SC-Earth: A Station-Based Serially Complete Earth Dataset from 1950 to 2019

GUOQIANG TANG,^{a,b} MARTYN P. CLARK,^{a,b} AND SIMON MICHAEL PAPALEXIOU^{b,c,d}

^a *University of Saskatchewan Coldwater Laboratory, Canmore, Alberta, Canada*

^b *Centre for Hydrology, University of Saskatchewan, Saskatoon, Saskatchewan, Canada*

^c *Department of Civil, Geological and Environmental Engineering, University of Saskatchewan, Saskatchewan, Canada*

^d *Faculty of Environmental Sciences, Czech University of Life Sciences Prague, Prague, Czech Republic*

(Manuscript received 25 January 2021, in final form 2 April 2021)

ABSTRACT: Meteorological data from ground stations suffer from temporal discontinuities caused by missing values and short measurement periods. Gap-filling and reconstruction techniques have proven to be effective in producing serially complete station datasets (SCDs) that are used for a myriad of meteorological applications (e.g., developing gridded meteorological datasets and validating models). To our knowledge, all SCDs are developed at regional scales. In this study, we developed the serially complete Earth (SC-Earth) dataset, which provides daily precipitation, mean temperature, temperature range, dewpoint temperature, and wind speed data from 1950 to 2019. SC-Earth utilizes raw station data from the Global Historical Climatology Network–Daily (GHCN-D) and the Global Surface Summary of the Day (GSOD). A unified station repository is generated based on GHCN-D and GSOD after station merging and strict quality control. ERA5 is optimally matched with station data considering the time shift issue and then used to assist the global gap filling. SC-Earth is generated by merging estimates from 15 strategies based on quantile mapping, spatial interpolation, machine learning, and multistrategy merging. The final estimates are bias corrected using a combination of quantile mapping and quantile delta mapping. Comprehensive validation demonstrates that SC-Earth has high accuracy around the globe, with degraded quality in the tropics and oceanic islands due to sparse station networks, strong spatial precipitation gradients, and degraded ERA5 estimates. Meanwhile, SC-Earth inherits potential limitations such as inhomogeneity and precipitation undercatch from raw station data, which may affect its application in some cases. Overall, the high-quality and high-density SC-Earth dataset will benefit research in fields of hydrology, ecology, meteorology, and climate. The dataset is available at <https://zenodo.org/record/4762586>.

KEYWORDS: Precipitation; Temperature; Wind; Databases; Machine learning; In situ atmospheric observations

1. Introduction

Ground stations are crucial sources of meteorological observations (New et al. 1999; Schamm et al. 2014; Shen and Xiong 2016; Fick and Hijmans 2017; Harris et al. 2020). Station observations for many environmental variables such as precipitation, temperature, wind, and humidity are used to calibrate and validate estimates from remote sensing and numerical models (Clark et al. 2006; Liu et al. 2018; Tang et al. 2020b; Hersbach et al. 2020). Remote sensing and numerical models often rely on station-based bias correction or assimilation to improve the quality of estimates, and thus have lower quality in periods and regions with few station observations (Tang et al. 2016; Li et al. 2020; Sheridan et al. 2020).

Ground station networks are insufficient in many regions of the world (Donat et al. 2013), caused by installation and maintenance costs and the complexity of topography and

climate (Muller et al. 2013; van de Giesen et al. 2014). Kidd et al. (2017) estimate that working rain gauges globally (~123 000) can cover a total area equivalent to less than half a football field. The number of stations has declined in recent decades and this reduces the capacity to monitor future climate in some regions (Stokstad 1999; Shiklomanov et al. 2002). Moreover, in some regions, not all station data are publicly available due to national restricting policies (Donat et al. 2013; Kidd et al. 2017). Temporal discontinuities, including missing records, missing seasons, poor data quality, and incomplete observation periods, further reduce the number of available stations (Eischeid et al. 2000; Feng et al. 2004; Wang et al. 2017). For example, the Global Historical Climatology Network–Daily (GHCN-D; Menne et al. 2012) provides a total of ~113 000 precipitation stations as of October 2020; the peak annual number of stations is only ~42 000 in 2012 whereas 90% of years before 2020 have fewer than 40 000 stations globally.

Serially complete datasets (SCDs) help to overcome the problem of temporal discontinuity and have proven to be effective in improving the quality of gridded meteorological estimates (Longman et al. 2020). SCDs fill or reconstruct gaps caused by missing values and absent periods to provide serially complete station records for practical applications. Approaches for gap filling can be classified into self-supporting infilling, spatial interpolation, quantile mapping (QM), and machine learning methods (Tang et al. 2020a). Self-supporting infilling imputes missing values based on data from previous and subsequent time steps and is suitable to fill data with high temporal

Denotes content that is immediately available upon publication as open access.

Supplemental information related to this paper is available at the Journals Online website: <https://doi.org/10.1175/JCLI-D-21-0067.s1>.

Corresponding author: Guoqiang Tang, guoqiang.tang@usask.ca

DOI: 10.1175/JCLI-D-21-0067.1

© 2021 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the AMS Copyright Policy (www.ametsoc.org/PUBSReuseLicenses).

autocorrelation (Teegavarapu and Chandramouli 2005; Simolo et al. 2010; Pappas et al. 2014). Spatial interpolation is probably the most common method. Simple interpolation methods such as inverse distance weighting (IDW; Shepard 1968) obtain values at the target station using information from neighboring stations. More sophisticated interpolation methods such as multiple linear regression and kriging can also utilize information from the target station to some extent and are more useful than self-contained infilling or simple interpolation methods (Eischeid et al. 2000; Kanda et al. 2018). QM fills missing data by matching the cumulative density function (CDF) of the target station and surrounding stations (Simolo et al. 2010; Newman et al. 2015, 2019; Devi et al. 2019). Machine learning builds nonlinear relationships between target and reference series, and in some cases are more effective than traditional gap-filling methods (Coulbaly and Evora 2007; Ustaoglu et al. 2008; Dastorani et al. 2010; Wambua et al. 2016; Serrano-Notivol et al. 2019). Gridded meteorological datasets such as the University of Delaware Dataset (Matsuura and Willmott 2017) and Climatic Research Unit gridded time series (CRU TS; Harris et al. 2020) use spatiotemporal infilling and thus are different from the station-based SCDs discussed here.

Many SCDs have been developed as independent datasets in different regions of the world, such as the western United States (Eischeid et al. 2000), northeast Spain (Vicente-Serrano et al. 2003), Italy (Di Piazza et al. 2011), the upper Blue Nile basin (Woldesenbet et al. 2017), and North America (Tang et al. 2020a). SCDs are also generated as intermediate outputs to produce gridded meteorological estimates (Di Luzio et al. 2008; Cornes et al. 2018; Longman et al. 2019, 2020; Newman et al. 2015, 2019). Among the previous studies, the SCD for North America (SCDNA; Tang et al. 2020a) covers the largest domain and provides precipitation and minimum and maximum temperature data. SCDNA combines station observations and three reanalysis products and uses 16 gap-filling strategies to generate daily precipitation and temperature data from 1979 to 2019. However, all previous SCDs, including SCDNA, are at regional or continental scales. Given the worldwide demand for SCDs in research and applications, the absence of a high-quality global SCD is a critical challenge to be solved.

Here we develop a global meteorological SCD, entitled the serially complete Earth dataset (SC-Earth), for five common meteorological variables—precipitation, mean daily temperature (Tmean), daily temperature range (Trange), dewpoint temperature (Tdew), and wind speed—from 1950 to 2019. SC-Earth follows the basic framework of SCDNA with several major updates (details in section 3) such as the first application of the Long-Short Term Memory (LSTM; Hochreiter and Schmidhuber 1997) network in gap filling. The 70-yr SC-Earth dataset will benefit researchers who have a demand for long-term gap-filled SCD data or quality-controlled raw station observations in diverse applications.

2. Datasets

a. Station data

We used station observations from GHCN-D (<https://www.ncdc.noaa.gov/ghcn-d-data-access>; Menne et al. 2012) and

the Global Surface Summary of the Day (GSOD; <https://catalog.data.gov/dataset/global-surface-summary-of-the-day-gsod>).

GHCN-D collects weather reports from numerous sources (Menne et al. 2012) and is the largest station repository of daily precipitation and temperature observations in the world (Donat et al. 2013). GHCN-D data undergo strict quality control including basic integrity checks, outlier checks, internal and temporal consistency checks, spatial consistency checks, and megaconsistency checks (Durre et al. 2010). GHCN-D has been widely used to produce gridded meteorological data (Yatagai et al. 2012; Livneh et al. 2015; Newman et al. 2015; Fick and Hijmans 2017), evaluate/correct existing datasets (Prat and Nelson 2015; Behnke et al. 2016; Beck et al. 2019), study climate changes (Alexander et al. 2006; Utsumi et al. 2011; Donat et al. 2013; Papalexiou and Montanari 2019; Papalexiou et al. 2018), and investigate the probabilistic properties of precipitation extremes (Nerantzaki and Papalexiou 2019; Papalexiou et al. 2013; Papalexiou and Koutsoyiannis 2016; Serinaldi and Kilsby 2014).

GSOD derives daily accumulated data from the Integrated Surface Hourly (ISH) dataset from 1929 to the present. The quality control of GSOD is not as strict as GHCN-D. The number of GSOD stations from 1950 to 2019 is 24 162, which is much fewer than that of GHCN-D. Some GSOD stations have been included in the GHCN-D repository (Menne et al. 2012). However, we still utilize GSOD in this study for two reasons: 1) some GSOD stations are not included in GHCN-D and thus can increase the number of available stations (Beck et al. 2019; Tang et al. 2020a) and 2) GSOD has some unique variables that are not included in GHCN-D. For example, GHCN-D provides wind speed data for some U.S. stations and does not provide Tdew data; in contrast, most GSOD stations provide Tdew and wind speed data.

Most stations record maximum and minimum daily temperature. The mean value and the difference between the two variables are used to compute Tmean and Trange in this study. Some stations, mostly from GSOD, only provide Tmean observations, resulting in a larger number of Tmean than Trange stations in the study. The spatial distributions of GHCN-D precipitation and temperature stations are shown in Figs. S1 and S2, and the distributions of GSOD stations are shown in Fig. S3 (see the online supplemental material). Both datasets start to reach a quasi-global coverage in the 1950s although the density in some regions is relatively low.

GHCN-D and GSOD do not adjust inhomogeneities caused by station relocations, instrument/shelter changes, environment changes, reporting time changes, and so on. Inhomogeneities in station records could result in biased climate trend estimates (Reeves et al. 2007; Vincent et al. 2009; El Kenaawy et al. 2013; Xu et al. 2013). Many homogenization methods have been developed to detect, adjust, and validate the changepoints in station climate series at different time scales (Wang et al. 2007, 2010; Mestre et al. 2011, 2013; Venema et al. 2012). However, most homogenized datasets are developed at the regional scale and with the assistance of metadata (Vincent et al. 2002, 2012; Xu et al. 2013; Li et al. 2016). The pairwise homogenization algorithm (PHA; Menne and Williams 2009) is the only method applied at the global

scale to homogenize monthly average surface air temperature for the GHCN-Monthly. For many regions of the world, the sparsity of stations and lack of metadata could introduce large uncertainties in automatic daily homogenization (Gubler et al. 2017). Therefore, global homogenization of multiple meteorological variables is very challenging and beyond the scope of this study.

b. Reanalysis product

The fifth generation of the European Centre for Medium-Range Weather Forecasts (ECMWF) atmospheric reanalyses of the global climate (ERA5; Hersbach et al. 2020) is used as auxiliary data for global gap filling. Although reanalysis datasets contain uncertainties caused by reasons such as observational constraints and model resolution/parameterization (Donat et al. 2014; Parker 2016), they are useful for gap filling in regions with sparse station networks.

ERA5 provides hourly estimates of precipitation, 2-m minimum and maximum temperature, 2-m Tdew, and 10-m u - and v -component wind speed at the $0.25^\circ \times 0.25^\circ$ resolution from 1950 to the present. Unlike SCDNA that simultaneously uses the ERA5, the Japanese 55-year Reanalysis (JRA-55; Kobayashi et al. 2015), and the Modern-Era Retrospective Analysis for Research and Applications, version 2 (MERRA-2; Gelaro et al. 2017), here we use only the ERA5 due to its long period and high spatiotemporal resolution. Moreover, the ERA5 product was shown to be better than other reanalysis products in the SCDNA study (Tang et al. 2020a). Existing twentieth-century reanalysis products have low spatial resolution and are not used in this study.

3. Methodology

a. Station–reanalysis data match

The spatial match between station and reanalysis data is straightforward. For every station, the nearest ERA5 grid is chosen. The temporal match considers that many stations adopt local reporting time, which usually does not correspond from 0000 to 2400 UTC. To achieve the optimal temporal station–reanalysis match, first, the 70-yr time series (1950–2019) of ERA5 is shifted by $-48, -47, -46, \dots, 0, \dots, 46, 47$, and 48 h. Then, for every station and every shift hour, the correlation coefficient (CC) between accumulated daily ERA5 estimates and station data is calculated. The shift hour with the highest CC is chosen for the final daily ERA5 estimates. This method is the same as Beck et al. (2019) and Yatagai et al. (2020). The number of shift hours is larger for precipitation because it is measured as the accumulation within 24 h (Fig. S4). Shift hours show notable national and regional differences and are different among the five variables (Fig. S4). This method is designed to optimize the use of reanalysis data for gap-filling purposes.

Although this method is suitable to achieve optimal station–reanalysis data match, the inferred shift hours may not represent the actual reporting time of stations because reanalysis estimates contain uncertainties and errors, such as inaccuracies in the diurnal cycle (Tang et al. 2020b) that may result in biased accumulated/averaged values. In addition, Tmean has a strong

autocorrelation. The final shift hour may just have a slightly higher CC than other shift hours, resulting in large uncertainties in shift hour estimation. This can partly explain why Tmean and Trange exhibit different shift hours (Fig. S4) although they are derived from the same variables. Moreover, the inferred shift hours exhibit strong annual variation (Fig. S5), and some stations exhibit shift hours up to or larger than ± 48 h. Due to those concerns, we do not consider possible historical changes of station reporting time such as the nationwide change in July 1961 in Canada (Vincent et al. 2012). Nevertheless, these problems have a limited influence on gap filling because our objective is to find the reference series with the highest correlation with station data, without considering the underlying reasons for the temporal mismatch.

b. Generate a unified station repository

GHCN-D and GSOD are merged by excluding overlapped stations. If the distance between two stations is less than 10 m, we keep the one with the longer period. If the distance is larger than 10 m but smaller than 25 km, the mean absolute difference and CC between the two stations are calculated. The CC value adopts the largest one using shift days of $-1, 0$, and 1 . If the absolute difference is smaller than 0.1 and the CC is larger than 0.9999, the station with a longer period is kept. The thresholds are determined through trial and error. The searching radius is set up to 25 km because the location information for a few old stations might be very different between GHCN-D and GSOD.

After the station merging, stations are quality controlled following the method used in SCDNA (Tang et al. 2020a), including integrity checks, outlier checks, internal and temporal consistency checks, spatial consistency checks, and extreme megaconsistency checks (Durre et al. 2010). Some additional checks are used for precipitation following Beck et al. (2019). Values failing quality control are treated as missing. Then, the time series for every station is divided into different segments if the gap between two observation records is larger than 365 days because some stations are operated in different periods and international data collection in some regions might be intermittent. A temporal segment is excluded if valid samples are fewer than 365 or the ratio of valid samples is smaller than 50%.

The elevation of stations, if missing, is estimated using the Multi-Error-Removed Improved-Terrain (MERIT) digital elevation model (DEM) at a 3-s (~ 90 m at the equator) resolution (Yamazaki et al. 2017). For a station, the DEM value from the closest MERIT grid within a radius of 1 km is used. This 1-km buffer is set to address the potential rounding bias of latitude and longitude information. If no DEM information is available within the searching radius, the station is abandoned.

Finally, stations that pass the quality control are included in the gap-filling procedure if they have more than 3000 valid samples for the whole period and more than 200 valid samples for every month (Tang et al. 2020a).

c. Produce the serially complete dataset

1) OVERALL GAP-FILLING PROCEDURE

The gap-filling procedures generally follow the framework of SCDNA. The SCD estimates are generated for each variable

and each day of the year (days 1–366), separately. The procedure contains nine steps that are briefly described below:

Step 1: Spatiotemporally concurrent ERA5 estimates are extracted for every station and the mean value of ERA5 is scaled to be the same as the station observations.

Step 2: For a target station, neighboring stations (at least 1 and at most 30) are found based on two criteria: an overlapping period of at least 8 years and a distance smaller than 200 km. Neighboring stations are sorted according to their correlation with the target station.

Step 3: The empirical CDFs of the target station, neighboring stations, and ERA5 estimates are obtained using data within a 31-day time window centered by the target day of the year.

Step 4: Estimates are generated using 15 strategies based on quantile mapping with neighboring stations (QMN), quantile mapping with reanalysis products (QMR), interpolation (INT), machine learning (MAL), and multi-strategy merging (MRG). The strategies are introduced in section 3c(2).

Step 5: Steps 3 and 4 are repeated by using 70% of station observations. The remaining 30% of observations are used for independent validation of the 15 strategies. Note that step 5 is only for validation purposes. The final SCD estimates still come from step 4 because models based on 100% observations are expected to be better than those based on 70% observations.

Step 6: The 15 strategies are ranked according to their accuracy metric derived from step 5 (section 3e). The strategy with the highest accuracy is adopted by the final SCD estimates.

Step 7: Estimates from step 6 are corrected to be as close to station observations as possible (section 3d).

Step 8: Since estimates are generated for every day, the estimates are replaced by observations whenever possible to generate the SC-Earth dataset.

Step 9: The SC-Earth dataset is quality controlled again to exclude problematic stations.

The quantile mapping/regression/machine learning models are trained for each station separately. The seasonality has been accounted for in gap filling since the nine steps are implemented for each day of the year. Please refer to section 3c in Tang et al. (2020a) for more details.

2) GAP-FILLING STRATEGIES

Quantile mapping with neighboring stations generates estimates by matching the CDF between the target station and neighboring stations. QMN-1 selects the highest correlated neighboring station with the target station and generates estimates using Eq. (1). QMN-2 is similar to QMN-1 but adopts the correlation-based weighted mean of all neighboring stations. QMN-3 adopts distance-based weighted mean of all neighboring stations. QMN-4 adopts the median of QMN-1 to QMN-3.

$$x_T = F_T^{-1}[F_N(x_N)], \quad (1)$$

where x_N is the observed value of the selected neighboring station, F_N is the empirical CDF of the neighboring station, F_T^{-1}

is the inverse CDF of the target station, and x_T is the estimated value at the target station.

Quantile mapping with reanalysis products (QMR) estimates missing values by matching the CDF between the target station and concurrent ERA5 estimates based on Eq. (1). Interpolations strategies 1 to 3 (INT-1 to INT-3) are based on multiple linear regression using the least absolute deviation criteria (Eischeid et al. 2000; Kanda et al. 2018), normal ratio (Young 1992), inverse distance weighting, respectively. INT-4 is the median of INT-1 to INT-3.

MAL-1 to MAL-3 use artificial neural networks, random forests (Breiman 2001), and LSTM, respectively, which train input features (various combinations of neighboring stations and ERA5 for each day of the year) to the target station series. MAL-4 is the median of MAL-1 to MAL-3. Artificial neural networks and random forests have been used in SCDNA (Tang et al. 2020a). LSTM is a recurrent neural network (RNN). It is a powerful tool for making predictions based on time series data because it learns the dependencies with unknown lags between time steps of a series. To our knowledge, LSTM has never been used in gap filling. Considering that environmental variables (particularly temperature) generally have strong temporal autocorrelation, LSTM is suitable for inferring missing station data. The input feature composes of data from the current day and five antecedent days. A single hidden LSTM layer is used; the number of hidden units is 80; the optimization algorithm is the adaptive moment estimation (AdaM); the mini batch size is set to 10% of total training samples and the maximum epoch number is 200. Although we find that LSTM is more sensitive to hyperparameters than artificial neural networks and random forests, those hyperparameters are constant for all stations because tuning global hyperparameters is too challenging. It is highly possible that LSTM can show better performance in regional studies through appropriate structure adjustment and parameter calibration.

The first multistrategy merging approach (MRG-1) is the weighted mean of the best 3 among 10 independent strategies (QMN-1 to -3, QMR, INT-1 to -3, and MAL-1 to -3), and MRG-2 is the median of the best three. The weight used by MRG-1 is calculated based on the accuracy metric introduced in section 3e.

The final SCD estimates are derived by selecting the best one among the 15 strategies for each day of the year according to the accuracy metric (section 3e). This merging design can improve the accuracy and reduce the spatial cross correlation of SCD estimates. Please refer to Tang et al. (2020a) for a more detailed description of the framework.

3) CORRECTION OF SCD ESTIMATES

Considering that all filling/reconstruction methods have uncertainties, it is useful to correct SCD estimates for each station using raw observations. Actually, the possibility of bias correction is an advantage of station-based gap filling compared to direct spatiotemporal interpolation. The correction method used by SC-Earth is a combination of QM and quantile delta mapping (QDM; Li et al. 2010; Cannon et al. 2015). QDM can preserve the relative or absolute changes in quantiles by adjusting the estimates from QM using the multiplicative

approach for nonnegative variables and the additive approach for other variables. QDM can partly consider the nonstationarity during the 70 years, which is beneficial to the correction of extreme events.

The correction is implemented in three steps for every target station. First, the time series of the target station is divided into different segments using a gap of 5 years, which is longer than that used in quality control (section 3b). Those segments are composed of observation periods with some missing values and blank periods without any observation. Around 25% of stations have at least two observation periods, which are mostly distributed in Asia and Russia (Fig. S6). Second, for every observation period, two empirical CDF curves are obtained: 1) based on station observations and 2) based on SCD estimates. QM-based correction is applied to SCD estimates by matching the two CDF curves since the fraction of missing values is relatively limited. Third, for every period of missing data, the nearest observation period with at least 10-yr valid samples is chosen to build the reference CDF curve, based on which the QDM correction is performed. Let o and e denote observational and estimated data, respectively, and r and b denote reference and blank periods, respectively, the QDM correction is implemented as below:

$$p_{e,b} = F_{e,b}(x_{e,b}), \quad (2)$$

where $F_{e,b}$ is the CDF in the blank period, and $x_{e,b}$ and $p_{o,r}$ is an estimated value and its nonexceedance probability, respectively. The changes in the blank and reference periods are expressed as

$$\begin{cases} \Delta_e^M = \frac{F_{e,b}^{-1}(p_{e,b})}{F_{e,r}^{-1}(p_{e,b})} = \frac{x_{e,b}}{F_{e,r}^{-1}(p_{e,b})} \\ \Delta_e^A = F_{e,b}^{-1}(p_{e,b}) - F_{e,r}^{-1}(p_{e,b}) = x_{e,b} - F_{e,r}^{-1}(p_{e,b}) \end{cases}, \quad (3)$$

where Δ_e^M is the multiplicative expression of the relative change suitable for precipitation, and Δ_e^A is the additive expression of the absolute change suitable for temperature. The bias corrected value $\hat{x}_{e,b}$ is obtained using Eq. (4):

$$\begin{cases} \hat{x}_{e,b} = F_{o,r}^{-1}(p_{e,b})\Delta_e^M \\ \hat{x}_{e,b} = F_{o,r}^{-1}(p_{e,b}) + \Delta_e^A \end{cases}, \quad (4)$$

where $F_{o,r}^{-1}$ is the inverse CDF of station observations in the reference period.

d. Differences between SC-Earth and SCDNA methods

Although SC-Earth generally follows the methodology framework of SCDNA, there are some major differences and improvements summarized as below.

- 1) SCDNA directly estimates minimum and maximum temperature, which are converted to Tmean and Trange in SC-Earth. This is done because constraining positive Trange is straightforward, and some stations only provide Tmean observations.
- 2) SCDNA has four QMR strategies (QMR-1 to 4) based on three reanalysis products, while SC-Earth only uses one QMR strategy based on ERA5 due to the data availability explained in section 2.

- 3) SC-Earth achieves optimal temporal station-reanalysis match by shifting the series of ERA5 estimates (section 3a), which is particularly useful for precipitation. SCDNA only uses the normal 0000–2400 UTC accumulation/average of reanalysis estimates.
- 4) For the preparation of the unified station repository, SC-Earth adopts an additional filtering criterion in quality control by dividing the station series into several segments. SC-Earth also adjusts the station merging strategy, which expands the searching radius and only keeps the station with longer series to identify and exclude overlapped stations more effectively.
- 5) SCDNA applies a downscaling method based on spatiotemporally varied lapse rates from monthly MERRA-2 vertical temperature profiles to account for the elevation difference between target and reference series. SC-Earth directly adjusts the mean value of neighboring stations or reanalysis estimates to be consistent with the target station using multiplicative or additive methods, which is simpler but more effective.
- 6) SC-Earth utilizes LSTM as a new machine learning method to fill gaps in station series.
- 7) SC-Earth adopts a combination of QM and QDM for bias correction to consider the nonstationarity during the 70 years, while SCDNA directly applies QM.
- 8) SCDNA uses the modified Kling-Gupta efficiency (KGE', Gupta et al. 2009; Kling et al. 2012) as the accuracy metric to calculate merging weights and rank different strategies. In this study, we propose another modified version of KGE (KGE'') to avoid the anomalously negative KGE' or KGE values when the mean value is close to zero (Santos et al. 2018).

$$\text{KGE}'' = 1 - \sqrt{(r-1)^2 + (\alpha-1)^2 + \beta^2}, \quad (5)$$

where r is the correlation coefficient between observations (o) and estimates (e), $\alpha = \sigma_e/\sigma_o$ is the variability ratio, $\beta = (\mu_e - \mu_o)/\sigma_o$ is the bias term, μ is the mean value, and σ is the standard deviation. Note that the original KGE uses the bias ratio μ_e/μ_o instead of $\beta = (\mu_e - \mu_o)/\sigma_o$. The range of KGE'' is from $-\infty$ to 1, where one represents perfect agreement between observations and estimates. KGE'' uses σ_o as the denominator because μ_o could be too close to zero in some cases (e.g., dry precipitation stations and temperature stations close to the zero isotherm), while σ_o of environmental variables is often notably larger than zero. The infilled data are generated over both missing and nonmissing periods, and KGE'' is calculated based on raw station observations and infilled estimates on nonmissing days. In section 4, SC-Earth is evaluated in different periods and temporal scales based on KGE'' to demonstrate the performance of SC-Earth from multiple aspects (e.g., overall performance, temporal stability, and seasonal variation).

4. Results

a. Comparison between SC-Earth and raw stations

Gap filling and reconstruction aim to build a SCD with better temporal continuity than raw station observations and

FIG. 1. (left) Monthly series of observed and gap-filled precipitation, Tmean, Trange, Tdew, and wind speed for one station (03749099999 from GSOD) located at 51.15°N, 1.57°W. Estimates shown are before bias correction. (right) A subperiod from 1992 to 1996 corresponding to the shaded area in the left panel. For station observations, a month must have valid samples for at least 25 days to be included. Gap-filled data (green lines) are generated from 1950 to 2019 and may overlap with station observations (red lines) in nonmissing periods.

thus improve station densities at all time steps. Figure 1 shows an example station comparing the monthly curves of SC-Earth estimates and raw observations. This station has incomplete observations from 1995 to 2019 for the five variables. After gap filling and reconstruction, SC-Earth estimates completely cover the period from 1950 to 2019. The daily curves show that station observations contain many gaps that are filled by SC-Earth (Fig. S7). The excellent agreement between raw observations and SC-Earth estimates is attributed to the advantage of SCDs against purely spatial interpolation; that is, the observational information from the target station is used to constrain SCD estimates.

Many precipitation and temperature (Tmean and Trange) stations are located in North America (Fig. 2). For Oceania, the number of precipitation stations is only second to North America, while the number of temperature stations is smaller

than in Europe and Asia. For Europe, temperature stations increase from 1950 to 2019, while precipitation stations increase since 1950 but decrease after 1990. Precipitation and temperature stations in Asia are comparable to or even greater in number than those in Europe before 1970 but become fewer since then. For South America, temperature stations are almost always the lowest among the six continents; precipitation stations show an increasing trend before 1984 but experience two significant drops on 1 January 1984 (from ~4100 to ~2800) and from 1 January 1998 to 1 January 2000 (from ~2600 to fewer than 700). For Africa, temperature stations are just slightly more than those in South America, and the precipitation stations experience an abrupt drop on 1 December 1997 (from ~800 to ~350). Regarding Tdew and wind, station numbers increase in all continents since 1950.

FIG. 2. Number of raw station observations for every day from 1950 to 2019. (bottom right) The numbers of SC-Earth stations, which remain unchanged from 1950 to 2019. All continents are included except Antarctica. The five variables are precipitation (Prpc), Tmean, Trange, Tdew, and wind speed (Wind).

Overall, the numbers of stations in different continents are variable due to factors such as station maintenance and data communication. In contrast, the SC-Earth station repository is much more constant and has higher effective numbers (Fig. 2). For example, the number of daily precipitation samples in North America reaches the peak ($\sim 16\,000$) in 2010 and is lower than 14 000 for most years. In contrast, the number of SC-Earth stations is 28 288 throughout the 1950–2019 period. This improvement is also notable for other variables and continents. The spatial distributions of SC-Earth stations are still uneven in the world (Fig. 3), while the spatiotemporal coverage is better than raw stations (Figs. S1, S2, and S3).

b. Statistical accuracy of SC-Earth

Here we use the KGE'' as the accuracy indicator (section 3e). The mean values of KGE'' based on global stations are 0.88, 0.97, 0.88, 0.97, and 0.85 for precipitation, Tmean, Trange, Tdew, and wind speed, respectively (Fig. 4). All variables show lower accuracy in the tropics and oceanic islands due to the stronger variability, lower density of stations, and lower quality of ERA5

estimates (see section 4d), which all affect the performance of gap filling. The seasonal cycle cannot explain the lower accuracy as the seasonal variation of the KGE'' values is much smaller than that of the KGE'' decrease from extratropics to tropics. The KGE'' for Tmean is much higher and more homogeneous than that for Trange because SCD and reanalysis Trange estimates are worse than Tmean estimates. Tdew (only 12 335 stations globally) and Tmean shows similar KGE'' , indicating that the station density has a limited influence on the gap filling of the two variables. For precipitation, the lowest KGE'' is observed for stations along the boundaries of the Sahara Desert. Some stations in the Tibetan Plateau and central Asia also show relatively low KGE'' . For wind speed, KGE'' is slightly lower than that for precipitation but has a more homogeneous spatial distribution. For example, KGE'' for wind speed in the Sahara, over islands, and on the Tibetan Plateau is comparable to other regions in the world.

SC-Earth estimates before bias correction (Fig. S8) show slightly lower KGE'' than the corrected estimates (Fig. 4). The degradation is more obvious in regions where gap filling performs worse, such as the Sahara and islands for precipitation.

FIG. 3. The global distributions of station densities at the resolution of $2^\circ \times 2^\circ$. The total number of stations is shown at the bottom-left corner of each panel. The five variables are precipitation (Prp), Tmean, Trange, Tdew, and wind speed (Wind).

We also performed independent evaluation by using 70% samples for gap filling and 30% samples for assessment (Fig. S9). Overall, Figs. S8 and S9 are similar, which is consistent with the evaluation of SCDNA estimates (Tang et al. 2020a). According to the three components of KGE'' of SC-Earth before bias correction (Fig. S10), the correlation coefficient shows similar distributions with KGE'' . The variability of precipitation is underestimated particularly in regions with lower station densities due to the smoothing effect of interpolation and machine learning techniques. The variability of Tmean and Tdew is only slightly underestimated in the tropics, whereas for Trange and wind speed the underestimation is more obvious than Tmean and Tdew. The mean value-based bias term is small in most regions and larger in the tropics. The bias of variability and mean value can be effectively eased after bias correction (Tang et al. 2020a). The spatial distributions of root-mean-square error and mean error are shown in Figs. S11 and S12. All variables show low mean errors. The root-mean-square error is larger in tropical regions for precipitation due to the high amount of precipitation and in high-latitude regions for Tmean due to the strong variation and low station density. For Trange and wind, the root-mean-square error also depends on the magnitude of the two

variables and the spatial distributions would be closer to KGE'' after normalization.

Although all the 15 gap-filling strategies contribute to final SC-Earth estimates, their performance among the variables differs. Figure 5 shows that the KGE'' for precipitation from 15 gap-filling strategies. Figures S13 to S16 are for Tmean, Trange, Tdew, and wind speed, respectively. Table 1 summarizes the KGE'' for all variables and strategies. For precipitation, all strategies achieve high KGE'' in regions with high station density such as North America, western Europe, India, Japan, and the eastern and western coasts of Australia. Although all strategies perform worse in regions with fewer stations such as South America, Africa, and central Asia, the degradation of INT-1 and MAL-1 is the most significant, probably because the multiple linear regression and artificial neural network have a high demand for training samples. In contrast, random forest-based MAL-2 and LSTM-based MAL-3 perform much better than MAL-1, indicating that random forest and LSTM are more effective than the artificial neural network in gap filling. QMR shows the lowest mean KGE'' , indicating that ERA5 is not accurate enough for gap filling of precipitation even after the optimal temporal match.

FIG. 4. The distributions of KGE'' for the final SC-Earth precipitation, Tmean, Trange, Tdew, and wind speed estimates. The mean KGE'' value of all stations is shown at the bottom-left corner of each panel.

However, ERA5 enables QMR, MAL, and MRG to generate estimates in remote islands, where QMN and INT cannot generate complete series (Fig. 4). Overall, LSTM is the most effective individual gap-filling strategy for precipitation with a mean KGE'' of 0.75, which agrees with the spatial distribution of best strategy groups in Fig. S17. MRG, particularly MRG-1, shows the highest KGE'' , benefiting from merging multiple strategies.

For Tmean and Tdew, the 15 strategies all achieve high accuracy and are generally comparable. However, LSTM-based MAL-3 is slightly worse than random forest-based MAL-2 for other variables unlike for precipitation. A major reason is that the performance of LSTM is more sensitive to network parameters (e.g., epoch numbers and batch size) than the random forest approach. We adopted a uniform setting for the global training of LSTM, which limits its accuracy for these variables. For regional studies, it is possible to further improve the performance of LSTM in gap filling. Overall, MAL-2 is always one of the best strategies for all variables, indicating that random forest has the advantages of high accuracy and broad applicability in global gap filling.

c. Temporal variations of SC-Earth accuracy

It is important to assess whether the quality of SC-Earth does not change over the 70-yr period. For precipitation,

KGE'' and its three components do not show notable temporal variations (Fig. 6). The lower KGE'' from 2015 to 2019 is mainly caused by the decrease in station numbers (Fig. 2). The variability ratio (Fig. 6c) is generally around 0.94, indicating that SC-Earth slightly underestimates the variability of precipitation. This is expected as interpolation and machine learning methods tend to generate more moderate estimates than extreme estimates. The bias term shows that SC-Earth slightly overestimates precipitation amounts, while this problem has been solved after the mean value adjustment (Tang et al. 2020a). South America, Africa, and Asia show KGE'' below the median level of global stations. Oceania shows the highest KGE'' followed by Europe. For Tmean, Trange, and Tdew (Figs. S18, S19, and S20), KGE'' shows a very weak increasing trend from 1950 to 2019, which can be attributed to the slightly increased station numbers (Fig. 2).

Wind speed is the only variable showing notable temporal variations in the accuracy. The median KGE'' increases from 1950 to 1975 and decreases from 1995 to 2019 (Fig. 7a). This is mainly caused by increasing variability (Fig. 7c) and increasing bias before 2010 and decreasing bias after 2010 (Fig. 7d). Both the variability and bias terms change from underestimation to overestimation during the study period. The dispersion of

FIG. 5. The distributions of KGE'' for precipitation estimates from 15 gap-filling strategies (section 3c). The mean KGE'' value of all stations is shown at the bottom-left corner of each panel.

KGE'' and its three components also becomes larger in recent years than early years. According to KGE'' and its three components for all gap-filling strategies (Fig. S20), QMR, MAL, and MRG strategies show significant temporal variations particularly for the variability and bias terms, which

agrees with the pattern in Fig. 7. The common point among those strategies is the usage of ERA5 wind estimates. For other variables, the most important role of ERA5 is to ensure the completeness of final SC-Earth estimates, and the accuracy of QMR is often lower than most strategies. In contrast, for wind,

TABLE 1. The mean and median KGE'' values of the 15 gap-filling strategies and the final SC-Earth dataset based on global stations. The best strategy (from QMN-1 to MRG-2) is highlighted in boldface font.

	Precipitation		Tmean		Trange		Tdew		Wind	
	Mean	Median	Mean	Median	Mean	Median	Mean	Median	Mean	Median
QMN-1	0.655	0.695	0.925	0.958	0.762	0.802	0.915	0.946	0.668	0.706
QMN-2	0.670	0.705	0.931	0.961	0.778	0.818	0.920	0.951	0.696	0.740
QMN-3	0.688	0.731	0.929	0.960	0.772	0.815	0.918	0.949	0.682	0.730
QMN-4	0.686	0.727	0.931	0.961	0.783	0.824	0.920	0.950	0.697	0.742
QMR	0.565	0.598	0.909	0.947	0.658	0.703	0.913	0.944	0.702	0.755
INT-1	0.645	0.697	0.955	0.984	0.798	0.844	0.948	0.976	0.730	0.774
INT-2	0.686	0.722	0.940	0.973	0.793	0.833	0.932	0.965	0.701	0.740
INT-3	0.696	0.735	0.937	0.973	0.770	0.806	0.928	0.962	0.683	0.719
INT-4	0.685	0.725	0.949	0.980	0.801	0.842	0.938	0.971	0.716	0.756
MAL-1	0.634	0.691	0.915	0.961	0.738	0.806	0.922	0.958	0.680	0.737
MAL-2	0.701	0.749	0.931	0.969	0.791	0.845	0.938	0.967	0.748	0.791
MAL-3	0.752	0.800	0.920	0.970	0.753	0.815	0.934	0.969	0.709	0.763
MAL-4	0.689	0.740	0.923	0.967	0.762	0.824	0.932	0.965	0.710	0.763
MRG-1	0.815	0.846	0.963	0.987	0.840	0.876	0.960	0.984	0.810	0.845
MRG-2	0.780	0.820	0.961	0.987	0.833	0.873	0.959	0.983	0.798	0.839
SC-Earth	0.880	0.905	0.972	0.991	0.878	0.907	0.972	0.988	0.837	0.864

FIG. 6. Temporal variations of (a) KGE'' and (b)–(d) its three components: the correlation coefficient (perfect value: 1), variability term (perfect value: 1), and bias term (perfect value: 0), respectively. The variable is precipitation, and the estimates are before correction. Only stations with at least 50-yr observations are involved. The line within the box is the median. KGE'' is calculated within each 5-yr interval. The upper and lower edges of the box represent the 25th and 75th percentiles, respectively. Values more than 1.5 times the interquartile range away from the upper or lower edges (i.e., vertical dotted error range) are outliers (not shown to be clean). Each box represents a 5-yr period: $\geq$ left bound year and $<$ right bound year. The colored dots show the median value over the six continents.

ERA5-based QMR shows the highest KGE'' for $\sim 40\%$ of stations. Considering that MAL and MRG may also rely on ERA5, the quality of SC-Earth estimates is largely affected by ERA5 wind estimates.

GSOD wind speed data show a significant decreasing trend from 1950 to ~ 2011 and reversely increasing trend in recent years (Zeng et al. 2019), whereas the trend of ERA5 estimates is weak, accompanied by ERA5's overestimation/underestimation of mean wind speed before/after 1973 (Fig. 8). According to the spatial comparison (Fig. S22), ERA5 shows inverse trends with GSOD observations for many regions in the world such as the United States, western Russia, and China. The inconsistency of the wind speed trend between reanalysis models and station observation and between different reanalysis models has also been noted in previous studies (Torralba et al. 2017; Miao et al. 2020; Fan et al. 2021). The reasons could be the quality of the data assimilation methods, vertical extrapolation of wind speed, scale differences between station and reanalysis data, the quality of the numerical weather prediction model used in the reanalysis, and continuous development of wind speed measurement technology, which may affect the homogeneity of data

used by models (Fan et al. 2021). Figure S23 shows the comparison between GSOD and ERA5 annual wind speed for nine example stations. GSOD stations show abnormal drops (i.e., inhomogeneities) of wind speed in different years, which results in extremely large decreasing trends from 1950 to 2019. We also manually inspected the series of GSOD stations in western Russia and China and found that some stations show a similar issue. Therefore, the inhomogeneities in GSOD data and uncertainties in ERA5 estimates jointly result in temporal variations in the performance of SC-Earth wind speed estimates. For other variables (precipitation and temperature), the inconsistency between ERA5 and station data is not significant according to previous regional and global validation studies (Donat et al. 2014; Sheridan et al. 2020; Tang et al. 2020b). Further investigation is needed to understand the reliability of GSOD and ERA5 wind speed for climate studies.

d. Attribution of SC-Earth performance

In addition to the effect of gap-filling methods, many other factors such as station density, climate variability, topographic complexity, and changes in the landscape around stations may

FIG. 7. As in Fig. 6, but for wind speed.

affect the quality of SC-Earth estimates. Those factors are embodied in the closeness between the reference series (i.e., neighboring stations and ERA5) and target series (i.e., observation at the target station). Defining CC_{near} as the mean CC between the target station and its 10 closest neighboring stations, and CC_{ERA5} as the CC between the target station and its concurrent ERA5 estimates, CC_{near} and CC_{ERA5} agree well with the KGE'' of SC-Earth estimates (Fig. 9a).

The deterministic coefficients between CC_{near} and KGE'' and between CC_{ERA5} and KGE'' are 0.698 and 0.599, respectively. A closer relationship between CC_{near} and KGE'' indicates that neighboring stations are better sources of reference series in gap filling of precipitation than ERA5 estimates, even though CC_{near} is generally lower than CC_{ERA5} (Figs. 9b,c). The high CC_{ERA5} is partly attributed to the optimal temporal station–reanalysis match in section 3a,

FIG. 8. (a) Mean value and (b) standard deviation of wind speed for every year from 1950 to 2019. Only GSOD stations with at least 50-yr observations are involved.

FIG. 9. For a target station and precipitation, let CC_{near} be the mean correlation coefficient between the station and its 10 closest neighboring stations, and CC_{ERA5} be the correlation coefficient between the station and its concurrent ERA5 estimates. (a) The scatter density plots between CC_{near} and SC-Earth KGE'' (blue) and CC_{ERA5} and SC-Earth KGE'' (green). The determination coefficients (R^2) between CC and KGE'' are shown at the bottom-right and top-left corners of (a). The density value of a point represents the number of nearby points within the radius of 0.01 KGE'' . (b),(c) The spatial distributions of CC_{near} and CC_{ERA5} , respectively.

which can notably improve the correlation between ERA5 estimates and station data. CC_{near} and CC_{ERA5} both show lower values in the tropics, leading to the lower KGE'' of SC-Earth estimates. ERA5 can have a large contribution to gap filling in regions such as islands, eastern Australia, and high latitudes, where CC_{near} is low.

For Tmean, the correlation between CC_{near}/CC_{ERA5} and KGE'' is much stronger with the deterministic coefficient being ~ 0.85 (Fig. 10). The degradation of CC_{near} and CC_{ERA5} for Tmean in the tropics is more evident compared to precipitation, which explains the worse performance of SC-Earth Tmean estimates in this region. CC_{ERA5} is higher than CC_{near} in some sparsely gauged regions such as the mid/high-latitude islands and the Antarctic where ERA5 can contribute more than ground stations in SC-Earth estimates. Trange, Tdew, and wind speed (Figs. S24, S25, and S26) also show a similar relationship and spatial patterns.

Both CC_{near} and CC_{ERA5} show notable seasonal variations for all variables and both hemispheres (Fig. 11), with higher values in cold seasons and lower values in warm seasons due to the natural variability. Trange, different from other variables, shows an obvious bimodal pattern with the highest CC_{near} and CC_{ERA5} in spring and autumn, particularly in the Northern Hemisphere. The shapes of KGE'' curves

resemble those of CC_{near} and CC_{ERA5} . The seasonal variation of KGE'' is small in magnitude, indicating that the degradation of SC-Earth estimates in warm seasons is not substantial. The terms CC_{near}^* and CC_{ERA5}^* are defined as the CC between KGE'' and CC_{near} , and between KGE'' and CC_{ERA5} , respectively; CC_{near}^* shows subtle seasonal variations in Northern Hemisphere and slight seasonal variation in the Southern Hemisphere, whereas CC_{ERA5}^* shows notable seasonal variations for precipitation and Trange in the Northern Hemisphere. The magnitude of CC^* is a useful indicator of the relative importance of neighboring stations and ERA5 in gap filling. For example, CC_{near}^* is notably larger than CC_{ERA5}^* for precipitation from April to October in the Northern Hemisphere, and for Trange in all cases, indicating that the contribution of neighboring stations is larger than that of ERA5. In contrast, CC_{ERA5}^* is larger than CC_{near}^* for Tdew and wind speed because their station densities are low and thus neighboring stations are not as useful as ERA5 in gap filling. This agrees with the findings of wind speed in section 4c.

5. Summary and conclusions

We developed a global serially complete dataset entitled SC-Earth for the 70-yr period from 1950 to 2019. Raw station data

FIG. 10. As in Fig. 9, but for Tmean.

are assembled from GHCN-D and GSOD and undergo strict quality control. ERA5 estimates are optimally matched with station observations to assist the production of SC-Earth. Fifteen gap-filling strategies based on quantile mapping, spatial interpolation (multiple linear regression, IDW, and normal ratio), machine learning (artificial neural network, random forest, and LSTM), and multistrategy merging are used to generate daily estimates. QM and QDM are used for bias correction of the generated estimates. The final numbers of stations used in the SC-Earth dataset are 64 399, 35 925, 34 851, 12 310, and 12 872 for precipitation, Tmean, Trange, Tdew, and wind speed, respectively.

Tmean and Trange are derived from observed minimum and maximum daily temperature. Tdew and wind speed stations exclusively come from GSOD and thus have lower numbers. The spatial distribution of stations is uneven with many of them located in North America. The lowest numbers of stations are typically observed in Africa and South America, while precipitation stations have sharp drops since 1950, and remain at low levels until the 1990s. Compared to raw station data, SC-Earth effectively increases the station density at all historical periods.

The accuracy of SC-Earth estimates is high around the globe. The mean KGE'' values of all stations are 0.880, 0.972, 0.878, 0.972, and 0.837 for precipitation, Tmean, Trange, Tdew, and wind speed, respectively. The quality of SC-Earth precipitation estimates is lower in regions with low station densities and high climate variabilities such as oceanic islands, tropical regions, the Sahara Desert, and the Tibetan Plateau. The other variables are

also less accurate in tropical regions. Seasonal analyses over the Northern and Southern Hemispheres show that SC-Earth estimates are slightly worse in warm seasons and better in cold seasons, while Trange shows a special bimodal pattern with peak KGE'' occurring in spring and autumn. The performance of SC-Earth estimates has a close relation to the correlation between the target series (i.e., the observations at the station to be filled or reconstructed) and the reference series (i.e., the observations from neighboring stations or concurrent ERA5 estimates). In regions (e.g., tropics and islands) where various factors (e.g., low station density, complex topography, and climate, and degraded ERA5 quality) cause a weak correlation between the target and reference series, SC-Earth estimates could be degraded.

SC-Earth precipitation, Tmean, Trange, and Tdew estimates show stable performance from 1950 to 2019. Wind speed estimates, however, exhibit notable temporal variations in the KGE'' and its three components. There are two possible reasons: 1) ERA5 wind speed estimates have a large contribution to SC-Earth estimates but are characterized by a much smaller decreasing trend of mean value and variability than GSOD data, and 2) some GSOD stations contain large inhomogeneities, resulting in the mismatch between GSOD data and ERA5 estimates. Therefore, users should be cautious when using SC-Earth or GSOD wind speed data.

Since raw station observations are used as the training target and bias correction reference, SC-Earth inherits limitations of station observations, such as inhomogeneity caused by station relocation, instrument, and environment changes (Vincent et al. 2002; Venema et al. 2012). Undercatch of precipitation is

FIG. 11. Seasonal variations of (left) CC_{near} and CC_{ERA5} and (right) KGE'' and CC^* for the Northern and Southern Hemispheres. Rows show (top to bottom) Precipitation, Tmean, Trange, Tdew, and wind speed, respectively. The definitions of CC_{near} and CC_{ERA5} follow Fig. 9. KGE'' represents the accuracy of gap-filled SC-Earth estimates. CC_{near} , CC_{ERA5} , and KGE'' curves use the mean values of all stations. CC_{near}^* and CC_{ERA5}^* represent the correlation coefficient between KGE'' and CC_{near} , and between KGE'' and CC_{ERA5} , respectively.

not corrected, and thus SC-Earth may underestimate precipitation, especially snowfall in windy regions (Yang et al. 2005; Wang et al. 2017). The reporting time of stations varies in different regions, indicating that treating daily precipitation as 0000–2400 UTC accumulation could be biased (Vincent et al. 2009; Yatagai et al. 2020). These problems are important but still challenging for both raw station datasets and SC-Earth. In addition, SC-Earth estimates still contain uncertainties, which is partly revealed by the inconsistency between different gap-filling strategies. Accurate and useful

estimation of uncertainties is a direction for the gap filling and reconstruction of station data.

The SC-Earth dataset offers complete time series for a period of 70 years and it can facilitate regional and global hydrological, meteorological, and climate studies; it is an open access and freely available dataset found at <https://zenodo.org/record/4762586>.

Acknowledgments. This study is funded by the Global Water Futures project. This research was enabled by

High-performance computing (HPC) provided by Copernicus in the University of Saskatchewan and Compute Canada (www.computeCanada.ca). SMP acknowledges the support of the Natural Sciences and Engineering Research Council of Canada (NSERC Discovery Grant: RGPIN-2019-06894). The authors appreciate the extensive efforts from the developers of GHCN-D, GSOD, and ERA5 for making their datasets available.

REFERENCES

- Alexander, L. V., and Coauthors, 2006: Global observed changes in daily climate extremes of temperature and precipitation. *J. Geophys. Res. Atmos.*, **111**, D05109, <https://doi.org/10.1029/2005JD006290>.
- Beck, H. E., E. F. Wood, M. Pan, C. K. Fisher, D. G. Miralles, A. I. J. M. van Dijk, T. R. McVicar, and R. F. Adler, 2019: MSWEP V2 global 3-hourly 0.1° precipitation: Methodology and quantitative assessment. *Bull. Amer. Meteor. Soc.*, **100**, 473–500, <https://doi.org/10.1175/BAMS-D-17-0138.1>.
- Behnke, R., S. Vavrus, A. Allstadt, T. Albright, W. E. Thogmartin, and V. C. Radeloff, 2016: Evaluation of downscaled, gridded climate data for the conterminous United States. *Ecol. Appl.*, **26**, 1338–1351, <https://doi.org/10.1002/15-1061>.
- Breiman, L., 2001: Random forests. *Mach. Learn.*, **45**, 5–32, <https://doi.org/10.1023/A:1010933404324>.
- Cannon, A. J., S. R. Sobie, and T. Q. Murdock, 2015: Bias correction of GCM precipitation by quantile mapping: How well do methods preserve changes in quantiles and extremes? *J. Climate*, **28**, 6938–6959, <https://doi.org/10.1175/JCLI-D-14-00754.1>.
- Clark, M. P., A. G. Slater, A. P. Barrett, L. E. Hay, G. J. McCabe, B. Rajagopalan, and G. H. Leavesley, 2006: Assimilation of snow covered area information into hydrologic and land-surface models. *Adv. Water Resour.*, **29**, 1209–1221, <https://doi.org/10.1016/j.advwatres.2005.10.001>.
- Cornes, R. C., G. van der Schrier, E. J. M. van den Besselaar, and P. D. Jones, 2018: An ensemble version of the E-OBS temperature and precipitation data sets. *J. Geophys. Res. Atmos.*, **123**, 9391–9409, <https://doi.org/10.1029/2017JD028200>.
- Coulibaly, P., and N. D. Evora, 2007: Comparison of neural network methods for infilling missing daily weather records. *J. Hydrol.*, **341**, 27–41, <https://doi.org/10.1016/j.jhydrol.2007.04.020>.
- Dastorani, M. T., A. Moghadamnia, J. Piri, and M. Rico-Ramirez, 2010: Application of ANN and ANFIS models for reconstructing missing flow data. *Environ. Monit. Assess.*, **166**, 421–434, <https://doi.org/10.1007/s10661-009-1012-8>.
- Devi, U., M. S. Shekhar, G. P. Singh, N. N. Rao, and U. S. Bhatt, 2019: Methodological application of quantile mapping to generate precipitation data over northwest Himalaya. *Int. J. Climatol.*, **39**, 3160–3170, <https://doi.org/10.1002/joc.6008>.
- Di Luzio, M., G. L. Johnson, C. Daly, J. K. Eischeid, and J. G. Arnold, 2008: Constructing retrospective gridded daily precipitation and temperature datasets for the conterminous United States. *J. Appl. Meteor. Climatol.*, **47**, 475–497, <https://doi.org/10.1175/2007JAMC1356.1>.
- Di Piazza, A., F. L. Conti, L. V. Noto, F. Viola, and G. La Loggia, 2011: Comparative analysis of different techniques for spatial interpolation of rainfall data to create a serially complete monthly time series of precipitation for Sicily, Italy. *Int. J. Appl. Earth Obs. Geoinf.*, **13**, 396–408, <https://doi.org/10.1016/j.jag.2011.01.005>.
- Donat, M. G., L. V. Alexander, H. Yang, I. Durre, R. Vose, and J. Caesar, 2013: 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>.
- , J. Sillmann, S. Wild, L. V. Alexander, T. Lippmann, and F. W. Zwiers, 2014: Consistency of temperature and precipitation extremes across various global gridded in situ and reanalysis datasets. *J. Climate*, **27**, 5019–5035, <https://doi.org/10.1175/JCLI-D-13-00405.1>.
- Durre, I., M. J. Menne, B. E. Gleason, T. G. Houston, and R. S. Vose, 2010: Comprehensive automated quality assurance of daily surface observations. *J. Appl. Meteor. Climatol.*, **49**, 1615–1633, <https://doi.org/10.1175/2010JAMC2375.1>.
- Eischeid, J. K., P. A. Pasteris, H. F. Diaz, M. S. Plantico, and N. J. Lott, 2000: Creating a serially complete, national daily time series of temperature and precipitation for the western United States. *J. Appl. Meteor.*, **39**, 1580–1591, [https://doi.org/10.1175/1520-0450\(2000\)039<1580:CASCND>2.0.CO;2](https://doi.org/10.1175/1520-0450(2000)039<1580:CASCND>2.0.CO;2).
- El Kenawy, A. E., J. I. López-Moreno, P. Stepanek, and S. M. Vicente-Serrano, 2013: An assessment of the role of homogenization protocol in the performance of daily temperature series and trends: Application to northeastern Spain. *Int. J. Climatol.*, **33**, 87–108, <https://doi.org/10.1002/joc.3410>.
- Fan, W., Y. Liu, A. Chappell, L. Dong, R. Xu, M. Ekström, T.-M. Fu, and Z. Zeng, 2021: Evaluation of global reanalysis land surface wind speed trends to support wind energy development using in situ observations. *J. Appl. Meteor. Climatol.*, **60**, 33–50, <https://doi.org/10.1175/JAMC-D-20-0037.1>.
- Feng, S., Q. Hu, and W. Qian, 2004: Quality control of daily meteorological data in China, 1951–2000: A new dataset. *Int. J. Climatol.*, **24**, 853–870, <https://doi.org/10.1002/joc.1047>.
- Fick, S. E., and R. J. Hijmans, 2017: WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas. *Int. J. Climatol.*, **37**, 4302–4315, <https://doi.org/10.1002/joc.5086>.
- Gelaro, R., and Coauthors, 2017: The Modern-Era Retrospective Analysis for Research and Applications, version 2 (MERRA-2). *J. Climate*, **30**, 5419–5454, <https://doi.org/10.1175/JCLI-D-16-0758.1>.
- Gubler, S., and Coauthors, 2017: The influence of station density on climate data homogenization. *Int. J. Climatol.*, **37**, 4670–4683, <https://doi.org/10.1002/joc.5114>.
- Gupta, H. V., H. Kling, K. K. Yilmaz, and G. F. Martinez, 2009: Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *J. Hydrol.*, **377**, 80–91, <https://doi.org/10.1016/j.jhydrol.2009.08.003>.
- Harris, I., T. J. Osborn, P. Jones, and D. Lister, 2020: Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset. *Sci. Data*, **7**, 109, <https://doi.org/10.1038/s41597-020-0453-3>.
- Hersbach, H., and Coauthors, 2020: The ERA5 global reanalysis. *Quart. J. Roy. Meteor. Soc.*, **146**, 1999–2049, <https://doi.org/10.1002/qj.3803>.
- Hochreiter, S., and J. Schmidhuber, 1997: Long short-term memory. *Neural Comput.*, **9**, 1735–1780, <https://doi.org/10.1162/neco.1997.9.8.1735>.
- Kanda, N., H. S. Negi, M. S. Rishi, and M. S. Shekhar, 2018: Performance of various techniques in estimating missing climatological data over snowbound mountainous areas of Karakoram Himalaya. *Meteor. Appl.*, **25**, 337–349, <https://doi.org/10.1002/met.1699>.
- Kidd, C., A. Becker, G. J. Huffman, C. L. Muller, P. Joe, G. Skofronick-Jackson, and D. B. Kirschbaum, 2017: So, how

- much of the Earth's surface is covered by rain gauges? *Bull. Amer. Meteor. Soc.*, **98**, 69–78, <https://doi.org/10.1175/BAMS-D-14-00283.1>.
- Kling, H., M. Fuchs, and M. Paulin, 2012: Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios. *J. Hydrol.*, **424**–**425**, 264–277, <https://doi.org/10.1016/j.jhydrol.2012.01.011>.
- Kobayashi, S., and Coauthors, 2015: The JRA-55 reanalysis: General specifications and basic characteristics. *J. Meteor. Soc. Japan*, **93**, 5–48, <https://doi.org/10.2151/jmsj.2015-001>.
- Li, H., J. Sheffield, and E. F. Wood, 2010: Bias correction of monthly precipitation and temperature fields from Intergovernmental Panel on Climate Change AR4 models using equidistant quantile matching. *J. Geophys. Res. Atmos.*, **115**, D10101, <https://doi.org/10.1029/2009JD012882>.
- Li, Z., L. Cao, Y. Zhu, and Z. Yan, 2016: Comparison of two homogenized datasets of daily maximum/mean/minimum temperature in China during 1960–2013. *J. Meteor. Res.*, **30**, 53–66, <https://doi.org/10.1007/s13351-016-5054-x>.
- , M. Chen, S. Gao, Z. Hong, G. Tang, Y. Wen, J. J. Gourley, and Y. Hong, 2020: Cross-examination of similarity, difference and deficiency of gauge, radar and satellite precipitation measuring uncertainties for extreme events using conventional metrics and multiplicative triple collocation. *Remote Sens.*, **12**, 1258, <https://doi.org/10.3390/rs12081258>.
- Liu, Z., Y. Liu, S. Wang, X. Yang, L. Wang, M. H. A. Baig, W. Chi, and Z. Wang, 2018: Evaluation of spatial and temporal performances of ERA-Interim precipitation and temperature in mainland China. *J. Climate*, **31**, 4347–4365, <https://doi.org/10.1175/JCLI-D-17-0212.1>.
- Livneh, B., T. J. Bohn, D. W. Pierce, F. Munoz-Arriola, B. Nijssen, R. Vose, D. R. Cayan, and L. Brekke, 2015: A spatially comprehensive, hydrometeorological data set for Mexico, the U.S., and southern Canada 1950–2013. *Sci. Data*, **2**, 150042, <https://doi.org/10.1038/sdata.2015.42>.
- Longman, R. J., and Coauthors, 2019: High-resolution gridded daily rainfall and temperature for the Hawaiian Islands (1990–2014). *J. Hydrometeorol.*, **20**, 489–508, <https://doi.org/10.1175/JHM-D-18-0112.1>.
- , A. J. Newman, T. W. Giambelluca, and M. Lucas, 2020: Characterizing the uncertainty and assessing the value of gap-filled daily rainfall data in Hawaii. *J. Appl. Meteor. Climatol.*, **59**, 1261–1276, <https://doi.org/10.1175/JAMC-D-20-0007.1>.
- Matsuura, K., and C. J. Willmott, 2017: Terrestrial air temperature: 1900–2017 gridded monthly time series. Data are available at http://climate.geog.udel.edu/~climate/html_pages/download.html; the read-me file is at http://climate.geog.udel.edu/~climate/html_pages/Global2017/README.GlobalTsT2017.html.
- Menne, O. J., and C. N. Williams, 2009: Homogenization of temperature series via pairwise comparisons. *J. Climate*, **22**, 1700–1717, <https://doi.org/10.1175/2008JCLI2263.1>.
- , I. Durre, R. S. Vose, B. E. Gleason, and T. G. Houston, 2012: 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>.
- Mestre, O., C. Gruber, C. Prieur, H. Caussinus, and S. Jourdain, 2011: SPLIDHOM: A method for homogenization of daily temperature observations. *J. Appl. Meteor. Climatol.*, **50**, 2343–2358, <https://doi.org/10.1175/2011JAMC2641.1>.
- , and Coauthors, 2013: HOMER: a homogenization software—Methods and applications. *Időjárás*, **117**, 47–67, https://www.researchgate.net/publication/281471961_HOMER_A_homogenization_software_-_methods_and_applications.
- Miao, H., D. Dong, G. Huang, K. Hu, Q. Tian, and Y. Gong, 2020: Evaluation of Northern Hemisphere surface wind speed and wind power density in multiple reanalysis datasets. *Energy*, **200**, 117 382, <https://doi.org/10.1016/j.energy.2020.117382>.
- Muller, C. L., L. Chapman, C. S. B. Grimmer, D. T. Young, and X. Cai, 2013: Sensors and the city: A review of urban meteorological networks. *Int. J. Climatol.*, **33**, 1585–1600, <https://doi.org/10.1002/joc.3678>.
- Nerantzaki, S. D., and S. M. Papalexiou, 2019: Tails of extremes: Advancing a graphical method and harnessing big data to assess precipitation extremes. *Adv. Water Resour.*, **134**, 103448, <https://doi.org/10.1016/j.advwatres.2019.103448>.
- New, M., M. Hulme, and P. Jones, 1999: Representing twentieth-century space–time climate variability. Part I: Development of a 1961–90 mean monthly terrestrial climatology. *J. Climate*, **12**, 829–856, [https://doi.org/10.1175/1520-0442\(1999\)012<0829:RTCSTC>2.0.CO;2](https://doi.org/10.1175/1520-0442(1999)012<0829:RTCSTC>2.0.CO;2).
- Newman, A. J., and Coauthors, 2015: Gridded ensemble precipitation and temperature estimates for the contiguous United States. *J. Hydrometeorol.*, **16**, 2481–2500, <https://doi.org/10.1175/JHM-D-15-0026.1>.
- , M. P. Clark, R. J. Longman, E. Gilleland, T. W. Giambelluca, and J. R. Arnold, 2019: Use of daily station observations to produce high-resolution gridded probabilistic precipitation and temperature time series for the Hawaiian Islands. *J. Hydrometeorol.*, **20**, 509–529, <https://doi.org/10.1175/JHM-D-18-0113.1>.
- Papalexiou, S. M., and D. Koutsoyiannis, 2016: A global survey on the seasonal variation of the marginal distribution of daily precipitation. *Adv. Water Resour.*, **94**, 131–145, <https://doi.org/10.1016/j.advwatres.2016.05.005>.
- , and A. Montanari, 2019: Global and regional increase of precipitation extremes under global warming. *Water Resour. Res.*, **55**, 4901–4914, <https://doi.org/10.1029/2018WR024067>.
- , D. Koutsoyiannis, and C. Makropoulos, 2013: How extreme is extreme? An assessment of daily rainfall distribution tails. *Hydrol. Earth Syst. Sci.*, **17**, 851–862, <https://doi.org/10.5194/hess-17-851-2013>.
- , A. AghaKouchak, K. E. Trenberth, and E. Foufoula-Georgiou, 2018: Global, regional, and megacity trends in the highest temperature of the year: Diagnostics and evidence for accelerating trends. *Earth's Future*, **6**, 71–79, <https://doi.org/10.1002/2017EF000709>.
- Pappas, C., S. M. Papalexiou, and D. Koutsoyiannis, 2014: A quick gap filling of missing hydrometeorological data. *J. Geophys. Res. Atmos.*, **119**, 9290–9300, <https://doi.org/10.1002/2014JD021633>.
- Parker, W. S., 2016: Reanalyses and observations: What's the difference? *Bull. Amer. Meteor. Soc.*, **97**, 1565–1572, <https://doi.org/10.1175/BAMS-D-14-00226.1>.
- Prat, O. P., and B. R. Nelson, 2015: Evaluation of precipitation estimates over CONUS derived from satellite, radar, and rain gauge data sets at daily to annual scales (2002–2012). *Hydrol. Earth Syst. Sci.*, **19**, 2037–2056, <https://doi.org/10.5194/hess-19-2037-2015>.
- Reeves, J., J. Chen, X. L. Wang, R. Lund, and Q. Q. Lu, 2007: A review and comparison of changepoint detection techniques for climate data. *J. Appl. Meteor. Climatol.*, **46**, 900–915, <https://doi.org/10.1175/JAM2493.1>.
- Santos, L., G. Thirel, and C. Perrin, 2018: Technical note: Pitfalls in using log-transformed flows within the KGE criterion. *Hydrol. Earth Syst. Sci.*, **22**, 4583–4591, <https://doi.org/10.5194/hess-22-4583-2018>.

- Schamm, K., M. Ziese, A. Becker, P. Finger, A. Meyer-Christoffer, U. Schneider, M. Schröder, and P. Stender, 2014: Global gridded precipitation over land: A description of the new GPCP First Guess Daily product. *Earth Syst. Sci. Data*, **6**, 49–60, <https://doi.org/10.5194/essd-6-49-2014>.
- Serinaldi, F., and C. G. Kilsby, 2014: Rainfall extremes: Toward reconciliation after the battle of distributions. *Water Resour. Res.*, **50**, 336–352, <https://doi.org/10.1002/2013WR014211>.
- Serrano-Notivol, R., S. Beguería, and M. de Luis, 2019: STEAD: A high-resolution daily gridded temperature dataset for Spain. *Earth Syst. Sci. Data*, **11**, 1171–1188, <https://doi.org/10.5194/essd-11-1171-2019>.
- Shen, Y., and A. Xiong, 2016: Validation and comparison of a new gauge-based precipitation analysis over mainland China. *Int. J. Climatol.*, **36**, 252–265, <https://doi.org/10.1002/joc.4341>.
- Shepard, D., 1968: A two-dimensional interpolation function for irregularly-spaced data. *Proc. 23rd ACM National Conf.*, Association for Computing Machinery, New York, NY, 517–524.
- Sheridan, S. C., C. C. Lee, and E. T. Smith, 2020: A comparison between station observations and reanalysis data in the identification of extreme temperature events. *Geophys. Res. Lett.*, **47**, e2020GL088120, <https://doi.org/10.1029/2020GL088120>.
- Shiklomanov, A. I., R. B. Lammers, and C. J. Vörösmarty, 2002: Widespread decline in hydrological monitoring threatens pan-Arctic research. *Eos, Trans. Amer. Geophys. Union*, **83**, 13–17, <https://doi.org/10.1029/2002EO000007>.
- Simolo, C., M. Brunetti, M. Maugeri, and T. Nanni, 2010: Improving estimation of missing values in daily precipitation series by a probability density function-preserving approach. *Int. J. Climatol.*, **30**, 1564–1576, <https://doi.org/10.1002/joc.1992>.
- Stokstad, E., 1999: Scarcity of rain, stream gages threatens forecasts. *Science*, **285**, 1199–1200, <https://doi.org/10.1126/science.285.5431.1199>.
- Tang, G., Z. Zeng, D. Long, X. Guo, B. Yong, W. Zhang, and Y. Hong, 2016: Statistical and hydrological comparisons between TRMM and GPM level-3 products over a midlatitude basin: Is day-1 IMERG a good successor for TMPA 3B42V7? *J. Hydrometeorol.*, **17**, 121–137, <https://doi.org/10.1175/JHM-D-15-0059.1>.
- , M. P. Clark, A. J. Newman, A. W. Wood, S. M. Papalexiou, V. Vionnet, and P. H. Whitfield, 2020a: SCDNA: A serially complete precipitation and temperature dataset for North America from 1979 to 2018. *Earth Syst. Sci. Data*, **12**, 2381–2409, <https://doi.org/10.5194/essd-12-2381-2020>.
- , —, S. M. Papalexiou, Z. Ma, and Y. Hong, 2020b: Have satellite precipitation products improved over last two decades? A comprehensive comparison of GPM IMERG with nine satellite and reanalysis datasets. *Remote Sens. Environ.*, **240**, 111697, <https://doi.org/10.1016/j.rse.2020.111697>.
- Teegavarapu, R. S. V., and V. Chandramouli, 2005: Improved weighting methods, deterministic and stochastic data-driven models for estimation of missing precipitation records. *J. Hydrol.*, **312**, 191–206, <https://doi.org/10.1016/j.jhydrol.2005.02.015>.
- Torralba, V., F. J. Doblas-Reyes, and N. Gonzalez-Reviriego, 2017: Uncertainty in recent near-surface wind speed trends: A global reanalysis intercomparison. *Environ. Res. Lett.*, **12**, 114019, <https://doi.org/10.1088/1748-9326/aa8a58>.
- Ustaoglu, B., H. K. Cigizoglu, and M. Karaca, 2008: Forecast of daily mean, maximum and minimum temperature time series by three artificial neural network methods. *Meteor. Appl.*, **15**, 431–445, <https://doi.org/10.1002/met.83>.
- Utsumi, N., S. Seto, S. Kanae, E. E. Maeda, and T. Oki, 2011: Does higher surface temperature intensify extreme precipitation? *Geophys. Res. Lett.*, **38**, L16708, <https://doi.org/10.1029/2011GL048426>.
- van de Giesen, N., R. Hut, and J. Selker, 2014: The Trans-African Hydro-Meteorological Observatory (TAHMO). *Wiley Interdiscip. Rev. Water*, **1**, 341–348, <https://doi.org/10.1002/wat2.1034>.
- Venema, V. K. C., and Coauthors, 2012: Benchmarking homogenization algorithms for monthly data. *Climate Past*, **8**, 89–115, <https://doi.org/10.5194/cp-8-89-2012>.
- Vicente-Serrano, S. M., M. A. Saz-Sanchez, and J. M. Cuadrat, 2003: Comparative analysis of interpolation methods in the middle Ebro Valley (Spain): Application to annual precipitation and temperature. *Climate Res.*, **24**, 161–180, <https://doi.org/10.3354/cr024161>.
- Vincent, L. A., X. Zhang, B. R. Bonsal, and W. D. Hogg, 2002: Homogenization of daily temperatures over Canada. *J. Climate*, **15**, 1322–1334, [https://doi.org/10.1175/1520-0442\(2002\)015<1322:HOBTOC>2.0.CO;2](https://doi.org/10.1175/1520-0442(2002)015<1322:HOBTOC>2.0.CO;2).
- , E. J. Milewska, R. Hopkinson, and L. Malone, 2009: Bias in minimum temperature introduced by a redefinition of the climatological day at the Canadian synoptic stations. *J. Appl. Meteor. Climatol.*, **48**, 2160–2168, <https://doi.org/10.1175/2009JAMC2191.1>.
- , 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. Atmos.*, **117**, D18110, <https://doi.org/10.1029/2012JD017859>.
- Wambua, R. M., B. M. Mutua, and J. M. Raude, 2016: Prediction of missing hydro-meteorological data series using artificial neural networks (ANN) for upper Tana River Basin, Kenya. *Amer. J. Water Resour.*, **4**, 35–43.
- Wang, X. L., Q. H. Wen, and Y. Wu, 2007: Penalized maximal t test for detecting undocumented mean change in climate data series. *J. Appl. Meteor. Climatol.*, **46**, 916–931, <https://doi.org/10.1175/JAM2504.1>.
- , H. Chen, Y. Wu, Y. Feng, and Q. Pu, 2010: New techniques for the detection and adjustment of shifts in daily precipitation data series. *J. Appl. Meteor. Climatol.*, **49**, 2416–2436, <https://doi.org/10.1175/2010JAMC2376.1>.
- , H. Xu, B. Qian, Y. Feng, and E. Mekis, 2017: Adjusted daily rainfall and snowfall data for Canada. *Atmos.–Ocean*, **55**, 155–168, <https://doi.org/10.1080/07055900.2017.1342163>.
- Woldesenbet, T. A., N. A. Elagib, L. Ribbe, and J. Heinrich, 2017: Gap filling and homogenization of climatological datasets in the headwater region of the Upper Blue Nile Basin, Ethiopia. *Int. J. Climatol.*, **37**, 2122–2140, <https://doi.org/10.1002/joc.4839>.
- Xu, W., Q. Li, X. L. Wang, S. Yang, L. Cao, and Y. Feng, 2013: Homogenization of Chinese daily surface air temperatures and analysis of trends in the extreme temperature indices. *J. Geophys. Res. Atmos.*, **118**, 9708–9720, <https://doi.org/10.1002/jgrd.50791>.
- Yamazaki, D., and Coauthors, 2017: A high-accuracy map of global terrain elevations. *Geophys. Res. Lett.*, **44**, 5844–5853, <https://doi.org/10.1002/2017GL072874>.
- Yang, D., D. Kane, Z. Zhang, D. Legates, and B. Goodison, 2005: Bias corrections of long-term (1973–2004) daily precipitation data over the northern regions. *Geophys. Res. Lett.*, **32**, L19501, <https://doi.org/10.1029/2005GL024057>.

- Yatagai, A., K. Kamiguchi, O. Arakawa, A. Hamada, N. Yasutomi, and A. Kitoh, 2012: APHRODITE: Constructing a long-term daily gridded precipitation dataset for Asia based on a dense network of rain gauges. *Bull. Amer. Meteor. Soc.*, **93**, 1401–1415, <https://doi.org/10.1175/BAMS-D-11-00122.1>.
- , M. Maeda, S. Khadgarai, M. Masuda, and P. Xie, 2020: End of the day (EOD) judgment for daily rain-gauge data. *Atmosphere*, **11**, 772, <https://doi.org/10.3390/atmos11080772>.
- Young, K. C., 1992: A three-way model for interpolating for monthly precipitation values. *Mon. Wea. Rev.*, **120**, 2561–2569, [https://doi.org/10.1175/1520-0493\(1992\)120<2561:ATWMFI>2.0.CO;2](https://doi.org/10.1175/1520-0493(1992)120<2561:ATWMFI>2.0.CO;2).
- Zeng, Z., and Coauthors, 2019: A reversal in global terrestrial stilling and its implications for wind energy production. *Nat. Climate Change*, **9**, 979–985, <https://doi.org/10.1038/s41558-019-0622-6>.