 2019: Linear trends in temperature extremes in China, with an emphasis on non-Gaussian and serially dependent characteristics. *Climate Dyn.*, 53, 533–550, https://doi.org/10.1007/s00382-018-4600-x.
- Qian, Y., L. Ruby Leung, S. J. Ghan, and F. Giorgi, 2003: Regional climate effects of aerosols over China: Modeling and observation. *Tellus*, **55B**, 914–934, https://doi.org/10.3402/ tellusb.v55i4.16379.
- Ren, G., and Y. Zhou, 2014: Urbanization effect on trends of extreme temperature indices of national stations over mainland China, 1961–2008. J. Climate, 27, 2340–2360, https://doi.org/ 10.1175/JCLI-D-13-00393.1.
- —, —, Z. Chu, J. Zhou, A. Zhang, J. Guo, and X. Liu, 2008: Urbanization effects on observed surface air temperature trends in north China. J. Climate, 21, 1333–1348, https:// doi.org/10.1175/2007JCLI1348.1.
- —, and Coauthors, 2015: An integrated procedure to determine a reference station network for evaluating and adjusting urban bias in surface air temperature data. J. Appl. Meteor. Climatol., 54, 1248–1266, https://doi.org/10.1175/JAMC-D-14-0295.1.
- Ren, Y., and G. Ren, 2011: A remote-sensing method of selecting reference stations for evaluating urbanization effect on surface air temperature trends. J. Climate, 24, 3179–3189, https:// doi.org/10.1175/2010JCLI3658.1.
- Scikit-learn developers, 2019: Novelty and outlier detection. https://scikit-learn.org/stable/modules/outlier\_detection.html.
- Sen, P. K., 1968: Estimates of the regression coefficient based on Kendall's tau. J. Amer. Stat. Assoc., 63, 1379–1389, https:// doi.org/10.1080/01621459.1968.10480934.
- Squintu, A. A., G. van der Schrier, Y. Brugnara, and A. Klein Tank, 2019: Homogenization of daily temperature series in the European Climate Assessment & Dataset. *Int. J. Climatol.*, 39, 1243–1261, https://doi.org/10.1002/joc.5874.
- Su, C., and Y. Hu, 1988: Cold island effect over oasis and lake. *Chin. Sci. Bull.*, 33, 1023–1026.
- Sun, Y., T. Hu, X. Zhang, C. Li, C. Lu, G. Ren, and Z. Jiang, 2019: Contribution of global warming and urbanization to changes in temperature extremes in Eastern China. *Geophys. Res. Lett.*, 46, 11426–11434, https://doi.org/ 10.1029/2019GL084281.
- Trewin, B. C., 2001: Extreme temperature events in Australia. Ph.D. thesis, The University of Melbourne, 417 pp.
- —, 2013: A daily homogenized temperature data set for Australia. Int. J. Climatol., 33, 1510–1529, https://doi.org/ 10.1002/joc.3530.
- —, and Coauthors, 2020: An updated long-term homogenized daily temperature data set for Australia. *Geosci. Data J.*, 7, 149–169, https://doi.org/10.1002/gdj3.95.
- Tysa, S. K., G. Ren, Y. Qin, P. Zhang, Y. Ren, W. Jia, and K. Wen, 2019: Urbanization effect in regional temperature series based on a remote sensing classification scheme of stations. *J. Geophys. Res. Atmos.*, **124**, 10646–10661, https://doi.org/10.1029/ 2019JD030948.
- Vincent, L. A., and É. Mekis, 2006: Changes in daily and extreme temperature and precipitation indices for Canada over the twentieth century. *Atmos.–Ocean*, 44, 177–193, https://doi.org/ 10.3137/ao.440205.
- —, X. L. Wang, E. J. Milewska, H. Wan, F. Yang, and V. Swail, 2012: A second generation of homogenized Canadian monthly surface air temperature for climate trend analysis. *J. Geophys. Res.*, **117**, D18110, https://doi.org/10.1029/2012JD017859.

- von Storch, H., and A. Navarra, 1999: Analysis of Climate Variability: Applications of Statistical Techniques. Springer-Verlag, 346 pp.
- Wang, F., and Q. Ge, 2012: Estimation of urbanization bias in observed surface temperature change in China from 1980 to 2009 using satellite land-use data. *Chin. Sci. Bull.*, **57**, 1708– 1715, https://doi.org/10.1007/s11434-012-4999-0.
- Wang, X. L., 2008a: Penalized maximal F test for detecting undocumented mean shift without trend change. J. Atmos. Oceanic Technol., 25, 368–384, https://doi.org/10.1175/ 2007JTECHA982.1.
- —, 2008b: Accounting for autocorrelation in detecting mean shifts in climate data series using the penalized maximal t or F test. J. Appl. Meteor. Climatol., 47, 2423–2444, https://doi.org/ 10.1175/2008JAMC1741.1.
- —, and V. R. Swail, 2001: Changes of extreme wave heights in Northern Hemisphere oceans and related atmospheric circulation regimes. J. Climate, 14, 2204–2221, https://doi.org/ 10.1175/1520-0442(2001)014<2204:COEWHI>2.0.CO;2.
- —, and Y. Feng, 2013: RHtests V4 User Manual. Climate Research Division, Atmospheric Science and Technology Directorate. Science and Technology Branch, Environment Canada, 29 pp.
- Wen, K., G. Ren, J. Li, A. Zhang, Y. Ren, X. Sun, and Y. Zhou, 2019: Recent surface air temperature change over mainland China based on an urbanization-bias adjusted dataset. *J. Climate*, 32, 2691–2705, https://doi.org/10.1175/JCLI-D-18-0395.1.
- Xu, Y., W. Xu, Q. Li, and S. Yang, 2014: Report on the development and evaluation of global land daily temperature and precipitation data sets. National Meteorological Information Center of China Meteorological Administration, 21 pp.
- Yan, Z., and Z. Li, Q. li, and P. Jones, 2010: Effects of site change and urbanisation in the Beijing temperature series 1977– 2006. Int. J. Climatol., 30, 1226–1234, https://doi.org/10.1002/ joc.1971.
- Yang, X., Y. Hou, and B. Chen, 2011: Observed surface warming induced by urbanization in east China. J. Geophys. Res., 116, D14113, https://doi.org/10.1029/2010JD015452.
- —, L. R. Leung, N. Zhao, C. Zhao, Y. Qian, K. Hu, X. Liu, and B. Chen, 2017: Contribution of urbanization to the increase of extreme heat events in an urban agglomeration in east China. *Geophys. Res. Lett.*, 44, 6940–6950, https://doi.org/10.1002/ 2017GL074084.
- Zhang, H., and Coauthors, 2012: Simulation of direct radiative forcing of aerosols and their effects on East Asian climate using an interactive AGCM–aerosol coupled system. *Climate Dyn.*, 38, 1675–1693, https://doi.org/10.1007/s00382-011-1131-0.
- Zhang, L., G. Ren, J. Liu, Y. Zhou, Y. Ren, A. Zhang, and Y. Feng, 2011: Urban effect on trends of extreme temperature indices at Beijing Meteorological Station. *Chin. J. Geophys.*, 54, 1150– 1159, https://doi.org/10.3969/j.issn.0001-5733.2011.05.002.
- —, —, Y.-Y. Ren, A.-Y. Zhang, Z.-Y. Chu, and Y.-Q. Zhou, 2014: Effect of data homogenization on estimate of temperature trend: A case of Huairou station in Beijing Municipality. *Theor. Appl. Climatol.*, **115**, 365–373, https://doi.org/10.1007/ s00704-013-0894-0.
- Zhang, P., G. Ren, Y. Xu, X. L. Wang, Y. Qin, X. Sun, and Y. Ren, 2019: Observed changes in extreme temperature over the global land based on a newly developed station daily dataset. *J. Climate*, **32**, 8489–8509, https://doi.org/10.1175/JCLI-D-18-0733.1.
- Zhang, X., L. A. Vincent, W. D. Hogg, and A. Niitsoo, 2000: Temperature and precipitation trends in Canada during the

20th century. Atmos.-Ocean, 38, 395-429, https://doi.org/ 10.1080/07055900.2000.9649654.

- —, L. Alexander, G. C. Hegerl, P. Jones, A. Klein Tank, T. C. Peterson, B. Trewin, and F. W. Zwiers, 2011: Indices for monitoring changes in extremes based on daily temperature and precipitation data. *Wiley Interdiscip. Rev.: Climate Change*, 2, 851–870, https://doi.org/10.1002/wcc.147.
- Zhao, N., Y. Jiao, T. Ma, M. Zhao, Z. Fan, X. Yin, Y. Liu, and T. Yue, 2019: Estimating the effect of urbanization on extreme climate events in the Beijing-Tianjin-Hebei region, China. Sci.

*Total Environ.*, **688**, 1005–1015, https://doi.org/10.1016/j.scitotenv. 2019.06.374.

- Zhou, L., R. E. Dickinson, Y. Tian, J. Fang, Q. Li, R. K. Kaufmann, C. J. Tucker, and R. B. Myneni, 2004: Evidence for a significant urbanization effect on climate in China. *Proc. Natl. Acad. Sci. USA*, **101**, 9540–9544, https://doi.org/10.1073/ pnas.0400357101.
- Zhou, Y., and G. Ren, 2011: Change in extreme temperature event frequency over mainland China, 1961–2008. *Climate Res.*, 50, 125–139, https://doi.org/10.3354/cr01053.