

## A High-Resolution 1983–2016 $T_{\max}$ Climate Data Record Based on Infrared Temperatures and Stations by the Climate Hazard Center

CHRIS FUNK,<sup>a,b</sup> PETE PETERSON,<sup>b</sup> SETH PETERSON,<sup>b</sup> SHRADDHANAND SHUKLA,<sup>b</sup> FRANK DAVENPORT,<sup>b</sup> JOEL MICHAELSEN,<sup>b</sup> KENNETH R. KNAPP,<sup>c</sup> MARTIN LANDSFELD,<sup>b</sup> GREGORY HUSAK,<sup>b</sup> LAURA HARRISON,<sup>b</sup> JAMES ROWLAND,<sup>a</sup> MICHAEL BUDDE,<sup>a</sup> ALEX MEIBURG,<sup>b</sup> TUFA DINKU,<sup>d</sup> DIEGO PEDREROS,<sup>a</sup> AND NICHOLAS MATA<sup>a</sup>

<sup>a</sup> Earth Resources Observation and Science Center, U.S. Geological Survey, Sioux Falls, South Dakota

<sup>b</sup> Climate Hazards Center, University of California, Santa Barbara, Santa Barbara, California

<sup>c</sup> NOAA/National Centers for Environmental Information, Asheville, North Carolina

<sup>d</sup> International Research Institute for Climate and Society, Columbia University, Palisades, New York


(Manuscript received 18 October 2018, in final form 30 May 2019)


### ABSTRACT

Understanding the dynamics and physics of climate extremes will be a critical challenge for twenty-first-century climate science. Increasing temperatures and saturation vapor pressures may exacerbate heat waves, droughts, and precipitation extremes. Yet our ability to monitor temperature variations is limited and declining. Between 1983 and 2016, the number of observations in the University of East Anglia Climatic Research Unit (CRU)  $T_{\max}$  product declined precipitously (5900  $\rightarrow$  1000); 1000 poorly distributed measurements are insufficient to resolve regional  $T_{\max}$  variations. Here, we show that combining long (1983 to the near present), high-resolution (0.05°), cloud-screened archives of geostationary satellite thermal infrared (TIR) observations with a dense set of  $\sim 15\,000$  station observations explains 23%, 40%, 30%, 41%, and 1% more variance than the CRU globally and for South America, Africa, India, and areas north of 50°N, respectively; even greater levels of improvement are shown for the 2011–16 period (28%, 45%, 39%, 52%, and 28%, respectively). Described here for the first time, the TIR  $T_{\max}$  algorithm uses subdaily TIR distributions to screen out cloud-contaminated observations, providing accurate (correlation  $\approx 0.8$ ) gridded emission  $T_{\max}$  estimates. Blending these gridded fields with  $\sim 15\,000$  station observations provides a seamless, high-resolution source of accurate  $T_{\max}$  estimates that performs well in areas lacking dense in situ observations and even better where in situ observations are available. Cross-validation results indicate that the satellite-only, station-only, and combined products all perform accurately ( $R \approx 0.8$ – $0.9$ , mean absolute errors  $\approx 0.8$ – $1.0$ ). Hence, the Climate Hazards Center Infrared Temperature with Stations (CHIRTS<sub>max</sub>) dataset should provide a valuable resource for climate change studies, climate extreme analyses, and early warning applications.

## 1. Introduction and background

Increased temperature extremes are expected to be one major climate change impact (Diffenbaugh et al. 2017), contributing to disruptions of crop production

 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-18-0698.s1>.

Corresponding author: Chris Funk, cfunk@usgs.gov

both in highly productive regions in the Northern Hemisphere and in marginal farming areas of the tropics (Brown et al. 2015). By increasing saturation vapor pressures, temperature increases will also exacerbate droughts (Trenberth et al. 2014) and precipitation extremes (Trenberth et al. 2003). While there are well-developed systems for organizing, quality controlling, and gridding station data, such as those supported by the monthly Global Historical Climate Network (GHCN), version 3, archive (Peterson and Vose 1997), GHCN daily archive (Menne et al. 2012), the University of East Anglia's Climatic Research Unit (CRU) Time Series, version 4.01 (Harris et al. 2014), and the Berkeley Earth project (Rohde et al. 2013a,b), there are no global 2-m

DOI: 10.1175/JCLI-D-18-0698.1

© 2019 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy](#) (<http://www.ametsoc.org/PUBSReuseLicenses>).

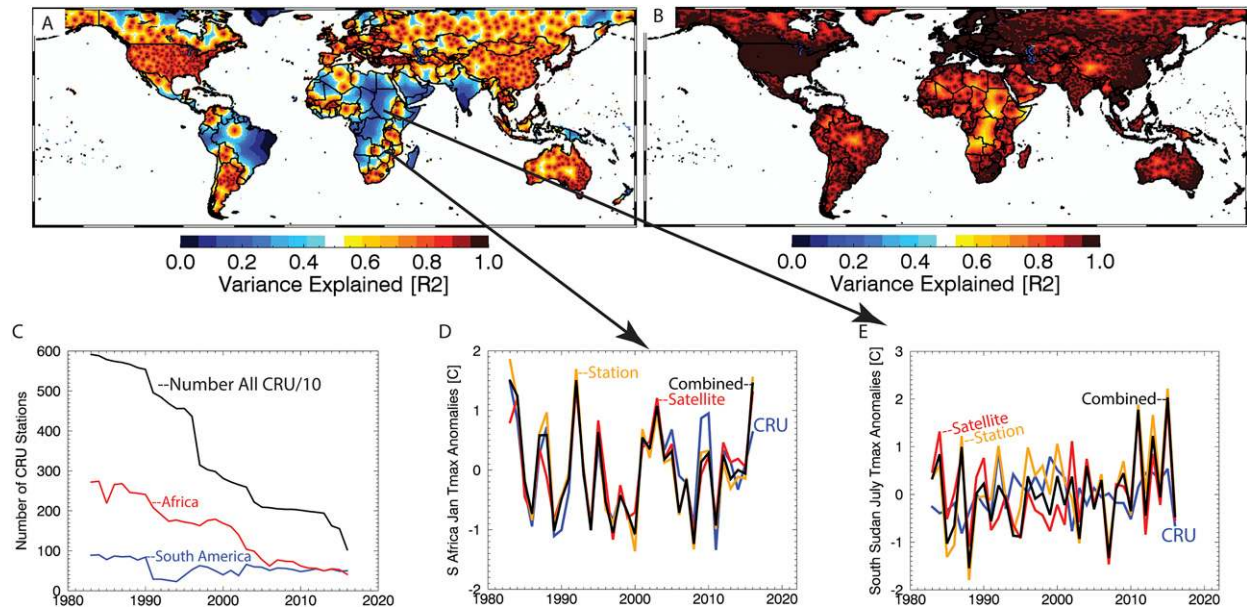


FIG. 1. Motivation for the development of the CHIRTS<sub>max</sub> dataset. (a) Estimated variance explained by CRU stations for January 2016. Mean  $R^2 = 0.61$ . (b) As in (a), but for CHIRTS<sub>max</sub>. Mean  $R^2 = 0.85$ . (c) Total number of January CRU station observations for the globe, Africa, and South America. (d) Time series of January  $T_{\max}$  anomalies for eastern southern (Zambia, Malawi, Mozambique, Botswana, eastern South Africa). (e) July  $T_{\max}$  anomalies for South Sudan and southern Sudan.

maximum temperature ( $T_{\max}$ ) products that directly combine satellite and station-based estimates of  $T_{\max}$  to produce routinely updated data to support the monitoring of temperature extremes. Here, we present such a product—the new Climate Hazards Center Infrared Temperature with Stations (CHIRTS<sub>max</sub>) Climate Data Record (CDR). Monthly maximum temperatures are the focus of this study because isolating a terrestrial  $T_{\max}$  signal from potential cloud contamination was easier than isolating minimum temperatures, which are harder to distinguish from clouds. Here, we focus on the 1983–2016 time frame, comparing our results with the widely used CRU archive. We stop the analysis in 2016 because that is the end of CRU archive. If we can produce a dataset similar to the CRU, with regularly available inputs with good spatial coverage, then updates to this dataset can be used to support early warning activities and the monitoring of temperature extremes. A similar approach was taken in the development of the Climate Hazards Center’s (CHC) Climate Hazards Infrared Precipitation with Stations (CHIRPS) rainfall product (Funk et al. 2015c). During the design stage, CHIRPS was shown to be similar to a gold standard dataset, the Global Precipitation Climatology Centre (GPCC) archive. CHIRPS is now regularly updated (<https://chc.ucsb.edu/data/chirps/>) and serves a wide community of end users. Here, we focus on describing the CHIRTS<sub>max</sub> algorithm and comparing the CHIRTS<sub>max</sub> with the CRU Time Series 4.01 product (CRU TS 4.01).

Since the origin of modern climate science in the nineteenth century, accurate data have been fundamental to advancing our understanding of the dynamics and physics of climate. In a rapidly warming world, the need for accurate data is increasing. For example, while there is an expectation that human-induced warming will likely exacerbate the impacts of rainfall deficits (Trenberth et al. 2014), data support for such evaluations is sparse. The World Climate Research Programme’s Grand Challenge on Weather and Climate Extremes (Zhang et al. 2013) seeks to understand the relative roles of large-scale, regional-, and local-scale processes, as well as their interactions, for the formation of extremes. This challenge also examines the contributors to observed extreme events and to changes in the frequency and intensity of the observed extremes. Unfortunately, sparse station networks means that gridded datasets like the CRU have large regions where little information is available. As motivation for this paper, Fig. 1a uses an empirical covariogram<sup>1</sup> and the distribution of CRU stations in January 2016 to estimate the variance explained across the globe. The mean covariance is 0.61, and many areas have variance-explained values of less than 0.5. Figure 1b shows similar estimates for our CHIRTS<sub>max</sub> product. A much more complete set of station observations combined with an independent

<sup>1</sup> Section 5d explains and expands these results.

satellite-based temperature estimate (i.e., CHIRTS<sub>max</sub>) performs much better (mean  $R^2$  of 0.91), which means analysis in Brazil, Africa, and India is now feasible.

The CRU product faces a substantial limitation arising from the “reporting crisis”—decaying observation networks combined with more restrictive data-sharing policies has resulted in a ~600% decline in total CRU stations between 1983 and 2016 (Fig. 1c). The numbers of CRU stations in Africa and South America have declined from 272 and 89 to 41 and 50, respectively. Independent satellite-based  $T_{\max}$  estimates can help overcome station data gaps while providing independent corroboration. Figures 1d and 1e show time series of  $T_{\max}$  data from South Sudan and southern Africa. Extreme temperature anomalies in these extremely food-insecure regions in 2016 and 2015 were not captured by the CRU, but were identified by both the station and satellite components of CHIRTS<sub>max</sub>. The satellite component is the first of its kind. While there are many global satellite-based precipitation datasets, we present here, for the first time, a global satellite-based temperature analog. While this  $T_{\max}$  dataset uses similar inputs (thermal infrared geostationary satellite observations), the algorithm is completely different from those used to estimate precipitation, since the goal is to sense the land, not the clouds or hydrometeors.

The CHIRTS<sub>max</sub> algorithm can be seen as an extension of traditional station-based climate gridding approaches. The CRU and Berkeley Earth processes begin by merging numerous in situ datasets from international and national data sources. These station data are quality controlled and checked for homogeneity. The CRU archive is gridded based on the climate anomaly method (CAM) (Peterson et al. 1998), which is similar to climatologically aided interpolation (CAI) (Willmott and Robeson 1995). The CAM process estimates the station mean over the 1961–90 time period, and then interpolates the arithmetic temperature anomalies. Stations without long periods of record are excluded, dramatically reducing the number of available observations in many developing countries. The Berkeley Earth process is substantially different (Rohde et al. 2013a,b). The data screening process is much less exclusive; questionable stations are identified and assigned low weights in an iterative geostatistical kriging process rather than being removed. Individual station time series are also decomposed into multiple baseline periods with accompanying anomalies, with the latter being interpolated (kriged).

The monthly CRU interpolation process for the CRU TS 4.01 begins by identifying all 2.5° grid cells with no nearby stations—defined as stations whose expected interstation correlations are not significant. These gridcell

centroids are populated with “dummy stations” with an assumed anomaly value of “0.” These values, along with “real” station anomaly values, are then interpolated using angular distance weighting as described in (New et al. 2000). The interpolated CRU anomalies are then combined with a high-resolution climatology (New et al. 2002) based on thin-plate smoothing splines (ANUSPLIN) (Hutchinson 1995). Thin-plate splines minimize the roughness of the interpolated surface while also seeking to minimize the at-station residuals from the fitted surface. The CRU climatology is based on a trivariate spline fit to latitude, longitude, and elevation.

Unfortunately, in many parts of the world, the utility of systems like the CRU interpolation scheme is limited by extremely low numbers of stations provided to public agencies such as the World Meteorological Organization (WMO). It is common to have large countries with only a few or no air temperature stations. For example, in January of 2016, Brazil, Peru, the Democratic Republic of Congo, Namibia, Kenya, Ethiopia, and India contained one or no CRU observations. This absence of stations limits our ability to monitor temperature anomalies and extremes in many parts of the world.

One valuable means of overcoming these data gaps is by using satellite information. Thermal infrared (11  $\mu\text{m}$ ) brightness temperatures have been used to estimate precipitation (Huffman et al. 1995; Xie and Arkin 1997), especially in tropical and subtropical regions where the frequency of very cold cloud tops tends to be a robust indicator of precipitation (Arkin and Meisner 1987). Simple linear relationships like the global precipitation index (Arkin and Meisner 1987), or more sophisticated approaches based on multisensor assimilation techniques (Adler et al. 2003; Hong et al. 2004; Huffman et al. 2009, 2007; Sorooshian et al. 2000), perform reasonably well in tropical and subtropical regimes. Global precipitation estimation has been greatly facilitated by the creation of merged intercalibrated geostationary thermal infrared (TIR) archives such as the Gridded Satellite (GridSat) archive produced by the National Centers for Environmental Information (Knapp et al. 2011), or the Climate Prediction Center multisatellite archive (Janowiak et al. 2001). While Advanced Very High Resolution Radiometer satellite retrievals are routinely used to estimate sea surface temperatures over the oceans (Reynolds et al. 2002), the application of polar orbiting or geosynchronous TIR observations to terrestrial temperature estimates has been less common.

Under clear-sky conditions, the at-sensor thermal energy received by a satellite will be a function of the surface emissivity and temperature modified by the local profile of atmospheric water vapor. The combined

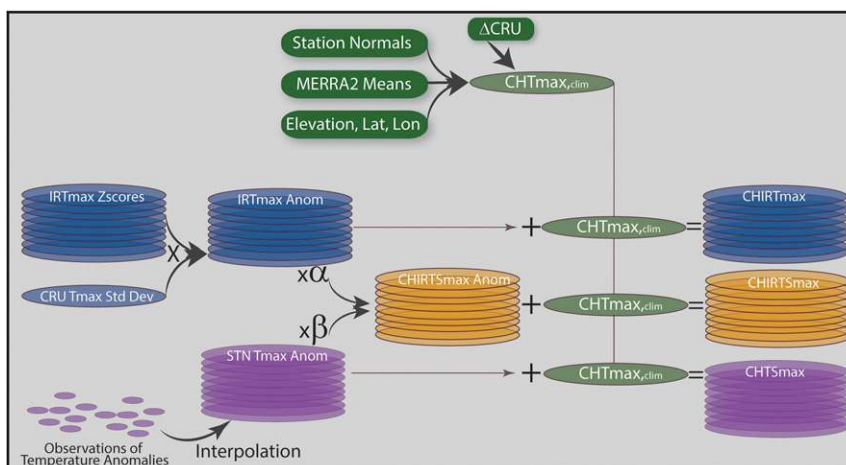


FIG. 2. Schema of  $\text{CHIRT}_{\text{max}}$ ,  $\text{CHIRTS}_{\text{max}}$ , and  $\text{CHTS}_{\text{max}}$  process. All data are at a monthly time scale. The gridded data use a common quasi-global  $0.05^\circ$  grid.

effects of emissivity and atmosphere make retrievals of land surface temperatures challenging (Li et al. 2013), and one of the most common approaches uses multiband optical observations to simultaneously estimate land surface temperatures (LST) and emissivities (Wan et al. 2004). Even once LST is known, it may have a complex relationship to 2-m air temperatures, since 2-m air temperatures are also affected by the specific details of the local atmospheric circulation. For example, over the area of Mount Kilimanjaro, there can be large discrepancies between Moderate Resolution Imaging Spectroradiometer (MODIS) LST and in situ 2-m air temperature observations (Pepin et al. 2016).

Despite these challenges, using TIR observations to guide temperature anomaly estimates in regions with few station observations could provide valuable information. Assuming a lack of cloud contamination, TIR observations have the distinct benefit of being just that—observations directly related to surface emission temperatures. Furthermore, if water vapor and emissivity effects are fairly constant in time (for a given month), the influence of these factors may be at least partially accounted for by working with infrared temperature (IRT) anomalies and scaling the variance of these anomalies using observed  $T_{\text{max}}$  standard deviations.

Another useful avenue for estimating air temperatures without station observations is through reanalysis systems, such as the Modern-Era Retrospective Analysis for Research and Applications, version 2 (MERRA-2). The MERRA-2 (Weller et al. 2016) provides physically consistent estimates of atmospheric and land surface conditions using numerical models constrained with a large suite of observed conditions, including a large number of satellite-based inputs such as radiances and satellite retrievals of

temperature profiles (Gelaro et al. 2017). We assessed the suitability of MERRA-2, as described below, and found it to be generally accurate. Ultimately, we did not use it in this study because we found several instances in which there were large, unexpected shifts in the  $T_{\text{max}}$  values estimated by the MERRA-2 system. More specifically, in data-sparse regions like Africa and South America, the precipitation assimilation process (Reichle et al. 2017) can induce large changes in local-surface energy balances and associated near-surface temperatures (Draper et al. 2018).

Following, we present a set of monthly 1983–2016  $T_{\text{max}}$  estimates based on a combination of IRT and interpolated monthly  $T_{\text{max}}$  station observations. These estimates are compared with monthly CRU TS 4.01 data. Three products are developed: a blended satellite-interpolated station dataset ( $\text{CHIRTS}_{\text{max}}$ ), a satellite-only Climate Hazards Center Infrared Temperature dataset ( $\text{CHIRT}_{\text{max}}$ ), and a station-only  $\text{CHTS}_{\text{max}}$ , but our validation focuses most closely on comparing  $\text{CHIRTS}_{\text{max}}$  with the CRU TS 4.01. Since the components used in  $\text{CHIRTS}_{\text{max}}$  should be able to be updated with relatively low latencies, a  $\text{CHIRTS}_{\text{max}}$  product similar to the CRU TS 4.01 will be a major benefit for analyzing temperature extremes and associated impacts, especially in data-sparse areas.

## 2. Overview

The  $\text{CHIRTS}_{\text{max}}$  development process (Fig. 2) broadly follows the data development strategy used to develop our CHIRPS 1981 to the present precipitation dataset (Funk et al. 2015c). The basic idea behind CHIRPS was to produce a dataset suitable for monitoring hydrologic extremes in areas with limited in situ observations.

This approach begins with a space–time partitioning similar to that used in CAI and CAM. The long-term (1981–2010) mean and time-varying monthly anomalies are estimated separately and then combined. For the  $\text{CHIRTS}_{\text{max}}$  long-term mean, we build a high-resolution ( $0.05^\circ$  latitude–longitude grid) global climatology based on climate normals (long-term monthly means), physiographic indicators (latitude, longitude, and elevation), and monthly average MERRA-2 (Weller et al. 2016) 2-m maximum air temperature estimates. Rather than use the tridimensional ANUSPLIN approach used by the CRU, we use a newer technique, moving window regression (Funk et al. 2015a), which builds localized regressions between the long-term monthly means of a given station and our predictors: MERRA-2 means, elevation, latitude, and longitude. This process produces the monthly  $0.05^\circ$  Climate Hazards Center  $T_{\text{max}}$  climatology ( $\text{CHT}_{\text{clim}}$ ).

The  $\text{CHT}_{\text{clim}}$  is then combined with two independent sets of 2-m  $T_{\text{max}}$  air temperature anomalies to produce the three different sets of  $T_{\text{max}}$  estimates: satellite-only  $\text{CHIRT}_{\text{max}}$ , station-only  $\text{CHTS}_{\text{max}}$  and blended satellite–station  $\text{CHIRTS}_{\text{max}}$ . While we anticipate that many users will be primarily interested in  $\text{CHIRTS}_{\text{max}}$ , the station and satellite-only products allow for independent analysis of extremes. In regions with dense observational networks (e.g., the United States and Europe), the station-only  $\text{CHTS}_{\text{max}}$  may be preferable because (i) stations provide a direct measurement of  $T_{\text{max}}$ , and (ii) our  $\text{CHIRT}_{\text{max}}$  estimates are based on intercalibrated TIR values from multiple satellites.

The first source of anomalies is derived from TIR-based monthly  $T_{\text{max}}$  anomalies. The TIR-based monthly  $T_{\text{max}}$  values are translated into standardized anomalies, and then converted into units of 2-m air temperature anomalies by multiplying these fields by observed standard deviations (which are derived from the 1983–2016 CRU TS 4.01). A separate set of estimates is derived from interpolated, quality-controlled station observations obtained from the Berkeley Earth project (<http://berkeleyearth.org>) and augmented by the Global Telecommunication System (GTS) archive. These station data are interpolated with a spherical inverse distance weighting procedure (Raskin et al. 1997). Then the TIR-based and interpolated station-based anomalies are blended based on their expected contributions to explaining variance at a given location, which is based on evaluations using station data. The expected covariance for the station-based estimates is derived from empirical covariograms that quantify the expected distance decay function. The covariance for the  $\text{CHIRT}_{\text{max}}$  component is based on the observed correlation between the station

observations and collocated  $\text{CHIRT}_{\text{max}}$  values. The blended TIR/station anomalies are then added to the  $\text{CHT}_{\text{clim}}$  to produce time-varying  $\text{CHIRTS}_{\text{max}}$  estimates. We describe the process below and conclude with a brief comparison of the  $\text{CHIRTS}_{\text{max}}$ ,  $\text{CHIRT}_{\text{max}}$ ,  $\text{CHTS}_{\text{max}}$ , MERRA-2, and CRU datasets.

### 3. Data

#### a. The geostationary thermal infrared satellite archive

The GridSat B1 dataset (Knapp et al. 2011) provides the key input to the time-varying  $\text{CHIRT}_{\text{max}}$ ,  $\text{CHIRTS}_{\text{max}}$  datasets. GridSat is composed of intercalibrated thermal infrared ( $11\ \mu\text{m}$ ) brightness temperatures. These Earth observations are drawn from the “geostationary quilt” composed of a suite of geostationary satellites that provide relatively continuous coverage of most of the Earth’s surface since the early 1980s. We begin the  $\text{CHIRTS}_{\text{max}}$  dataset in 1983—the onset of fairly complete coverage. The GridSat archive provides two sets of intercalibrated brightness temperatures—the “best” set of observations, defined as the satellite observations from the most nadir looking geostationary satellite, and a “next best” set of observations. The “next best” set of observations are those from the satellite that is second nearest to each location. In general, we have used the most nadir looking geostationary satellite with one exception. When available, we use observations from the *Meteosat-7* satellite. Launched in September of 1997, *Meteosat-7* was originally located at the prime meridian ( $0^\circ\text{E}$ ) and then moved to  $63^\circ\text{E}$ , over the Indian Ocean, in July of 2006. By selecting *Meteosat-7* observations when they are available, and even when newer observations from *Meteosat-8* or *Meteosat-9* are available, we obtain a relatively stable, long-term satellite record for parts of Africa and Eurasia. The GridSat archive provides 8 instantaneous 3-hourly TIR observations every day at  $\sim 7\text{-km}$  resolution. These data are resampled to a  $0.5^\circ$  grid spanning  $60^\circ\text{S}$ – $70^\circ\text{N}$ .

To minimize the effects of shifts between satellites, the monthly mean CRU and TIR-based  $T_{\text{max}}$  values were calculated at each location for each satellite in the GridSat geostationary quilt. The TIR-based  $T_{\text{max}}$  anomalies were then adjusted by a linear shift to align with the CRU anomalies. Both the *Meteosat-5* and *Meteosat-7* observations were divided into two different “satellites” for this correction procedure. One CRU correction was applied when *Meteosat-5* and *Meteosat-7* orbited over the prime meridian ( $0^\circ\text{E}$ ), and another correction was used when they were moved to orbits over the Indian Ocean.

### b. MERRA-2

We compare CHIRTS<sub>max</sub> to MERRA-2 (Gelaro et al. 2017) 2-m  $T_{\max}$  fields. We also use MERRA-2  $T_{\max}$  estimates as a background field when developing our  $T_{\max}$  climatology. The MERRA-2 reanalysis incorporates a large number of satellite-based inputs, including radiances and satellite retrievals of temperature profiles (Gelaro et al. 2017). MERRA-2 also incorporates in situ station observations from the National Centers for Environmental Prediction, as well as a large number of satellite observations of surface and cloud radiance values and satellite sounder-based temperature profiles. Radiosonde and other sources of information are also utilized. These information sources are then assimilated using the physically based framework provided by the Goddard Earth Observing System model.

### c. Datasets used to derive the long-term $T_{\max}$ climatology

This work builds on multiple sources of station observations. In general, more station data are available in the form of long-term averages (climate normals) than time-varying monthly data. To guide the development of our long-term mean field, we used a fairly dense (~12 300 stations) set of long-term  $T_{\max}$  averages obtained from the Food and Agricultural Organization (FAO) Loc Clim Package (Raymo et al. 1996). We also utilized elevation calculated from global 30-arc-s elevation dataset (GTOPO30) grids, following the methodology developed for the HYDRO1K (Verdin and Greenlee 1996). We transformed these data by adding the minimum elevation and then taking the square root. We used this transform because without it, very high elevations ended up being unrealistically cold. Weather stations are almost always located in valleys, which made it difficult for our statistical model to capture perfect lapse rate relationships in data-sparse mountainous areas (like the Himalayas). Latitude and longitude were also included as potential predictor variables. In addition to location and elevation, MERRA-2 (Gelaro et al. 2017) 2-m maximum air temperature was also used to guide our interpolation. Because the FAO station normals generally correspond to a 1950–80 climate period, the most recent version of the CRU maximum temperatures (CRU TS 4.01) (Harris et al. 2014) was used to adjust the FAO-based climatology to a 1981–2010 baseline. This adjustment was done by (i) calculating the mean CRU values between the periods of 1981–2010 and 1950–80, (ii) taking the difference between these climatologies, (iii) regridding this difference field to the CHT<sub>clim</sub> resolution, and (iv) adding the difference field to produce an adjusted CHT<sub>clim</sub>. Later, we show a

high level of correspondence between the CRU climatology and CHT<sub>clim</sub>. This correspondence arises because independent modeling efforts, driven with similar sets of in situ station normals, arrive at similar climatologies, and not because we force the CHT<sub>clim</sub> to match the CRU climatology.

### d. Berkeley station observations and gridded station data

To guide our time-varying CHIRTS<sub>max</sub> dataset, we use two sources of station data: the Berkeley Earth archive and the GTS archive obtained from the Climate Prediction Center (CPC). Our primary source of information was the Berkeley Earth breakpoint-adjusted station archive (Rohde et al. 2013b). The Berkeley station dataset draws from the Colonial Era Weather Archive, the Global Climate Observing System, the Global Historical Climate Network, the Global Summary of the Day, Hadley Centre CRU data, Monthly Climatic Data of the World, the U.S. Cooperative Summaries of the Day and Month, the World Monthly Surface Station Climatology, and the World Weather Records. During the Berkeley Earth averaging process, each station is compared with other nearby stations. Discontinuities and other heterogeneities are identified, and then a set of breakpoint-adjusted data is created. From 2002 onward, this dataset is augmented by GTS  $T_{\max}$  observations obtained from the CPC.

## 4. Methods

### a. The CHIRTS<sub>max</sub> modeling process

The CHIRTS<sub>max</sub> modeling process combines three components (Fig. 2): 1) CHT<sub>clim</sub>, a high-resolution (0.05°) monthly climatology of  $T_{\max}$  based on FAO station normals, long-term average MERRA-2 2-m temperature estimates, latitude, longitude, and elevation, 2) CHIRTS<sub>max</sub>, which are time-varying (1983–2016), high-resolution (0.05°) monthly satellite-based  $T_{\max}$  anomaly fields, and 3) CHTS<sub>max</sub>, interpolated monthly anomaly, based on Berkeley-GTS  $T_{\max}$  air temperature observations.

The process proposed is similar to but not identical with the process used to derive the CHIRPS precipitation product (Funk et al. 2015c). The major differences are (i) a simpler combination process (additive, not multiplicative), and (ii) the development of independent satellite and station-based anomaly fields. In the CHIRPS procedure, a satellite-only product (CHIRP) is developed and then adjusted using station data, producing CHIRPS. Experimentation revealed that this strategy, when applied to temperature estimation,

exhibited large potential errors associated with mismatches between the long-term mean values of in situ observations and satellite-based estimates. The large spatial heterogeneity of mean  $T_{\max}$  values induced frequent, large differences between the station observations and satellite estimates. This can be contrasted with the satellite-only CHIRP estimates, which typically exhibit very low bias (Dinku et al. 2018). Changing station networks combined with potentially large mean differences can create large, spurious variations in temperature over time. Working independently in station and satellite anomaly space removed this source of uncertainty, and also allowed each data source to be evaluated independently. The latter characteristic is very useful when evaluating extremes in applied early warning contexts.

By allowing  $\mathbf{C}$ ,  $\mathbf{V}$ , and  $\mathbf{S}$  to represent monthly grids of climatological averages, satellite  $T_{\max}$  anomalies, and interpolated station anomalies, the final CHIRTS<sub>max</sub> estimate ( $\mathbf{T}$ ) is derived from a simple linear combination:

$$\mathbf{T} = \bar{\mathbf{C}} + \alpha\mathbf{V} + \beta\mathbf{S}, \tag{1}$$

where  $\alpha$  and  $\beta$  are weights, summing to 1, based on the expected covariance of the anomalies fields and the true  $T_{\max}$  values at each location. We briefly describe the derivation of each component below.

*b. The  $T_{\max}$  climatology*

The climatology modeling methodology is a more sophisticated version of the process used to derive the precipitation climatology underlying the CHIRPS product (Funk et al. 2015a). The description that follows parallels closely to that of Funk et al. (2015a), and the following text is derived from there with minor modifications. There have been, however, two major advances in our climatological modeling procedure. First, rather than dividing the globe into 56 modeling regions, a modification of the algorithm now uses a continuous modeling regime to define locally fit moving window regressions (MWR). Second, the improved modeling process also uses a dynamically determined modeling radius. Areas with dense observation networks (like the United States) use a small modeling radius ( $d_{\max} \sim 50$  km), allowing the modeling process to capture fine details. Areas with sparse observation networks use a large radius ( $d_{\max} \sim 300$  km).

The modeling process first uses a series of locally fit MWR models to create an initial prediction of a 0.05°-resolution global temperature grid. The second step then calculates and interpolates the at-station residuals, with the residuals being station observations minus the regression estimates. These residuals are interpolated

using a modified inverse-distance weighting (IDW) interpolation scheme. These gridded residuals are added to the MWR model estimates to create the CHT<sub>clim</sub> climatology. This climatology is then adjusted to a 1981–2010 baseline using gridded CRU temperature estimates. The original FAO training data have a ~1951–80 baseline, so the change in the CRU archive between the periods of 1981–2010 and 1951–80 is added, resulting in our final monthly CHT<sub>clim</sub> climatologies.

1) LOCALIZED REGRESSION ESTIMATES

As first described in Funk et al. (2015a), and briefly summarized below, the CHC climatology modeling procedure is based on local regressions between a target variable, represented by a set of in situ climate normals, and spatially continuous background fields. Here, we begin by describing the bivariate case of this process, which can be extended to the multivariate instance in a straightforward manner.

For a given location, we first identify a set of in situ observations that fall within a certain distance ( $d_{\max}$ ) and calculate their distance weighted (localized) regression slope  $b(d)$  with a predictand. As in Funk et al. (2015a), this study uses a cubic function of the distance ( $d$ ) and a user-defined, regionally variable, maximum distance ( $d_{\max}$ ):

$$w(d) = 0, \quad d > d_{\max}, \quad \text{and} \\ w(d) = [(1 - d/d_{\max})^3]^3, \quad d \leq d_{\max}. \tag{2}$$

These weights are then used to estimate a localized regression slope where

$$\bar{x} = (n - 1)^{-1} \left[ \sum_{i=1}^n w(d_i) \right]^{-1} \sum_{i=1}^n w(d_i)x_i, \\ \bar{y} = (n - 1)^{-1} \left[ \sum_{i=1}^n w(d_i) \right]^{-1} \sum_{i=1}^n w(d_i)y_i, \quad \text{and} \\ b_{x,y} = (n - 1)^{-1} \left[ \sum_{i=1}^n w(d_i) \right]^{-1} \sum_{i=1}^n w(d_i)[x_i - \bar{x}][y_i - \bar{y}][x_i - \bar{x}]^{-2}. \tag{3}$$

The localized slope ( $b_{x,y}$ ) at some location ( $x, y$ ) corresponds with the cross product of the neighboring points weighted by their distance scaled by predictand sum of squares.

2) LOCALIZED MOVING WINDOW REGRESSIONS

The local bivariate slope estimation [Eq. (3)] is extended to a multivariate case using weighted least

squares regression model estimates. At each center point, station values within the radius ( $d_{\max}$ ) are collected and a regression model is fit. The  $d_{\max}$  values are defined individually for each location based on the density of the available data. Latitude, longitude, elevation, and long-term (1983–2016) average MERRA-2 2-m temperatures are used as inputs into the MWR, which is trained using long-term climate normals (green inputs Fig. 2).

### 3) INTERPOLATION OF MODEL RESIDUALS

Following the MWR modeling procedure, at-station anomalies (the arithmetic difference between the FAO station normals and the nearest  $0.05^\circ$  regression estimate) are calculated and interpolated using a modified IDW interpolation procedure. For each  $0.05^\circ$  grid cell, the cube of inverse distances is used to produce a weighted average of the surrounding station residuals,  $r$ . This value is then modified based on a local interpolation radius  $d_{\text{IDW}}$  and the distance to the closest neighboring station ( $d_{\min}$ ):

$$r^* = \left(1 - \frac{d_{\min}}{d_{\text{IDW}}}\right)r. \quad (4)$$

This simple thresholding procedure forces the interpolated residual field to relax toward zero based on the distance to the closest station. The  $d_{\min}$  values change based on the local-station density.

### 4) ADJUSTING THE $\text{CHT}_{\text{CLIM}}$ CLIMATOLOGY

The  $\sim 12\,300$  FAO station normals used to drive the MWR and IDW steps of the climatology have good spatial coverage in data-sparse areas like South America, Africa, and Central Asia. The dataset has a drawback in that the time period of the normals is poorly specified. Most of the data come from a 1950–80 time period, an era when data were shared more freely. This selection bias creates a tendency to modestly underestimate air temperatures in the 1981–2010 time period. We account for this bias by adjusting our penultimate climatology by the differences between the mean of the gridded 1981–2010 CRU TS 4.01 dataset and the 1950–80 spatially varying gridded CRU fields. This temperature change field is regridded to a  $0.05^\circ$  latitude–longitude grid using parametric cubic convolution with a  $-0.3$  interpolation parameter.

#### *c. Deriving daily and monthly $\text{CHIRT}_{\max}$ values*

One underlying basis for this work is the strong relationship between geostationary cloud-screened satellite brightness temperatures and observed monthly  $T_{\max}$

observations. Brightness temperatures (such as those observed in the  $11\text{-}\mu\text{m}$  channel from geostationary satellites) are a function of the radiation emitted from the Earth's surface, which is a direct function of temperature and emissivity, as specified by the Stefan–Boltzmann law ( $\text{rad} = \varepsilon\sigma T_e^4$ ), where  $\varepsilon$  is emissivity,  $\sigma$  the Stefan–Boltzmann constant, and  $T_e$  the emission temperature. The brightness temperature is estimated by taking the fourth root of the satellite-observed radiation ( $T_e = \text{rad}^{0.25}$ ).

While emissivity will modulate the total energy received by a passive geostationary satellite, temporal changes in the brightness temperature are mostly driven by changes in surface temperature, not emissivity. These changes will arise from variations in the surface temperature (the diurnal cycle and weather) as well as the elevation of the emitting surface: brightness temperatures arising from energy emitted from cloud tops will be lower because temperatures rapidly diminish with height.

One core element of the  $\text{CHIRT}_{\max}$  processing stream is a series of filtering procedures that attempt to minimize the influence of cloud contamination. These procedures are similar to commonly used maximum compositing algorithms used in estimating vegetation indices like the normalized difference vegetation index (NDVI). Atmospheric water vapor contaminates the NDVI signal, consistently reducing NDVI values (Van Leeuwen et al. 1999). The systematic nature of this contamination, with water vapor always reducing NDVI values, provides an avenue for correction by filtering out low NDVI values. Here, we face a similar problem and apply a similar solution.

With observations from a single channel, we cannot confidently identify all cloud contributions to the TIR signal. Even under ideal conditions, issues associated with mixed pixels or low-lying fog, for example, can produce suppressed temperature conditions, which may be impossible to discriminate from “true” cold weather. The objective of the  $\text{CHIRT}_{\max}$  screening process is not to totally remove such influences, but to reduce their effects to a point where the screened TIR is a useful predictor of true 2-m air temperatures. The general approach has three steps: (i) the estimation of daily  $\text{CHIRT}_{\max}$  values, (ii) spatial maxima filtering of daily  $\text{CHIRT}_{\max}$  values, and (iii) the translation of monthly Z scores into 2-m  $T_{\max}$  estimates.

#### 1) ESTIMATING DIURNAL $T_{\max}$ VALUES

The first step of the  $\text{CHIRT}_{\max}$  screening process builds a historical probability density function (PDF) based on histograms of the observed brightness temperatures at each pixel for a specific month and 3-hourly



period. Figure 3 shows an example histogram based on 1983–2016 GridSat B1 data—for 1200 UTC, for April—for a pixel located in western Kenya. The PDF exhibits substantial skew, with a long tail on the low end, presumably associated with cloud contamination. We know that this area exhibits warm near-surface conditions and limited air temperature fluctuations because of its tropical location. Temperatures drop rapidly with height, and the thermal energy emitted from clouds is much lower than surface emissions, leading to cooler brightness temperatures.

Assuming that cloud-free conditions are the most common state, we use the mode of the distribution as a center point of our distribution (1 in Fig. 3). Values cooler than this central value are considered potential clouds. As described below, the distribution of values above the mode is used to define a cutoff value for cloud screening. While this screening may create a warm bias (preferencing the selection of warmer surface TIR observations), the final estimates will be based on the standardized monthly estimates, effectively removing this influence. The temperature range (in kelvins) between the central value and the 99th percentile value (between 1 and 2 in Fig. 3) provides the expected range of cloud-free observations. Subtracting this range from the central value gives a lower threshold for plausible cloud-free observations (3 in Fig. 3). We are using the distance, in degrees Celsius, between the median and the 99th percentile to estimate the temporal variance of the land surface. Using other thresholds (90th, 95th, 98th) produced similar results. Any TIR values less than this lower threshold are assumed to be contaminated by clouds and are treated as missing and removed. This screening, of course, could be imperfect for a number of reasons; our central value may be contaminated by clouds, small fractions of cloud contamination may have an undetectable cooling effect, or emissions from near-surface clouds and water vapor may be undetectable. On the other hand, if we can screen most of the cloud values from our data, we receive a signal that has an immediate relationship to the surface temperature.

Once subdaily data have been screened, we then carry out a second screening process that estimates a daily  $\text{CHIRT}_{\text{max}}$  value. Since we are unsure whether or not any individual 3-hourly observation has been partially contaminated by clouds, we identify each day's warmest 3-hourly anomaly and use that value to estimate the daily TIR-based  $T_{\text{max}}$ . To do this, we start by screening out all observations found to be below the lower thresholds described above. We then translate the remaining 3-hourly values into anomalies, by subtracting each location's long-term mean based on the cloud-free data for that 3-hourly period and month. Then, we

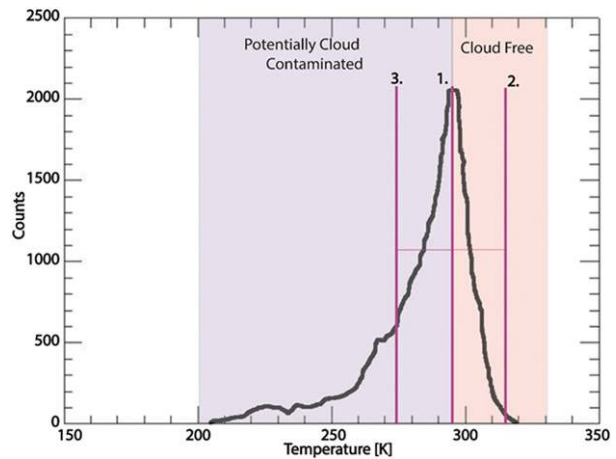


FIG. 3. Histogram of 1983–2016 GridSat B1 data, for 1200 UTC, for April, for a pixel located in western Kenya. Values colder than the cutoff at 3 were screened as cloudy. Values warmer than the cutoff at 3 were treated as valid observations.

convert the subdaily anomalies into estimates of daily maximum temperatures by adding the maximum subdaily anomaly to the maximum 3-hourly temperature for that pixel's climatological diurnal cycle. This means that we might take a warm nighttime temperature anomaly and add it to that location's climatological maximum daytime temperature. We do so because we are focused on producing less cloud-influenced monthly  $\text{CHIRT}_{\text{max}}$  data, while using as many cloud-free observations as possible. With only eight 3-hourly observations to work with, exploratory analysis revealed very poor daytime sampling in tropical cloudy regions. This poor daytime sampling is likely exacerbated by the fact that in warm tropical climates, clouds often form in conjunction with diurnal increases in land surface temperatures. Note that, at this stage, if no noncloudy subdaily temperature values are available, we flag the daily  $\text{CHIRT}_{\text{max}}$  value as missing.

## 2) SPATIAL MAXIMA FILTERING OF DAILY $\text{CHIRT}_{\text{MAX}}$ VALUES

Next, we filter the daily TIR-based  $T_{\text{max}}$  estimates spatially. For each day, we use running  $5 \times 5$  pixel filters to select the warmest daily  $T_{\text{max}}$  value across a  $25 \text{ km} \times 25 \text{ km}$  window. The selected value is then assigned to the central pixel of the  $5 \times 5$  window, and the process repeated for all of the  $0.05^\circ$   $\text{CHIRT}_{\text{max}}$  grid cells. This selection process focuses on the warmest observations, preferencing the selection of less-cloudy or cloud-free pixels and diminishing the potential influence of cloud contamination. This  $5 \times 5$  window was selected based on validation analyses comparing filtered results using  $3 \times 3$ ,  $5 \times 5$ ,  $7 \times 7$ ,  $9 \times 9$ ,  $11 \times 11$ ,  $13 \times 13$ , and  $15 \times 15$  windows.

These results (not shown) compared monthly maximum-filtered  $\text{CHIRT}_{\text{max}}$  values to CRU observations. The filtering process increased the correlation with the CRU data and removed clearly visible reductions due to cloud contamination. No clear benefit was obtained using windows larger than  $5 \times 5$ , hence this spatial filter was selected as the best trade-off between resolution and accuracy.

### 3) CREATING MONTHLY $\text{CHIRT}_{\text{max}}$ ESTIMATES

While the GridSat archive brightness temperatures (Knapp et al. 2011) were calibrated using satellite-specific information and geometry, evaluations of brightness temperature data indicated inhomogeneities in some places and time periods. For example, *Meteosat-2* brightness temperatures in the Sahel in the early 1980s appeared unrealistically warmer than later temperatures from *Meteosat-7*. These systematic biases were estimated by compositing daily TIR  $T_{\text{max}}$  and CRU TS 4.01 temperatures for each month, location, and satellite combination. Each month's daily TIR  $T_{\text{max}}$  values were then adjusted by the difference between the CRU and satellite TIR  $T_{\text{max}}$  averages for each location.

Then, for each location and month, if at least 10-daily noncloudy  $\text{CHIRT}_{\text{max}}$  estimates were available, a monthly  $\text{CHIRT}_{\text{max}}$  was calculated from the average of these values. The 1983–2016 time period was then used to calculate monthly  $\text{CHIRT}_{\text{max}}$  climatologies (mean fields) and standard deviations. These values were used to translate the monthly varying  $\text{CHIRT}_{\text{max}}$  values into standardized anomalies ( $Z$  scores). These standardized anomalies were clipped to  $-3.5$  and  $4.0$ . This clipping was used to screen for implausible TIR values, since there could be unrealistically high or low TIR values in the B1 archive. A higher threshold was used for warm anomalies because the Earth is warming.

While the diurnal and spatial filtering process used in the  $\text{CHIRT}_{\text{max}}$  development algorithm tended to preference the selection of less-cloudy or cloud-free pixels (creating a tendency to overestimate the mean of “true” surface temperature), this bias was not carried through to our final product since only monthly  $\text{CHIRT}_{\text{max}}$  standardized anomalies are used.

Anomaly fields were then estimated and translated into estimates of 2-m maximum air temperatures. For each month, this translation was achieved by multiplying  $\text{CHIRT}_{\text{max}}$   $Z$  scores by the local 1983–2016 standard deviation of the gridded CRU  $T_{\text{max}}$  product. These monthly standardized deviations were calculated at the native CRU resolution ( $0.5^\circ$ ) and then resampled to a  $0.05^\circ$  resolution. Following the nomenclature from Eq. (1),

$$I' = Z_{\text{IRT}} \sigma_{\text{CRU}}, \quad (5)$$

where  $Z_{\text{IRT}}$  are the standardized anomalies based on the  $\text{CHIRT}_{\text{max}}$   $Z$  scores, and  $\sigma_{\text{CRU}}$  are the CRU TS 4.01  $T_{\text{max}}$  standard deviations. The expected standard deviation is 1.

### d. Quality controlling and interpolating Berkeley-GTS station data

For each month, candidate Berkeley-GTS station data locations were derived by identifying stations with at least seven observations over the 14 years between 2003 and 2016. The rationale for this selection process was that we are most interested in temperature variations over recent years in locations where we can plausibly calculate median temperature values. Additional tests were developed through statistical and visual analysis of the Berkeley-GTS station data. The final tests were applied to identify good station observations: (i) the standardized difference between the recent (2003–16) median and mean must be less than 4, (ii) the standardized difference between the 2003–16 and 1983–96 mean must be less than 7, (iii) the standardized difference between the maximum and median air temperature must be less than 5, (iv) the standardized difference between the minimum and median air temperature must be greater than  $-5$ , and (v) the standardized difference between the station median and local  $\text{CHT}_{\text{clim}}$  value must be less than  $\pm 6$ . Stations failing these criteria were excluded, which typically reduced the total number of stations by approximately 10% to 12%. Thresholds for each of these five tests were derived via extensive visual examination of individual station time series, combined with sensitivity analyses that looked at the number and locations of stations lost when each threshold was altered. Values were selected that removed obviously bad data but left as much station coverage as possible.

For the screened stations, monthly  $T_{\text{max}}$  values were translated into anomalies by subtracting the 2003–16 median from each time series. Additional filtering was used to identify and remove potential erroneous extremes. Anomalies with absolute values of greater than  $\pm 8^\circ\text{C}$  were removed, as were anomalies with standardized values of greater than  $+4$  or less than  $-3.5$ . Asymmetric thresholds were used based on the possible influences of climate change. Individual monthly  $\text{CHIRT}_{\text{max}}$  anomalies, expressed as differences from the  $\text{CHIRT}_{\text{max}}$  2003–16 median, were compared with the station anomalies. Station observations with differences exceeding  $\pm 10^\circ\text{C}$  were excluded. Each threshold described above was identified by examining maps, time

series, and histograms of the underlying data and screening criteria.

A spherically correct spatial interpolation toolkit (Spherkit; Raskin et al. 1997) was used to interpolate the monthly station anomalies using IDW with an angular correction to reduce the influence of clustered station observations. Spherkit uses direction cosines to calculate correct geospatial Earth distances. This algorithm is similar to the commonly used Spheremap procedure. The interpolators' neighborhood function sought an average of 10 neighbors, although the number of neighbors was allowed to range between 3 and 20.

*e. Combining the  $CHIRT_{max}$  and  $CHTS_{max}$  anomalies*

Combinations of the two datasets were produced based on the expected variance explained by each estimate:

$$\alpha + \beta = 1 \alpha \propto R^2(IRT_{max}), \quad \beta \propto R^2(IDW StnT_{max}),$$

$$R^2(IRT_{max}) = 0.56, \quad \text{and}$$

$$R^2(IDW StnT_{max}) = \exp\left(-\frac{d_{neighbor}}{\text{Range km}}\right). \quad (6)$$

The relative weighting of the  $CHIRT_{max}$  ( $\alpha$ ) and  $CHTS_{max}$  station fields ( $\beta$ ) are proportional to their expected variance explained values ( $R^2$ ). The expected variance, explained by  $CHIRT_{max}$ , is derived from comparisons with independent station data, a single expected  $R^2$  value is used (0.56) based on validation analyses reported below in section 5a and in Figs. 4e and 4f. Comparisons with Berkeley-GTS station data and gridded CRU time series consistently indicate that the  $CHIRT_{max}$  typically explains a bit more than half the variance of the observations. The  $CHTS_{max}$  variance explained estimates are based on the distance to the nearest neighboring station,  $d_{neighbor}$  in Eq. (6). When this distance is 0, the expected variance explained is 1. At distances greater than 0, an exponential decay function is used to reduce the expected variance explained. The structure of this decay function is based on empirical spatial covariograms, as described in section 5a and in Figs. 4e and 4f.

When the nearest station is far away from a given grid point,  $\alpha$  will reach 1 and  $\beta$  will reach 0. At a station location,  $\alpha$  and  $\beta$  will have values of 0.36 and 0.64, that is,  $0.56/(0.56 + 1)$  and  $1/(0.56 + 1)$ . Hence, the  $CHIRTS_{max}$  process is not an exact interpolator. Even at a station location, a substantial  $CHIRT_{max}$  contribution will be included. In areas with missing  $CHIRT_{max}$  data,  $\alpha$  and

$\beta$  will be set to 1 and 0, respectively. The final  $CHIRTS_{max}$  estimates are then produced:

$$CHIRTS_{max} = CHT_{clim} + \alpha CHIRTS_{max} Anom + \beta IDW_{Stn} Anom. \quad (7)$$

## 5. Results

### a. Validation analyses based on station data

We begin with an analysis of the correlation of the  $CHIRT_{max}$ , MERRA-2, and  $CHIRTS_{max}$  databased on the quality-controlled Berkeley-GTS station archive. Spatial similarities of CRU,  $CHIRTS_{max}$  and MERRA-2 are compared in our Figs. S1–S3 in the online supplemental material. Figure 4a shows the location where at least seven Berkeley-GTS observations existed between 2003 and 2016. Since we are interested in producing an effective monitoring product, it is important that we find a large number of available stations: 19606 stations, to be precise. For comparison, we also show the corresponding locations based on the CRU dataset (blue dots in Fig. 4a). The CRU has about one-tenth the number of observing stations (2068), and these stations are poorly distributed in most continents. While the heavily curated CRU archive is well suited for estimating temperature trends in areas with good station density, the Berkeley-GTS archive may be better suited for monitoring temperature extremes outside of Europe and North America.

July results for 2016 (Fig. 4b) indicate similar results. Figure 4b shows the location of every Berkeley-GTS and CRU station observation in July of 2016. The temporal coverage of Berkeley-GTS archive is far more complete. Figures 4c and 4d displays time series of the total number of Berkeley-GTS and CRU observations globally and for Africa. While the number of Berkeley-GTS observations differs substantially over time, the archive consistently contains many more observations than the CRU TS 4.01.

Table 1 compares MERRA-2,  $CHIRTS$ , and CRU means and standard deviations. Table 2 displays monthly correlation and mean absolute error (MAE) statistics for the  $CHIRT_{max}$ , cross-validated  $CHTS_{max}$  station estimates,  $CHIRTS_{max}$ , and MERRA-2 datasets. These statistics were based on stations at least 150 km from the nearest neighbor. The  $CHTS_{max}$  and  $CHIRTS_{max}$  validation results are based on cross-validated IDW estimates for which the corresponding station observations have been withheld. MAE values are based on the difference between the observed and estimated monthly  $T_{max}$  anomalies. Figure 5 shows scatterplots for the January and July station-only, satellite-only, and blended satellite–station temperature estimates.

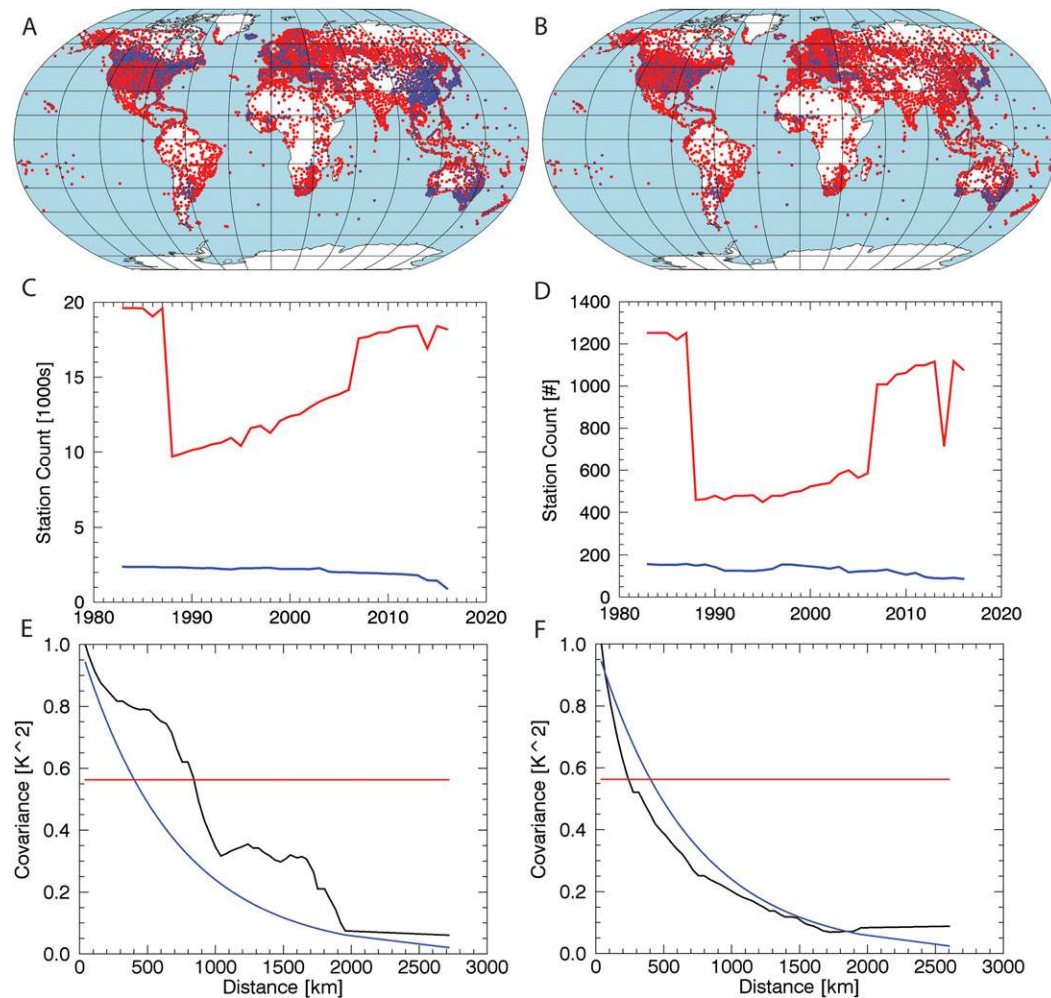


FIG. 4. (a) Red dots indicate where the Berkeley-GTS station database had at least 7 observations in July between 2003 and 2016; blue dots indicate the same for CRU stations. Blue dots are drawn on top of the Berkeley-GTS station locations. (b) Locations of Berkeley-GTS (red) and CRU (blue) observations for July 2016. Blue dots are drawn on top of the Berkeley-GTS station locations. (c),(d) Counts of Berkeley-GTS and CRU stations for (c) Africa and (d) the globe. (e) Observed (black line) and modeled covariogram (blue) for January. Red line shows assumed covariance for  $\text{CHIRT}_{\max}$ ,  $\text{Cov}(\text{CHIRT}_{\max}) = 0.75^2$ . (f) As in (e), but for July.

Overall, the performance of the  $\text{CHIRT}_{\max}$  anomalies is quite good for a space-based estimator, with median correlation and MAE values of  $0.75^\circ$  and  $0.85^\circ\text{C}$  when compared to the Berkeley-GTS station observations.

The performance of MERRA-2 is slightly better (median correlations of 0.8 and MAE values of  $\sim 0.75^\circ\text{C}$ ). However, note that since the MERRA-2 assimilates station observations, the MERRA-2 statistics may be

TABLE 1. Comparison of 1983–2016 MERRA-2,  $\text{CHIRTS}_{\max}$ , and CRU TS 4.01 terrestrial means and standard deviations, and pattern correlations for the terrestrial mean and standard deviation fields.

	Means ( $^\circ\text{C}$ )			Standard deviations ( $^\circ\text{C}$ )			Correlations: means				Correlations: standard deviations	
	MERRA-2	$\text{CHIRTS}_{\max}$	CRU	MERRA-2	$\text{CHIRTS}_{\max}$	CRU	MERRA-2– CRU	$\text{CHIRTS}_{\max}$ – CRU	MERRA-2– CRU	$\text{CHIRTS}_{\max}$ – CRU	MERRA-2– CRU	$\text{CHIRTS}_{\max}$ – CRU
Annual	16.4	17.8	17.1	0.90	0.5	0.6	0.99	0.99	0.45	0.91		
Jan	6.2	7.4	7.2	2.3	1.9	2.0	0.99	0.99	0.85	0.93		
Jul	25.7	27.0	26.0	1.5	1.1	1.1	0.97	0.93	0.72	0.95		

TABLE 2. Comparison of correlations and mean absolute errors of  $\text{CHIRT}_{\max}$ ,  $\text{CHTS}_{\max}$ ,  $\text{CHIRTS}_{\max}$ , and MERRA-2 with respect to Berkeley-GTS stations observations when the nearest neighboring station value is at least 150 km away. The 150 km threshold was selected to limit the influence of heavily instrumented regions. The Cov Model row shows the ranges used in Eq. (6) for each month. This range controls the distance decay from each station.

		Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Median
Correlation	$\text{CHIRT}_{\max}$	0.76	0.76	0.78	0.76	0.73	0.69	0.66	0.68	0.71	0.73	0.76	0.75	0.75
	$\text{CHTS}_{\max}$	0.9	0.88	0.87	0.87	0.84	0.81	0.81	0.82	0.85	0.86	0.87	0.84	0.86
	$\text{CHIRTS}_{\max}$	0.89	0.89	0.88	0.87	0.84	0.82	0.81	0.81	0.84	0.85	0.87	0.86	0.86
	MERRA-2	0.85	0.85	0.84	0.82	0.78	0.76	0.73	0.75	0.78	0.79	0.82	0.8	0.80
MAE	$\text{CHIRT}_{\max}$	1.1	1	0.9	0.9	0.8	0.8	0.8	0.8	0.8	0.8	0.9	1	0.85
	$\text{CHTS}_{\max}$	0.7	0.8	0.7	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.7	0.8	0.6
	$\text{CHIRTS}_{\max}$	0.75	0.8	0.7	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.7	0.8	0.6
	MERRA-2	0.9	0.9	0.8	0.8	0.7	0.7	0.7	0.7	0.7	0.7	0.8	0.9	0.75
Cov Model	Range	1100	900	700	700	700	700	700	700	700	700	800	900	

overestimating the true accuracies. While the performance of MERRA-2 is impressive, extreme trends in MERRA-2 temperatures were also found in parts of Africa and Central/South America (Africa example shown in Fig. S4). These physically implausible tendencies are thought to be caused by shifts in surface energy balances associated with changes in forcing datasets (Reichle et al. 2017; Draper et al. 2018).

The cross-validated  $\text{CHTS}_{\max}$  station anomalies and  $\text{CHIRTS}_{\max}$  estimates have similar accuracy statistics: median correlations and MAE values of  $0.86^\circ$  and  $0.6^\circ\text{C}$  (Table 2). Statistically, the inclusion of TIR data does little to improve our observed levels of accuracy, largely

stations are spatially clustered, potentially skewing these results. Because stations tend to be clumped geographically, cross-validated estimates of station-based accuracies tend to be optimistic, in the sense that they tend to quantify accuracies in areas that have stations. In an applied setting, we may often be interested in extremes at remote distances from a station. Hence, the blended product ( $\text{CHIRTS}_{\max}$ ) is expected to perform relatively well in data-sparse regions.

Figures 4e and 4f show spatial covariograms for January and July, respectively. These plots are based on the Berkeley-GTS station archives and interstation distances and covariances. At a distance of 0, the observed

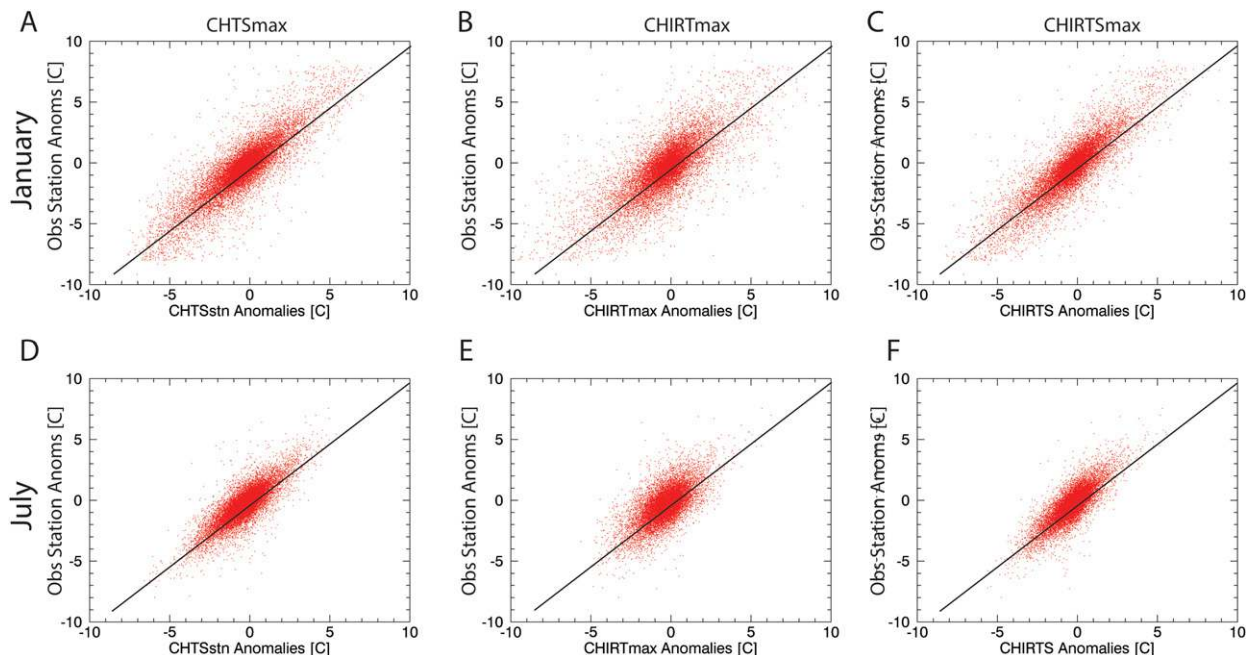


FIG. 5. Validation results based on comparison with Berkeley-GTS stations. (a) Cross-validated interpolated  $T_{\max}$  estimates and observed station anomalies for January. (b) January  $\text{CHIRT}_{\max}$  and station  $T_{\max}$  anomalies. (c) Same for  $\text{CHIRTS}_{\max}$ . (d)–(f) As in (a)–(c), but for July.

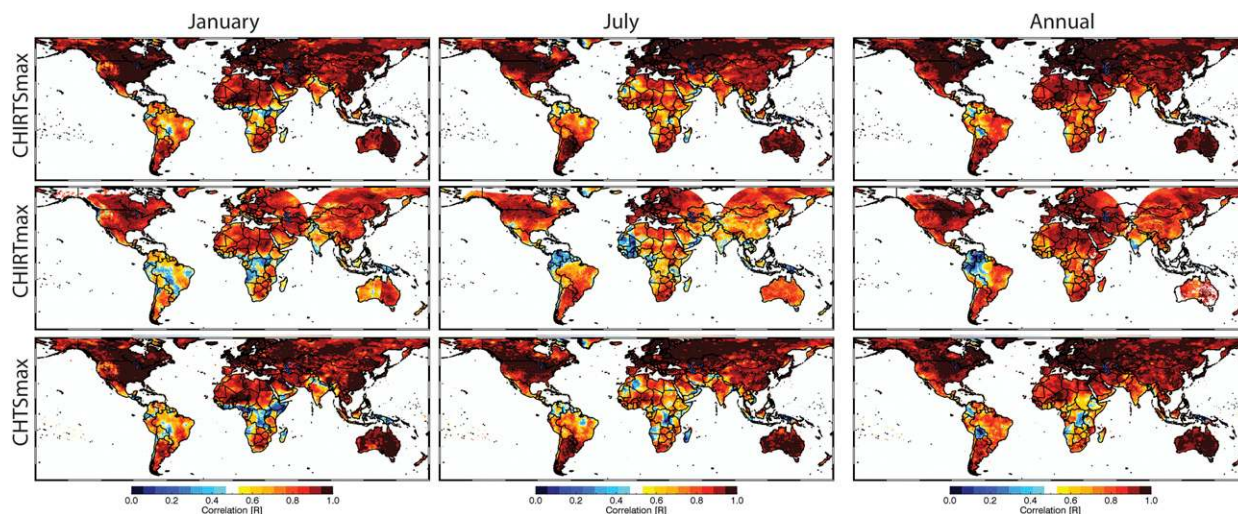


FIG. 6. 1983–2016 Correlations between (top) CRU TS 4.01 and  $\text{CHIRTS}_{\text{max}}$ , (middle) CRU TS 4.01 and  $\text{CHIRT}_{\text{max}}$ , and (bottom) CRU TS 4.01 and  $\text{CHTS}_{\text{max}}$ .  $\text{CHIRT}_{\text{max}}$  correlations with less than 7 observations set to missing.

variance explained approaches 1, and this value decays exponentially with increasing distance. An examination of all months (summarized in Table 2) found exponential decay ranges between 700 and 1100. A fixed value of 700 km was used to estimate variances explained in the calculation of the blending weights  $a$  and  $b$  [Eq. (6)]. A fixed value of  $0.75^2$  was used as the expected variance explained by the  $\text{CHIRT}_{\text{max}}$  anomaly field.

#### b. $\text{CHIRTS}_{\text{max}}$ /CRU TS 4.01 correlation maps

Figure 6 shows correlations between the  $\text{CHIRTS}_{\text{max}}$ ,  $\text{CHIRT}_{\text{max}}$  and  $\text{CHTS}_{\text{max}}$  data versus CRU TS 4.01 fields for January, July, and annual averages. Overall, we find high levels of agreement. This is important, because it means that we should be able to produce a  $T_{\text{max}}$  monitoring product that is similar to the current “gold standard” CRU dataset. There are areas, especially in tropical Africa and South America, where correlations are low. There are also areas (the western Sahel, Western Australia) where the  $\text{CHIRT}_{\text{max}}$ /CRU correlations are particularly low. These differences likely points to continued systematic errors associated with the  $\text{CHIRT}_{\text{max}}$  calibration. These errors are introduced when transitioning between different geostationary satellites.

Note that we do use the CRU to adjust  $\text{CHIRT}_{\text{max}}$  to intercalibrate between geostationary satellites. Most regions were covered by six or seven different geostationary satellites between 1983 and 2016, so each  $\text{CHIRT}_{\text{max}}$  time series was adjusted six to seven times using the CRU data. The correlation maps shown in Fig. 6 may be modestly optimistic given this intercalibration. Comparisons with cross-validated  $\text{CHTS}_{\text{max}}$

estimates, however, produce similar accuracy results. Both the  $\text{CHIRT}_{\text{max}}$  and  $\text{CHTS}_{\text{max}}$  effectively reproduce the variability expressed in the CRU archive. This implies that future extensions of the  $\text{CHIRTS}_{\text{max}}$  monitoring framework should be able to provide high-quality  $T_{\text{max}}$  estimates with relatively low latencies on the order of 1 or 2 months.

In general, tropical areas tend to have lower correlations. CRU coverage in these regions is often limited (Fig. 4), and high levels of cloud and water vapor contamination may erode the accuracy of the satellite retrievals used to derive the  $\text{CHIRT}_{\text{max}}$ .

#### c. CRU TS 4.01 and $\text{CHIRTS}_{\text{max}}$ temperature changes

We turn next to a brief examination of temporal changes in annual  $\text{CHIRTS}_{\text{max}}$  and CRU TS 4.01  $T_{\text{max}}$ . Our objective is to assess whether the warming patterns identified within the CRU TS 4.01 are similar to those in  $\text{CHIRTS}_{\text{max}}$ . Note that these datasets rely on similar station inputs, and therefore, where stations are dense, they will likely provide similar results. Conversely, both datasets have areas of sparse station coverage, especially in Africa and South America. Because linear trend estimates may be strongly influenced by outliers and end points, we use a simple nonparametric approach based on the difference between the beginning and end of the dataset.

Figure 7 shows maps contrasting mean temperatures during the last 16 years of data (2001–16) and the first 16 years (1983–98). First, we note that most of the globe has warmed substantially in line with climate change model simulations, such as those from the Intergovernmental Panel on Climate Change (Taylor et al. 2012). There is a

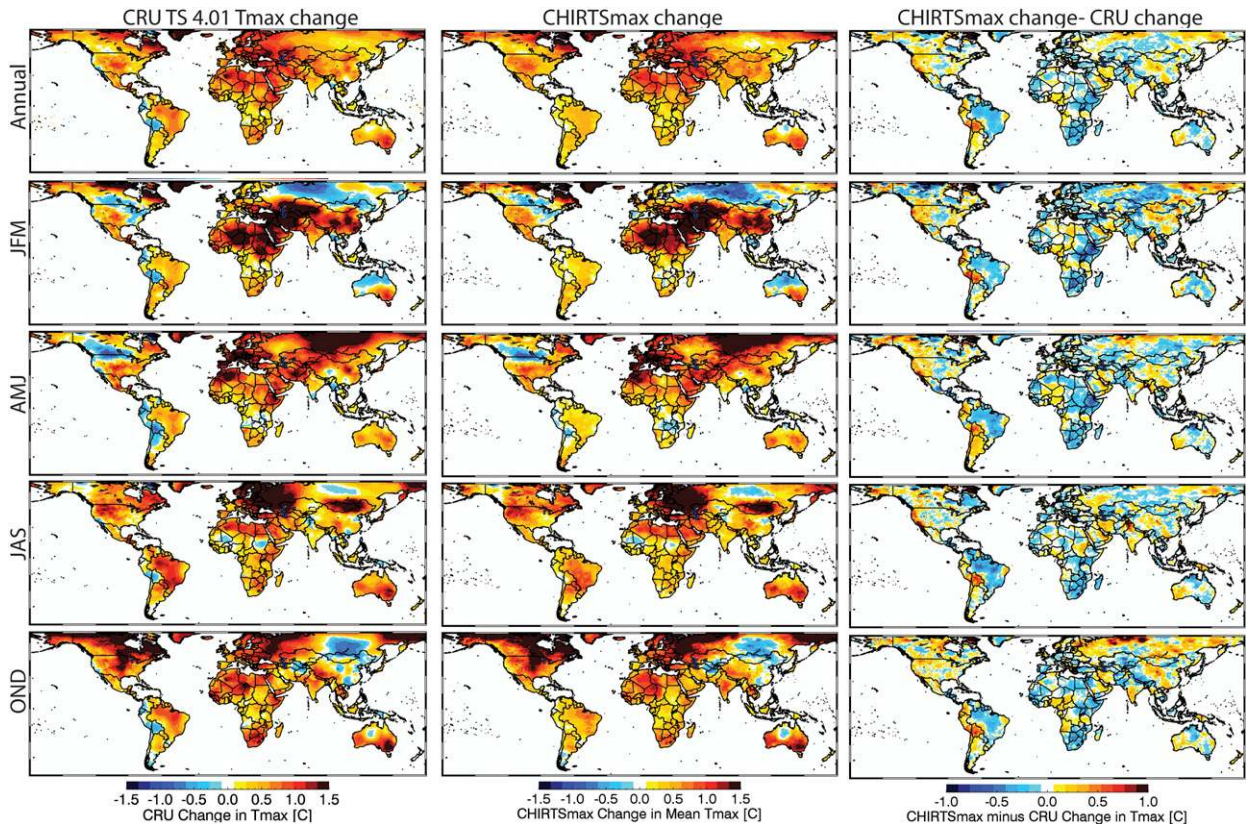


FIG. 7. (left),(middle) Changes in 2001–16 vs 1983–98 mean temperatures for the CRU TS 4.01 and CHIRTS<sub>max</sub> datasets for annual, January–March, April–June, June–September, and October–December. (right) Changes in CHIRTS<sub>max</sub> minus the changes in CRU.

strong level of agreement between the CRU and CHIRTS<sub>max</sub>, with pattern correlations for the annual, January–March, April–June, July–September, and October–December temperature changes of 0.94, 0.92, 0.94, and 0.94, respectively. This convergence boosts our confidence in both the CRU and CHIRTS<sub>max</sub>. Figure S5 shows similar variations in the satellite-only CHIRTS<sub>max</sub> and station-only CHIRTS<sub>max</sub>.

At global and continental scales, the MERRA-2 and CHIRTS<sub>max</sub> based annual mean  $T_{max}$  time series are highly correlated with the CRU (Table 3 and Fig. 8). CHIRTS<sub>max</sub> annual correlation values vary from 0.90 for South America to 0.97–0.99 for the other continents and the globe. Note that while there is some redundancy between the CHIRTS<sub>max</sub> and CRU, CHIRTS<sub>max</sub> has ~10 times more station observations. Once again, the performance of CHIRTS<sub>max</sub> and CRU is highly convergent.

While the MERRA-2 data are also highly correlated with the CRU (correlation values from 0.89 to 0.97), it appears MERRA-2 data substantially underestimate recent warming over many regions.

The seasonal progression of temperature changes across Asia is quite similar in the CHIRTS<sub>max</sub> and CRU datasets (Fig. 7). A north–south, cooling–warming dipole in January–March transitions to strong warming in April–June, with large temperature increases persisting into boreal summer near Mongolia. Large boreal summer temperature increases are also found over western Asia and eastern Europe.

*d. Maps and time series of estimated variance explained by CRU and CHIRTS<sub>max</sub>*

This section estimates the extra information included within the CHIRTS<sub>max</sub> versus the CRU station database.

TABLE 3. Correlations between time series of annual  $T_{max}$  CRU TS 4.01, MERRA-2, and CHIRTS<sub>max</sub> for selected regions.

	Globe	North America	South America	Africa	Eurasia	Australia
MERRA-2	0.97	0.93	0.89	0.88	0.95	0.93
CHIRTS <sub>max</sub>	0.99	0.98	0.90	0.97	0.99	0.99

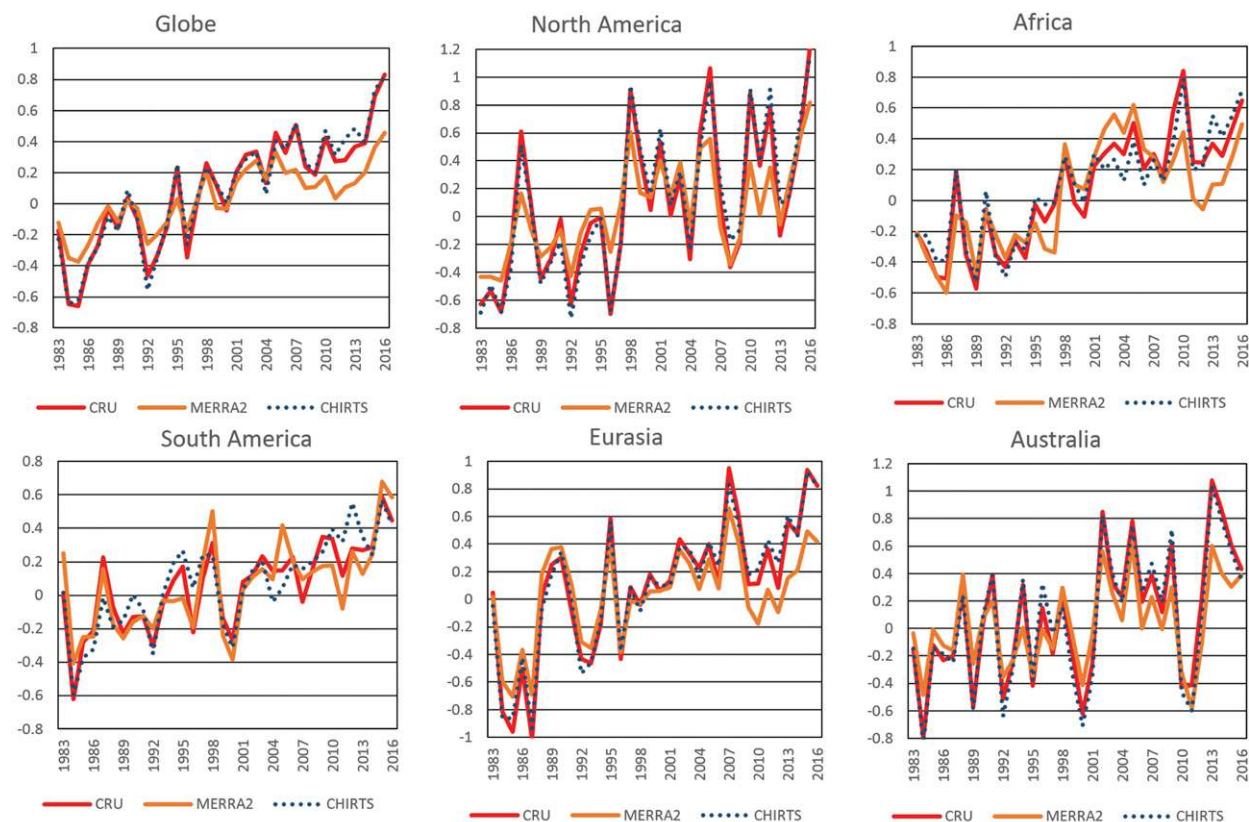


FIG. 8. Time series of CRU TS 4.01, MERRA-2, and CHIRTS<sub>max</sub> 1983–2016 annual  $T_{\max}$  anomalies averaged over the globe and across continents. Anomalies based on a 1981–2010 baseline.

This extra information arises from two sources: (i) the inclusion of a much denser set of station observations (Fig. 4), and (ii) the addition of CHIRTS<sub>max</sub>, which is found to have a good ability to explain the variance of the observed station data (Fig. 5,  $R^2 \sim 0.56$ ). We present only January results here, since all the months were similar. For the CRU, we use our empirical covariogram function from Eq. (6) [ $R^2(\text{IDW Stn } T_{\max}) = \exp[-(d_{\text{neighbor}}/\text{Range km})]$ ] to translate the distance to the closest station ( $d_{\text{neighbor}}$ ) into an estimate of the variance explained by the CRU. This relationship is based on the observed covariance structure (Fig. 4e). Figure 1a shows an example CRU variance explained map. For the CHIRTS<sub>max</sub>, we perform a similar calculation for the CHIRTS<sub>max</sub> component, assume the CHIRTS<sub>max</sub>  $R^2 = 0.56$ , and then use our  $\alpha$  and  $\beta$  weights [Eq. (6)] to produce a local estimate of the combined variance explained  $R^2 = \alpha(0.56)^{-0.5} + \beta(R_{\text{stn}}^2)^{-0.5}$ . Figure 1b shows an example of such a map for January of 2016. While these estimates assume that only the closest station is used in spatial interpolation, and hence are not exact estimates, they allow us to plausibly assess the implications of adding both a much greater number of stations and satellite-based IRT<sub>max</sub> fields.

Figure 9 shows the differences (CHIRTS<sub>max</sub> – CRU) for January for the 1983–2016 time period (Fig. 9a) and the 2011–16 time period (Fig. 9b). Averaging across these maps suggests that the CHIRTS<sub>max</sub> explains about 23% more of the variance. The 1983–2016 average CHIRTS<sub>max</sub> variance explained statistic is very good  $R^2 = 91\%$ . The CRU variance explained statistic is 68%. The corresponding values for 2011 are 92% versus 63%, with the CHIRTS<sub>max</sub> explaining +29% more variance.

Figure 10 shows average variance explained time series for the globe, South America, Africa, India and areas north of 50°N. Only January statistics are presented, but results for other months were similar. The CHIRTS<sub>max</sub> variance explained estimates are substantially higher, especially in low data density locations. For the globe, South America, Africa, India, and areas north of 50°N, estimates of variance explained for the 1983–2016 period for the CRU were 68%, 49%, 54%, 51%, and 70%, respectively. The corresponding CHIRTS<sub>max</sub> values were 91%, 89%, 84%, 92%, and 71%, respectively, with improvements ranging from 34% to 81%. Over the 2011–16 period, the CRU variance explained means for these same areas were 64%, 46%, 47%, 41%, and 66%, respectively, while the CHIRTS<sub>max</sub> values



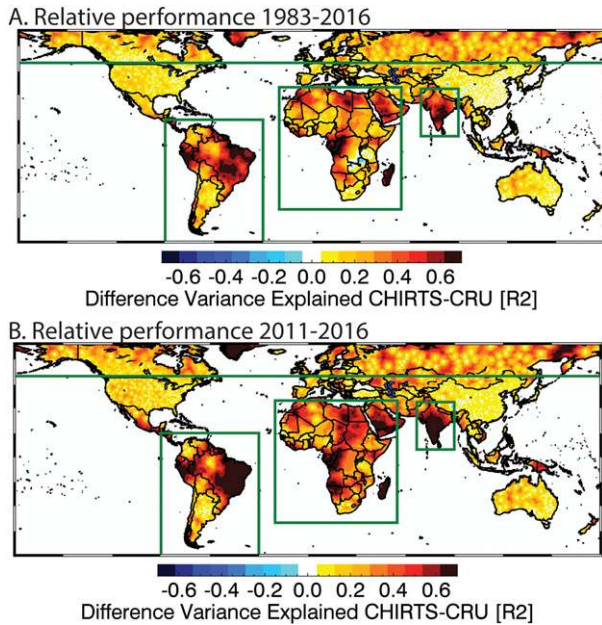


FIG. 9. Maps of differences in average January  $T_{max}$  variance explained ( $R^2$ ) for the CHIRTS<sub>max</sub> and CRU datasets. (a) Differences (CHIRTS – CRU) for the entire period of record. (b) Differences for 2011–16.

were 92%, 91%, 86%, 93%, and 94%, respectively. The CHIRTS<sub>max</sub> improvements over this later period ranged from 45% to 124%. Note that these improvements cover both tropical/subtropical and high-latitude regimes. Furthermore, the performance of the CHIRTS<sub>max</sub> appears much more consistent in the latter part of the record.

Our supplemental material provides a brief analysis focused on East Africa (Figs. S6 and S7) highlighting the potential value of the high-resolution of CHIRTS<sub>max</sub>. Time series of CRU and CHIRTS<sub>max</sub> air temperatures again indicate a high level of congruence but with a more consistent level of warming in the CHIRTS<sub>max</sub> since 2011. We also provide examples focused on monitoring extreme temperatures during severe droughts in Ethiopia and southern Africa (Figs. S8 and S9). These results suggest that the CHIRTS<sub>max</sub>, CHIRTS<sub>max</sub> and CHIRTS<sub>max</sub> can provide a useful basis for monitoring temperature extremes.

**6. Conclusions**

While the CRU effectively captures temperatures on continental and global scales, continents like South America and Africa are poorly instrumented (Fig. 1), and decreases in the number of reporting stations have resulted in reductions in variance explained (Fig. 10). With only 41 and 50 observations in Africa and South

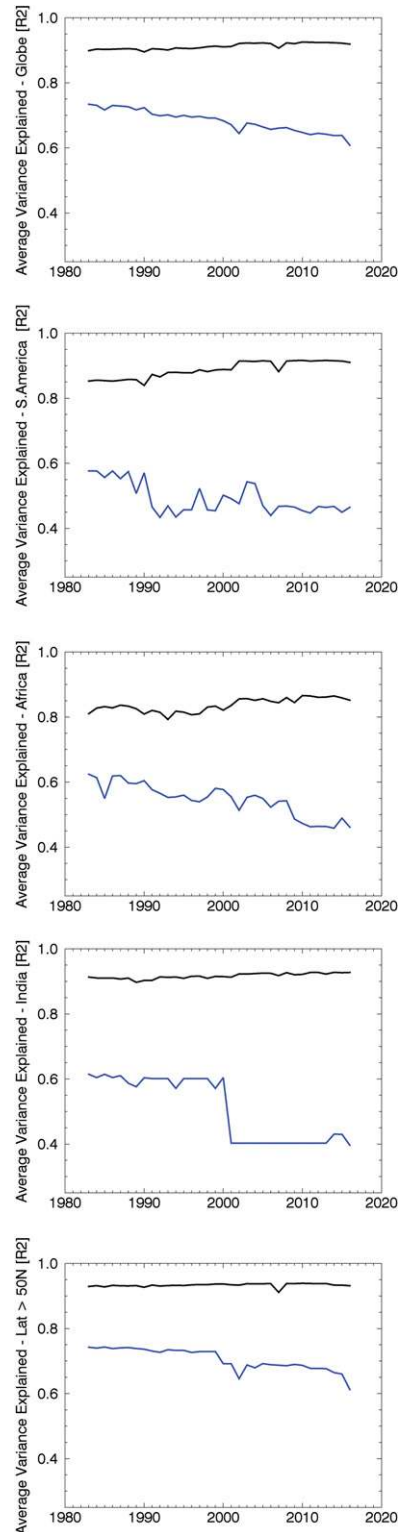


FIG. 10. Time series of average local 1983–2016 estimated variance explained for January CHIRTS<sub>max</sub> (black) and CRU (blue) for (from top to bottom) the Globe, South America, Africa, India, and regions north of 50°N. Regions are shown in Fig. 9.

America, respectively, in 2016, the CRU misses important extreme temperature events (Figs. 1d,e). For precipitation, it has become common to use satellite-based precipitation proxies to help overcome large data gaps. There has been little work, however, that uses TIR data to derive global  $T_{\max}$  proxies. Here, we have described such a process and found that it works quite well. Building on the extensive efforts that have gone into the development of the GridSat archive (Knapp et al. 2011), this algorithm allows us to estimate monthly  $T_{\max}$  in remote regions. We have also shown that reasonably dense station networks with good global coverage (Fig. 1b) can be derived by combining the Berkeley Earth and GTS networks.

It should be noted, however, that one systematic issue can be identified with the CHIRT<sub>max</sub> estimates. While comparison with the Berkeley-GTS archive identifies fairly high levels of correlation ( $\sim 0.75$ , Table 2), scatterplots of these estimates (Figs. 5b,e) exhibit an underestimation of the variance of  $T_{\max}$  anomalies. The CHIRT<sub>max</sub> anomalies fall above and below the one-to-one line, ranging from about  $-5^{\circ}$  to  $5^{\circ}\text{C}$  rather than  $-10^{\circ}$  to  $10^{\circ}\text{C}$ . We believe this underestimation is likely because the standard deviation of the gridded CRU data was used to convert standardized brightness temperature anomalies into estimates of 2-m maximum air temperature anomalies [Eq. (5)]. The standard deviation of gridded air temperatures will be lower than the standard deviation of air temperatures estimated at weather stations. Future work will examine this issue. One solution might be to build high-resolution climatologies of standard deviations.

Estimates of the relative variance explained by the CRU versus the CHIRTS<sub>max</sub> (Figs. 9, 10) suggest that the extra information contained in the Berkeley-GTS archive and CHIRT<sub>max</sub> estimates can explain much of the variability of regional  $T_{\max}$  variations. This is especially true in South America, Africa, the Middle East, and India, where many large countries have no or few CRU stations, especially in the past 5 or 10 years. The inclusion of reasonable satellite-based proxy can help overcome this limitation, and the CHIRTS<sub>max</sub> dataset should provide a valuable resource for climate change studies, climate extreme analyses, and early warning applications.

As the Earth warms rapidly, weather and climate extremes are contributing to extremely costly and sometimes deadly droughts, floods, and fires (Figueres et al. 2018). Understanding and predicting these extremes will be a grand challenge for twenty-first-century climate science (Zhang et al. 2013). Increasing air temperatures can increase the frequency or intensity of many types of extreme events (National Academies of Sciences,

Engineering, and Medicine 2016) including heat waves, pluvials and floods, fires, and droughts. The CHIRTS<sub>max</sub> product can help identify, quantify, and explore variations in these extreme temperatures. (The CHIRT, CHTS and CHIRTS<sub>max</sub> data can be accessed at <https://edcintl.cr.usgs.gov/downloads/sciweb1/shared/fews/web/global/monthly/> and [ftp://ftp.chg.ucsb.edu/pub/org/chg/products/Tmax\\_monthly/](ftp://ftp.chg.ucsb.edu/pub/org/chg/products/Tmax_monthly/)).

*Acknowledgments.* This paper has benefited greatly from the expert technical editing of Juliet Way-Henthorne and thoughtful comments from our anonymous reviewers and our editor Dr. Xianglei Huang. This research was supported by the U.S. Geological Survey's Drivers of Drought program and Cooperative Agreement G14AC00042, SERVIR-AST Grant NNN15ZDA001N-SERVIR, the NASA Harvest Program Grant Z60592017, and the NSF INFEWS/T1 project 2003169999: Understanding multi-scale resilience options for climate-vulnerable Africa. We gratefully acknowledge the support of the Defense Advanced Research Projects Agency (DARPA) World Modelers Program under Army Research Office (ARO) Prime Contract W911NF-18-C-0012. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the view of DARPA or ARO.

## REFERENCES

- Adler, R. F., and Coauthors, 2003: The Version-2 Global Precipitation Climatology Project (GPCP) monthly precipitation analysis (1979–present). *J. Hydrometeor.*, **4**, 1147–1167, [https://doi.org/10.1175/1525-7541\(2003\)004<1147:TVGPCP>2.0.CO;2](https://doi.org/10.1175/1525-7541(2003)004<1147:TVGPCP>2.0.CO;2).
- Arkin, P. A., and B. N. Meisner, 1987: The relationship between large-scale convective rainfall and cold cloud over the western hemisphere during 1982–84. *Mon. Wea. Rev.*, **115**, 51–74, [https://doi.org/10.1175/1520-0493\(1987\)115<0051:TRBLSC>2.0.CO;2](https://doi.org/10.1175/1520-0493(1987)115<0051:TRBLSC>2.0.CO;2).
- Brown, M., and Coauthors, 2015: Climate change, global food security, and the U.S. food system. U.S. Global Change Research Program, 146 pp., [https://www.usda.gov/oce/climate\\_change/FoodSecurity2015Assessment/FullAssessment.pdf](https://www.usda.gov/oce/climate_change/FoodSecurity2015Assessment/FullAssessment.pdf).
- Diffenbaugh, N. S., and Coauthors, 2017: Quantifying the influence of global warming on unprecedented extreme climate events. *Proc. Natl. Acad. Sci. USA*, **114**, 4881–4886, <https://doi.org/10.1073/pnas.1618082114>.
- Dinku, T., C. Funk, P. Peterson, R. Maidment, T. Tadesse, H. Gadain, and P. Ceccato, 2018: Validation of the CHIRPS satellite rainfall estimates over eastern Africa. *Quart. J. Roy. Meteor. Soc.*, **144** (Suppl. 1), 292–312, <https://doi.org/10.1002/qj.3244>.
- Draper, C. S., R. H. Reichle, and R. D. Koster, 2018: Assessment of MERRA-2 land surface energy flux estimates. *J. Climate*, **31**, 671–691, <https://doi.org/10.1175/JCLI-D-17-0121.1>.
- Figueres, C., C. Le Quéré, A. Mahindra, O. Bäte, G. Whiteman, G. Peters, and D. Guan, 2018: Emissions are still rising: Ramp

- up the cuts. *Nature*, **564**, 27–30, <https://doi.org/10.1038/d41586-018-07585-6>.
- Funk, C., A. Verdin, J. Michaelsen, P. Peterson, D. Pedreros, and G. Husak, 2015a: A global satellite assisted precipitation climatology. *Earth Syst. Sci. Data*, **7**, 275–287, <https://doi.org/10.5194/essd-7-275-2015>.
- , and Coauthors, 2015c: The climate hazards infrared precipitation with stations—A new environmental record for monitoring extremes. *Sci. Data*, **2**, 150066, <https://doi.org/10.1038/sdata.2015.66>.
- 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>.
- Harris, I., P. D. Jones, T. J. Osborn, and D. H. Lister, 2014: Updated high-resolution grids of monthly climatic observations—The CRU TS3.10 dataset. *Int. J. Climatol.*, **34**, 623–642, <https://doi.org/10.1002/joc.3711>.
- Hong, Y., K.-L. Hsu, S. Sorooshian, and X. Gao, 2004: Precipitation estimation from remotely sensed imagery using an artificial neural network cloud classification system. *J. Appl. Meteor.*, **43**, 1834–1853, <https://doi.org/10.1175/JAM2173.1>.
- Huffman, G. J., R. F. Adler, B. Rudolf, U. Schneider, and P. R. Keehn, 1995: Global precipitation estimates based on a technique for combining satellite-based estimates, rain gauge analysis, and NWP model precipitation information. *J. Climate*, **8**, 1284–1295, [https://doi.org/10.1175/1520-0442\(1995\)008<1284:GPEBOA>2.0.CO;2](https://doi.org/10.1175/1520-0442(1995)008<1284:GPEBOA>2.0.CO;2).
- , and Coauthors, 2007: The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. *J. Hydrometeorol.*, **8**, 38–55, <https://doi.org/10.1175/JHM560.1>.
- , R. F. Adler, D. T. Bolvin, and G. Gu, 2009: Improving the global precipitation record: GPCP version 2.1. *Geophys. Res. Lett.*, **36**, L17808, <https://doi.org/10.1029/2009GL040000>.
- Hutchinson, M. F., 1995: Interpolating mean rainfall using thin plate smoothing splines. *Int. J. Geogr. Inf. Syst.*, **9**, 385–403, <https://doi.org/10.1080/02693799508902045>.
- Janowiak, J. E., R. J. Joyce, and Y. Yarosh, 2001: A real-time global half-hourly pixel-resolution infrared dataset and its applications. *Bull. Amer. Meteor. Soc.*, **82**, 205–217, [https://doi.org/10.1175/1520-0477\(2001\)082<0205:ARTGHH>2.3.CO;2](https://doi.org/10.1175/1520-0477(2001)082<0205:ARTGHH>2.3.CO;2).
- Knapp, K. R., and Coauthors, 2011: Globally gridded satellite observations for climate studies. *Bull. Amer. Meteor. Soc.*, **92**, 893–907, <https://doi.org/10.1175/2011BAMS3039.1>.
- Li, Z.-L., and Coauthors, 2013: Satellite-derived land surface temperature: Current status and perspectives. *Remote Sens. Environ.*, **131**, 14–37, <https://doi.org/10.1016/j.rse.2012.12.008>.
- Menne, M. J., 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>.
- National Academies of Sciences, Engineering, and Medicine, 2016: *Attribution of Extreme Weather Events in the Context of Climate Change*. National Academies Press, 186 pp., <https://doi.org/10.17226/21852>.
- New, M., M. Hulme, and P. Jones, 2000: Representing twentieth-century space–time climate variability. Part II: Development of 1901–96 monthly grids of terrestrial surface climate. *J. Climate*, **13**, 2217–2238, [https://doi.org/10.1175/1520-0442\(2000\)013<2217:RTCSTC>2.0.CO;2](https://doi.org/10.1175/1520-0442(2000)013<2217:RTCSTC>2.0.CO;2).
- , D. Lister, M. Hulme, and I. Makin, 2002: A high-resolution data set of surface climate over global land areas. *Climate Res.*, **21**, 1–25, <https://doi.org/10.3354/cr021001>.
- Pepin, N., E. Maeda, and R. Williams, 2016: Use of remotely sensed land surface temperature as a proxy for air temperatures at high elevations: Findings from a 5000 m elevational transect across Kilimanjaro. *J. Geophys. Res. Atmos.*, **121**, 9998–10015, <https://doi.org/10.1002/2016jd025497>.
- Peterson, T. C., and R. S. Vose, 1997: An overview of the Global Historical Climatology Network temperature database. *Bull. Amer. Meteor. Soc.*, **78**, 2837–2849, [https://doi.org/10.1175/1520-0477\(1997\)078<2837:A00TGH>2.0.CO;2](https://doi.org/10.1175/1520-0477(1997)078<2837:A00TGH>2.0.CO;2).
- , T. R. Karl, P. F. Jamason, R. Knight, and D. R. Easterling, 1998: First difference method: Maximizing station density for the calculation of long-term global temperature change. *J. Geophys. Res.*, **103**, 25 967–25 974, <https://doi.org/10.1029/98JD01168>.
- Raskin, R. G., C. C. Funk, S. R. Webber, and C. J. Willmott, 1997: Spherkit: The Spatial Interpolation Toolkit. National Center for Geographic Information and Analysis Tech. Rep. 97–4, 45 pp.
- Raymo, M. E., B. Grant, M. Horowitz, and G. H. Rau, 1996: Mid-Pliocene warmth: Stronger greenhouse and stronger conveyor. *Mar. Micropaleontol.*, **27**, 313–326, [https://doi.org/10.1016/0377-8398\(95\)00048-8](https://doi.org/10.1016/0377-8398(95)00048-8).
- Reichle, R. H., Q. Liu, R. D. Koster, C. S. Draper, S. P. Mahanama, and G. S. Partyka, 2017: Land surface precipitation in MERRA-2. *J. Climate*, **30**, 1643–1664, <https://doi.org/10.1175/JCLI-D-16-0570.1>.
- Reynolds, R. W., N. A. Rayner, T. M. Smith, D. C. Stokes, and W. Wang, 2002: An Improved in situ and satellite SST analysis for climate. *J. Climate*, **15**, 1609–1625, [https://doi.org/10.1175/1520-0442\(2002\)015<1609:AHSAS>2.0.CO;2](https://doi.org/10.1175/1520-0442(2002)015<1609:AHSAS>2.0.CO;2).
- Rohde, R., and Coauthors, 2013a: A new estimate of the average Earth surface land temperature spanning 1753 to 2011. *Geoinf. Geostat.: Overview*, **1**, <https://doi.org/10.4172/2327-4581.1000101>.
- , and Coauthors, 2013b: Berkeley Earth Temperature Averaging Process. *Geoinf. Geostat.: Overview*, **1**, <https://doi.org/10.4172/2327-4581.1000103>.
- Sorooshian, S., K.-L. Hsu, X. Gao, H. V. Gupta, B. Imam, and D. Braithwaite, 2000: Evaluation of PERSIANN system satellite-based estimates of tropical rainfall. *Bull. Amer. Meteor. Soc.*, **81**, 2035–2046, [https://doi.org/10.1175/1520-0477\(2000\)081<2035:EOPSSSE>2.3.CO;2](https://doi.org/10.1175/1520-0477(2000)081<2035:EOPSSSE>2.3.CO;2).
- Taylor, K. E., R. J. Stouffer, and G. A. Meehl, 2012: An overview of CMIP5 and the experiment design. *Bull. Amer. Meteor. Soc.*, **93**, 485–498, <https://doi.org/10.1175/BAMS-D-11-00094.1>.
- Trenberth, K. E., A. Dai, R. M. Rasmussen, and D. B. Parsons, 2003: The changing character of precipitation. *Bull. Amer. Meteor. Soc.*, **84**, 1205–1217, <https://doi.org/10.1175/BAMS-84-9-1205>.
- , —, G. van der Schrier, P. D. Jones, J. Barichivich, K. R. Briffa, and J. Sheffield, 2014: Global warming and changes in drought. *Nat. Climate Change*, **4**, 17–22, <https://doi.org/10.1038/nclimate2067>.
- Van Leeuwen, W. J., A. R. Huete, and T. W. Laing, 1999: MODIS vegetation index compositing approach: A prototype with AVHRR data. *Remote Sens. Environ.*, **69**, 264–280, [https://doi.org/10.1016/S0034-4257\(99\)00022-X](https://doi.org/10.1016/S0034-4257(99)00022-X).
- Verdin, K. L., and S. K. Greenlee, 1996: Development of continental scale digital elevation models and extraction of hydrographic features. *Third Int. Conf./Workshop on Integrating GIS and Environmental Modeling*, Santa Fe, NM, National Center for Geographic Information and Analysis, 21–26.

- Wan, Z., Y. Zhang, Q. Zhang, and Z.-L. Li, 2004: Quality assessment and validation of the MODIS global land surface temperature. *Int. J. Remote Sens.*, **25**, 261–274, <https://doi.org/10.1080/0143116031000116417>.
- Weller, E., S.-K. Min, W. Cai, F. W. Zwiers, Y.-H. Kim, and D. Lee, 2016: Human-caused Indo-Pacific warm pool expansion. *Sci. Adv.*, **2**, e1501719, <https://doi.org/10.1126/sciadv.1501719>.
- Willmott, C. J., and S. M. Robeson, 1995: Climatologically aided interpolation (CAI) of terrestrial air temperature. *Int. J. Climatol.*, **15**, 221–229, <https://doi.org/10.1002/joc.3370150207>.
- Xie, P., and P. A. Arkin, 1997: Global precipitation: A 17-year monthly analysis based on gauge observations, satellite estimates, and numerical model outputs. *Bull. Amer. Meteor. Soc.*, **78**, 2539–2558, [https://doi.org/10.1175/1520-0477\(1997\)078<2539:GPAYMA>2.0.CO;2](https://doi.org/10.1175/1520-0477(1997)078<2539:GPAYMA>2.0.CO;2).
- Zhang, X., G. Hegerl, S. Seneviratne, R. Stewart, F. Zwiers, and L. Alexander, 2013: WCRP grand challenge: Understanding and predicting weather and climate extremes. World Climate Research Programme, 10 pp., [https://www.wcrp-climate.org/images/documents/grand\\_challenges/GC\\_Extremes\\_v2.pdf](https://www.wcrp-climate.org/images/documents/grand_challenges/GC_Extremes_v2.pdf).