• Data Description Article •

The Assessment of Global Surface Temperature Change from 1850s: The C-LSAT2.0 Ensemble and the CMST-Interim Datasets

Wenbin SUN[1,5], Qingxiang LI[*1,5], Boyin HUANG[2], Jiayi CHENG[1,5], Zhaoyang SONG[1,5], Haiyan LI[1,5], Wenjie DONG[1,5], Panmao ZHAI[3], and Phil JONES[4]

[1]*School of Atmospheric Sciences and Key Laboratory of Tropical Atmosphere−Ocean System*,
Ministry of Education, *Zhuhai* 519082, *China*

[2]*National Centers for Environmental Information*, *NOAA*, *Asheville*, *NC* 28801, *USA*

[3]*Chinese Academy of Meteorological Sciences*, *CMA*, *Beijing* 100081, *China*

[4]*Climatic Research Unit, University of East Anglia, Norwich NR*4 7*TJ, UK*

[5]*Southern Laboratory of Ocean Science and Engineering (Guangdong Zhuhai)*, *Zhuhai* 519082, *China*

ABSTRACT

Based on C-LSAT2.0, using high- and low-frequency components reconstruction methods, combined with observation constraint masking, a reconstructed C-LSAT2.0 with 756 ensemble members from the 1850s to 2018 has been developed. These ensemble versions have been merged with the ERSSTv5 ensemble dataset, and an upgraded version of the CMST-Interim dataset with 5° × 5° resolution has been developed. The CMST-Interim dataset has significantly improved the coverage rate of global surface temperature data. After reconstruction, the data coverage before 1950 increased from 78%−81% of the original CMST to 81%−89%. The total coverage after 1955 reached about 93%, including more than 98% in the Northern Hemisphere and 81%−89% in the Southern Hemisphere. Through the reconstruction ensemble experiments with different parameters, a good basis is provided for more systematic uncertainty assessment of C-LSAT2.0 and CMST-Interim. In comparison with the original CMST, the global mean surface temperatures are estimated to be cooler in the second half of 19th century and warmer during the 21st century, which shows that the global warming trend is further amplified. The global warming trends are updated from $0.085 \pm 0.004°C \ (10 \ yr)^{-1}$ and $0.128 \pm 0.006°C \ (10 \ yr)^{-1}$ to $0.089 \pm 0.004°C \ (10 \ yr)^{-1}$ and $0.137 \pm 0.007°C \ (10 \ yr)^{-1}$, respectively, since the start and the second half of 20th century.

Key words: C-LSAT2.0 ensemble datasets, CMST-Interim, EOTs, high- and low-frequency components, reconstruction

Dataset profile		
Dataset title	China-Land Surface Air Temperature 2.0	China Merged Surface Temperature-Interim
Time range	Monthly land surface temperature from 1850−2018	Monthly global surface temperature from 1854−2018
Geographic scope	Land only	global
Data format	".nc"	".nc"
Data volume	19.5 MB	20.0 MB
Data service system	http://atmos.sysu.edu.cn/ResearchDownload	
Sources of funding	Natural Science Foundation of China (Grant: 41975105) and the National Key R&D Program of China (Grant: 2018YFC1507705; 2017YFC1502301)	
Dataset composition	The dataset contains the LSAT file named "Reconstructed C-LSAT2.0.nc"	The dataset contains the ST file named "CMST-Interim.nc"

_____

* Corresponding author: Qingxiang LI
  Email: liqingx5@mail.sysu.edu.cn

1. Introduction

C-LSAT (China-Land Surface Air Temperature) and

CMST (China Merged Surface Temperature) are newly released Land Surface Air Temperature (LSAT) and global Surface Temperature (ST) datasets developed since 2018 (Xu et al., 2018; Yun et al., 2019). Previous studies (Li et al., 2020; 2021) have shown that the CMST dataset has similar trends and uncertainties of global ST when compared to other datasets [including HadCRUT4 (Morice et al., 2012), GISS (Hansen et al., 2010), NOAAGlobalTemp (Vose et al., 2012), and Berkeley Earth (BE) (Rohde et al., 2013)]. The significant ST trends during the "hiatus" period (1998−2012), which have been discussed in recent years, are identified by multiple datasets including CMST, Had-CRUTem4-Hybrid (infilled by the polar surface air temperature) (Cowtan and Way, 2014), and ERA-5 (Simmons et al., 2017). There are three reasons for the strong ST trend in the CMST dataset. First, ERSST of NOAA/NCEI is adopted in accordance with the CMST, NOAAGlobalTemp, and GISS datasets, which is different from the HadCRUTem4 and BE datasets, and the global sea surface temperature (SST) trend of ERSSTv5 from 1900 to 2018 is slightly higher than that of HadSST4 (Yun et al., 2019). Second, more station data collected by the International Surface Temperature Initiative (ISTI) (Thorne et al., 2011) have been added into C-LSAT2.0 ((Li et al., 2021). And third, the homogenization scheme for air temperature series adopted by CMA-LSAT differs from GHCNv4 (Xu et al., 2018). This indicates that both the C-LSAT2.0 and CMST datasets have some unique features in describing the recent global LSAT and ST changes. Nevertheless, C-LSAT2.0 and CMST are still not global "complete coverage" datasets, though ERSSTv5 integrated by CMST is a global "complete coverage" SST dataset.

The importance of a global "complete coverage" dataset is emphasized in recent studies of the "hiatus" period (1998−2012) (Cowtan and Way, 2014; Karl et al., 2015; Lewandowsky et al., 2016; Huang et al., 2017b; Medhaug et al., 2017; Simmons et al., 2017), especially for the observation of high-latitude regions such as the Arctic. There have been several attempts to develop a global "complete coverage" dataset, incorporating satellite observations (Cowtan and Way, 2014), reanalysis (Simmons et al., 2017), Arctic buoy observations (Huang et al., 2017b), the Deep Neural Network (DNN) algorithm (Kadow et al., 2020), and reconstructions by other statistical methods (Li and Tu, 2000; Cheng et al., 2020; Huang et al., 2020; Morice et al., 2020). These methods increase the potential coverage rate of global ST anomalies, but there are still some insufficiencies. For example, extending to the period before 1979 is difficult for satellite and reanalysis data, the air temperature data from buoys are different from traditional SST data merged with LSAT, and DNNs cannot reflect the changes of sea ice area in the Arctic along with global warming. The results of global ST trends from before and after reconstruction in Kadow et al. (2020) cannot replicate statistical differences. In other words, the difference in long-term trends before and after reconstruction does not exceed its uncertainty at the 95% confidence level, and the trends of the two could

be considered similar. Therefore, there is still a need to find methods for how to extend the benchmark data of limited regional coverage to global coverage and to quantify the uncertainties (Jones, 2016). Recently, all datasets mentioned above have been updated to the latest version (Rohde et al., 2013; Lenssen et al., 2019; Huang et al., 2020; Morice et al., 2020). The most significant change is that the coverage rate of the datasets used by IPCC AR5 has been greatly improved.

Considering the uncertainty of systematic bias in CRUTEM4, Morice et al. (2012) generated a 100-member ensemble dataset using different systematic bias correction settings, then combined the 100-member SST ensemble dataset with the HadSST3 dataset from Kennedy et al. (2011a, b) to form the HadCRUT4 global ST ensemble dataset and estimate its uncertainty. However, Morice et al. (2012) did not reconstruct some regions suffering from a lack of observations. More recently, based on the upgraded CRUTEM5 (Osborn et al., 2020), Morice et al. (2020) have developed two variants of the HadCRUT5 dataset. The first are the traditional gridded ST anomaly data where in situ observations are available, and the second extends temperature anomaly estimates into regions where the underlying measurements are informative using a Gaussian process-based statistical method.

The previous version of the ERSSTv3 dataset (Smith et al., 2008) divided the super-observations of SST anomalies into low- and high-frequency components and reconstructed them using spatial smoothing and empirical orthogonal teleconnections (EOTs, van den Dool et al., 2000), respectively. Therefore, the SST anomaly data have "complete coverage" for the global oceans. Based on this, Huang et al. (2015, 2017a) generated SST ensemble datasets ERSSTv4 and ERSSTv5 using different bias correction and reconstruction parameters and thus evaluated its uncertainty level. Huang et al. (2020) applied this method to the GHCNm4 dataset, merged it with the ERSSTv5 dataset to form a new NOAAGlobalTemp5 dataset, and evaluated its uncertainty level. Their results showed that the difference between the global ST dataset based on NOAAGlobalTemp5 and other datasets is within an uncertainty of 95% confidence level, which is consistent with Li et al. (2020).

Different methods have been used to interpolate continental and hemispheric variations to regions without observations (Li and Tu, 2000; Hansen et al., 2010; Cowtan and Way, 2014; Cheng et al., 2020; Huang et al., 2020; Kadow et al., 2020; Morice et al., 2020). This study uses EOT methods to reconstruct the C-LSAT2.0 dataset. We establish a "near-complete coverage" reconstruction of the LSAT ensemble dataset, then merge with the ERSSTv5 dataset to develop the reconstructed CMST-Interim dataset and evaluated its uncertainties.

Section 2 introduces the input datasets and the methodology of the reconstruction of the C-LSAT2.0 dataset and the uncertainty evaluation. Section 3 shows the process and results of the reconstruction of the C-LSAT 2.0 ensembles and the development of the CMST-Interim datasets. Section 4

introduces the reconstructed GMST series and its uncertainty based on the new CMST-Interim dataset. Section 5 gives a short summary, discussion, and suggests potential improvement for the future.

2. Data and methods

2.1. *Dataset inputs*

CMST uses C-LSAT (formerly named CMA-LSAT) as the LSAT component (Xu et al., 2018; Yun et al., 2019; Li et al., 2020). The current version is C-LSAT 2.0, with monthly LSAT anomalies on a 5° × 5° grid from 1850–2018. The C-LSAT2.0 dataset has certain advantages when compared to other datasets, such as GHCN4 and CRUTEM4, based on the number of observations (Xu et al., 2018; Yun et al., 2019). Because of these advantages, we are able to perform high- and low-frequency statistical reconstruction (see section 2.2; Smith et al., 2008) and develop an ensemble dataset for C-LSAT2.0, which further improves the spatial coverage rate of the dataset to a certain extent. The ocean component of CMST still uses NOAA/NCEI's ERSSTv5 (Huang et al., 2015), which uses statistical models to reconstruct the ocean surface temperature (except sea ice surface) data and develops an ensemble dataset. The uncertainty of global SST observations was systematically assessed. In addition, the ERA5 dataset (Hersbach et al., 2020) was used to train the EOT modes in the C-LSAT2.0 reconstruction, considering that its representativeness of global LSAT since 1979 is significantly better than other reanalysis datasets on global and regional scales (Chao et al., 2020).

2.2. *Methods*

2.2.1. *Reconstruction of the LSAT*

Similar to Smith et al. (2008) and Huang et al. (2015), the reconstruction of LSAT and C-LSAT2.0 adopts the method of reconstructing high- and low-frequency components separately. The low-frequency component mainly relies on the temporal and spatial moving average processing of the existing data, while the high-frequency component uses the EOTs reconstruction method. Then the high- and low-frequency components are synthesized into the reconstructed dataset. The reconstructed data are masked (section 3.4) through certain regular observation constraints to ensure the representativeness of the reconstructed temperature anomaly data.

First, we processed the low-frequency component of the original data, which represents large-scale changes of the anomaly of LSAT, in terms of both time and space. To separate low-frequency components, the original data were averaged to a 25° × 25° spatial running average. Second, we calculated the annual mean anomalies for LSAT, using a minimum number of two months of valid data in a year. Then, the annual anomaly LSAT was filtered with a 15-year median filter. After that, we used a 15° × 25° spatial running average, in latitude and longitude respectively, a 9-point binomial filter in space, and a 3-point binomial filter in time to fill the missing data. Finally, a 15° × 25° spatial running average was used to smooth the spatial distribution of annual anomaly LSAT.

The high-frequency component of LSAT was defined as the difference of the original data with the low-frequency component. We then fit the difference to a maximum of 100 modes of EOTs trained by the ERA5 reanalysis from 1979–2018. The selection of modes is based on the variance ratio according to observational coverage, and therefore the maximum number of modes is not sensitive to the final result. Before fitting, we localized EOT modes within a specified space. The localization in the longitudinal direction was to maintain EOT modes within the distance of 500 km from the base point, set it to zero when the distances are larger than 2500 km from the base point, and linearly decrease to zero in the range of 500 km and 2500 km from the base point. The localization in the latitudinal direction is similar to that for latitude, with the exception of varying damping distances. It linearly decreased in the range of 2000 km and 4000 km from the base point in the low-latitude region (22.5°N–22.5°S), in the range of 1000 km and 3000 km in the mid-latitude region (22.5°–57.5°N/S), and in the range of 500 km and 1500 km in the high-latitude (57.5°–90°N/S) region, while setting to zero when distances exceeded the maximum range in the three latitude zones and keeping the original EOT modes when distances were less than the minimum range. The method of fitting to EOT modes is:

$$R(x) = \sum_i f_i \Psi_i(x) , \qquad (1)$$

where $R(x)$ is reconstructed data, $\Psi_i(x)$ is ith EOTs mode, and f_i is the fitting coefficient, which is calculated by solving linear equations using the lower upper (LU) decomposition method (Press et al., 1992).

To avoid the influence of EOTs in sparse regions with low reliability, for the reconstruction, the EOT modes were accepted only if supported by sufficient observations, defined by Eq. (2):

$$r_i = \frac{\sum_x \Psi_i^2(x)\delta_x \cos\varphi_x}{\sum_x \Psi_i^2(x)\cos\varphi_x} , \qquad (2)$$

where r_i is the variance ratio of ith EOTs mode, δ_x is 0 with missing grid boxes, otherwise equal to 1, and $\cos\varphi_x$ is the cosine of latitude.

2.2.2. *Merging of the CMST-Interim dataset and assessing uncertainties*

The merging method and related steps of the CMST dataset follow Yun et al. (2019) and Li et al. (2020). The LSAT dataset use the above-mentioned C-LSAT2.0 dataset (1850–2018), and the SST data still use the ERSSTv5 dataset (1854–2018), so the temporal coverage of the CMST-Interim dataset after merging is still from January 1854 to December 2018.

For the C-LSAT2.0 component, through the above-mentioned EOTs reconstruction based on different parameters, C-LSAT2.0 has generated 756 ensemble members (3 × 3 × 3 × 7 × 4, Table 2). The parameter uncertainty estimation method is similar to that of the NOAAGloalTemp5 dataset (Huang et al., 2020). As the missing data after reconstruction are greatly reduced, to a certain extent, the reconstruction uncertainty can be significantly reduced, particularly for the globally averaged temperature. Therefore, this study only considers its parameter uncertainty. For the ERSSTv5 component, both types of uncertainties are considered like Huang et al. (2020), and the uncertainties for the final merged CMST-Interim dataset have been evaluated using the NOAAGloalTemp5 method (Huang et al., 2020; Li et al., 2020). According to Huang et al. (2020), the parameter uncertainty (ε_p) is the area-averaged LSAT uncertainty, as in Eq. (3) and Eq. (4):

$$\varepsilon_\mathrm{p}^2(t) = \frac{1}{M} \sum\nolimits_{m=1}^{M} \left[(A_\mathrm{g})_m(t) - \overline{A_\mathrm{g}(t)} \right]^2 , \qquad (3)$$

$$\overline{A_\mathrm{g}} = \frac{1}{M} \sum\nolimits_{m=1}^{M} (A_\mathrm{g})_m(t) , \qquad (4)$$

where M is the number of ensemble members (for this study M=756), $(A_\mathrm{g})_m$ represents global LSAT of m-member ensemble, $\overline{A_\mathrm{g}}$ is the average of all ensembles, and t represents temporal variations.

The annual uncertainty of GMST is composed of uncertainties from the land and marine components (Huang et al., 2020; Li et al., 2020). The annual uncertainty of the land component (GLSAT series) (U_L) is based on C-LSAT2.0. Accordingly, the uncertainty of C-LSAT2.0, assessed by the 5%−95% uncertainty levels, results from an ensemble approach of the parametric uncertainties. The estimation of the 5%−95% annual uncertainty range for the marine component (GSST series) (U_S) (based on ERSSTv5) (Huang et al., 2020) uses an ensemble approach of combining the reconstruction and parametric uncertainties together. We finally synthesized the total global annual uncertainty of the GLSAT series (U_G) (based on CMST-Interim) by using Eq. (5):

$$U_\mathrm{G}^2 = (0.29 \times U_\mathrm{L})^2 + (0.71 \times U_\mathrm{S})^2 , \qquad (5)$$

where 0.29 and 0.71 are the proportion of land and ocean areas to global area, respectively.

3. The reconstruction of C-LSAT2.0 and the merging of CMST-Interim

Considering that the ERSSTv5 is already a reconstructed ensemble dataset, we only reconstruct the LSAT dataset in this study. In order to show the effectiveness and reliability of reconstruction, this section gives a comparison of reconstruction results for several selected months: January 1850 (representing a period with very little data), January 1900 (representing a period with slightly more data), and January 1960 (represents a period with much more data). The specific process and results are explained in Section 3.1.

3.1. *Low-frequency component reconstruction*

Figure 1 shows the spatial distribution of the low-frequency component of the global LSAT dataset for the selected three months mentioned above (Figs. 1a−c) and the global average annual series (Fig. 1d). From Figs. 1a−c, it is obvious that for the first two selected months (January 1850 and January 1900), when there are fewer observational data, the low-frequency component reflects the more spatially smoothed results at large scales, and therefore, the spatial distribution difference is small. From Fig. 1c (January 1960), the low-frequency component reflects more regional/local differences. The global average annual change series of low-frequency components (Fig. 1d) basically reflects the long-term change characteristics of global average LSAT during the period of 1850−2018 (Li et al., 2020). The cooling trend lasted for nearly 10 years in the late 1870s, which was followed by a significant warming period from the 1890s to the 1940s. The mid-to-late 1940s−1970s experienced a cooling period of nearly 30 years. The LSAT entered a rapid warming period of more than 50 years after the late 1970s, and there is no more sign of "hiatus" or "cooling" trends since then.

3.2. *High-frequency component reconstruction*

The high-frequency component reconstruction uses an EOTs reconstruction method similar to ERSSTv4 (Section 2.1). Since C-LSAT2.0 is missing observations at many grid points, it is not conducive to the training of EOTs. We used the ERA5 reanalysis data (Hersbach et al., 2020) to conduct EOT modes training. Since EOTs do not require two directions to be orthogonal like EOFs (EOTs only require one direction to be orthogonal), EOT is less restrictive than EOF. From this point of view, EOTs are more like Rotational Empirical Orthogonal Function (REOFs, Li and Tu, 2002), but it should be noted that the calculation of EOTs is completed in one step, with no need to rotate the spatial modes, and there is no need to consider whether principal components are truncated in the calculation. Therefore, it is a very convenient and useful method (van den Dool et al., 2000). Figures 2a−p show the spatial modal distribution of the first 16 EOT modes trained with ERA5 reanalysis data. Obviously, these modes each indicate a representative spatial correlation distribution of global LSAT changes. Figure 2q shows the area covered by the first 100 EOTs. It also shows that the first 100 EOTs mentioned above have already completely covered the entire land area of the world, which corresponds with the reconstruction needs of this paper.

3.3. *Combination of high- and low-frequency components*

The C-LSAT2.0 anomaly fields from January 1850 to December 2018 were reconstructed according to the methods and process described in section 2.2. Figures 3a−c show
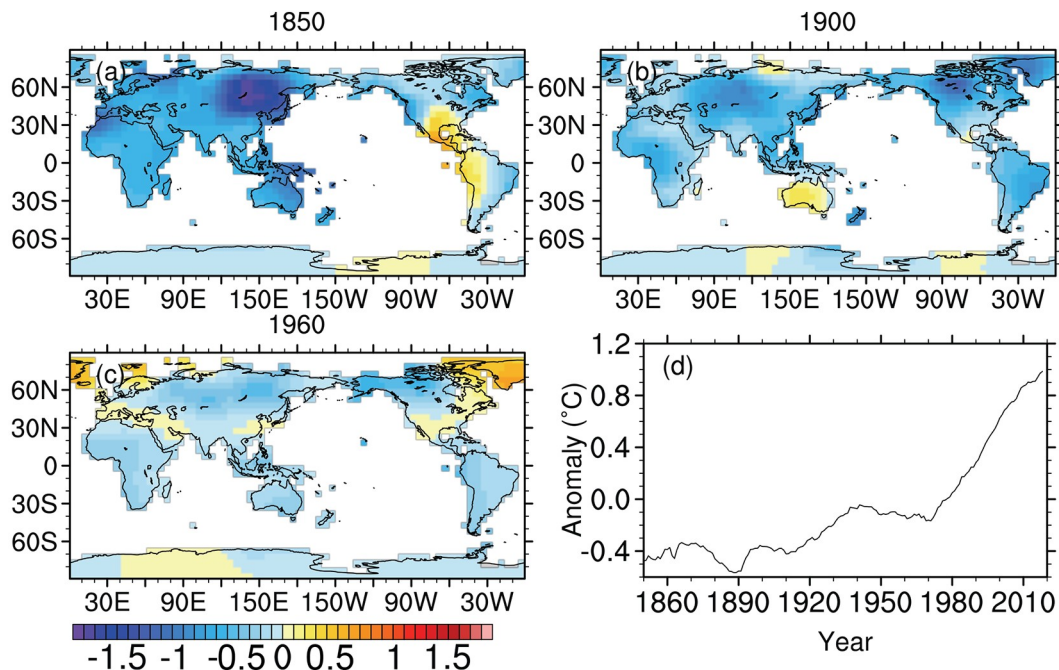
Fig. 1. The low-frequency component distribution of global LSAT anomalies at several typical times (1850.1, 1900.1 and 1960.1) and the global average low-frequency component series from 1850 to 2018.

the comparison examples between the reconstruction and the original observation results for the three aforementioned months (January 1850, January 1900, and January 1960). The cold and warm distribution of the reconstructed data and the magnitude of the high and low value centers are basically same as the original data, but the reconstructed data for years and regions with scarce data (such as in 1850s and in Antarctica) are mainly low-frequency data, which only reflects the long-term scale change characteristics. Obviously, the reconstruction can better reflect the anomaly distribution of observation information, and the grid boxes with the missing values are infilled and reconstructed, which has high reliability. Figure 3d shows the average deviation between the observational data and the reconstructed data (reduced to the grid boxes with observations) from January 1850 to December 2018. Obviously, as the coverage rate of data is further improved, this deviation is reduced. From the perspective of long-term changes, this deviation mainly reflects the fluctuations and oscillations near the zero value, and there is no obvious upward or downward trend. This also shows that the reconstruction described in this study has small changes in the long-term change trend of the global LSAT series, which further shows that the reconstruction is reasonable.

3.4. *Observation constraint masking of reconstructed data*

The analysis of the reconstruction process discussed above shows that the reconstruction and integration of high- and low-frequency components to reconstruct the current C-LSAT 2.0 better reflects the global LSAT trend and anomaly distribution. However, it also shows that the fewer the

observational data values, the lower the accuracy of the reconstructed data. Therefore, how to determine the extent to which the density of observation sites can be retained as reliable data for reconstruction will also be a very important issue.

Here, we present the designed principles to tailor (mask) the reconstructed data of C-LSAT2.0. When the correlation coefficient (explaining variance) between the observations and the reconstructed grid boxes series reaches a certain level, the reconstruction is considered reliable. Otherwise, the reconstruction is considered unreliable and masked out. HadCRUT5 adopts a parameter threshold masking method (Morice et al., 2020), but this study adopts a simpler method using a benchmark period from 1960 to 2018. It is generally believed that the observation stations are dense during this period and the reconstruction data are relatively reliable (Fig. 3d). We calculate the maximum value of the shortest distance between two grid boxes without missing observations during this period and use it as the mask threshold. Similar to Li et al. (2010), the corresponding relationship between the average distance between the grid boxes and the correlation coefficient is calculated. The comparison found that, except for individual grid points [the islands in Pacific (17.5°N, 157.5°W)] and the grid boxes in the Antarctic region, the maximum length between the grid boxes without missing observations is less than 1200 km. On average, this distance roughly corresponds to the correlation coefficient $r = 0.5$, and the explained variance between the two grid points series reaches 25%, which is a useful value. Based on this, in this study, the average distance when $d = 1200$ km ($r = 0.5$) is used as the threshold of the observation constraint masking of the reconstructed data.
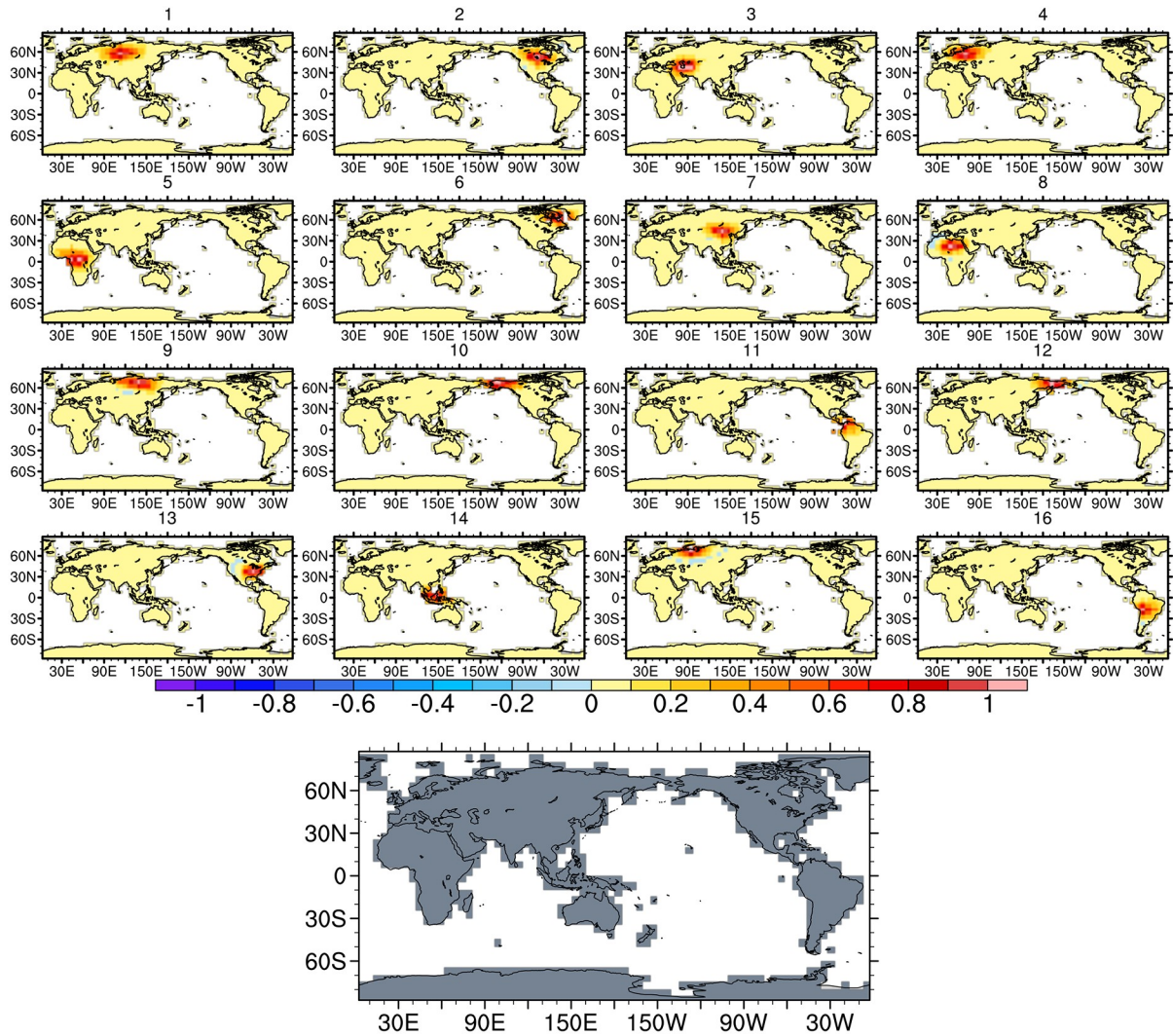
Fig. 2. The spatial modes (upper panel, 1−16) of the first 16 EOTs and the superimposed diagrams of all 100 EOTs (lower panel, 1−16).

Figure 4 also uses the above three months as examples to show comparison of the anomaly distribution of LSAT between before and after masking. Comparing Figs. 4a, 4b and 4c, it is clear that the observation constraint masks the anomaly data that only depends on the reconstruction of the low-frequency smooth component, while retaining the grid data reconstructed by both of the low- and the high-frequency components, which is more reasonable. It is worth noting that this kind of observation-constrained masking will hardly affect the establishment of global or regional temperature change series and long-term trend detection results.

Based on the reconstructed dataset after observation constraint masking, the 1850–2018 annual LSAT series of the eight regions in Xu et al. (2018) after reconstruction were calculated and compared with those before reconstruction (Fig. 5). Obviously, the long-term trend of the two are similar in all regions. However, the interannual variations in reconstructed series are significantly reduced before the 1900s and in the Antarctic region, which is mainly due to the sparse observations. Figure 5 also shows that the above

process still has certain problems with the reconstruction of the LSAT series in the Antarctic region. The main reason is that the observations in the Antarctic region are remarkably scarce. Even in the more completely data-covered period of January 1960, it is still relatively limited after the observation-constraint masking, and it was almost impossible to reconstruct reliable gridded data before that.

3.5. *LSAT ensemble dataset and uncertainty assessment*

When C-LSAT2.0 is reconstructed, the selection of different parameters will lead to specific uncertainties. Table 1 shows the different parameter settings during reconstruction, with a total of 756 ensemble members. Among them, the operational options use the intermediate values of various parameters (so-called "optimal" parameters). The dataset reconstructed with the operational optimal parameters is the basis for our daily product evaluation and scientific applications.

It can be seen from Fig. 6 that the differences in all 756 ensemble members obtained by the different parameters
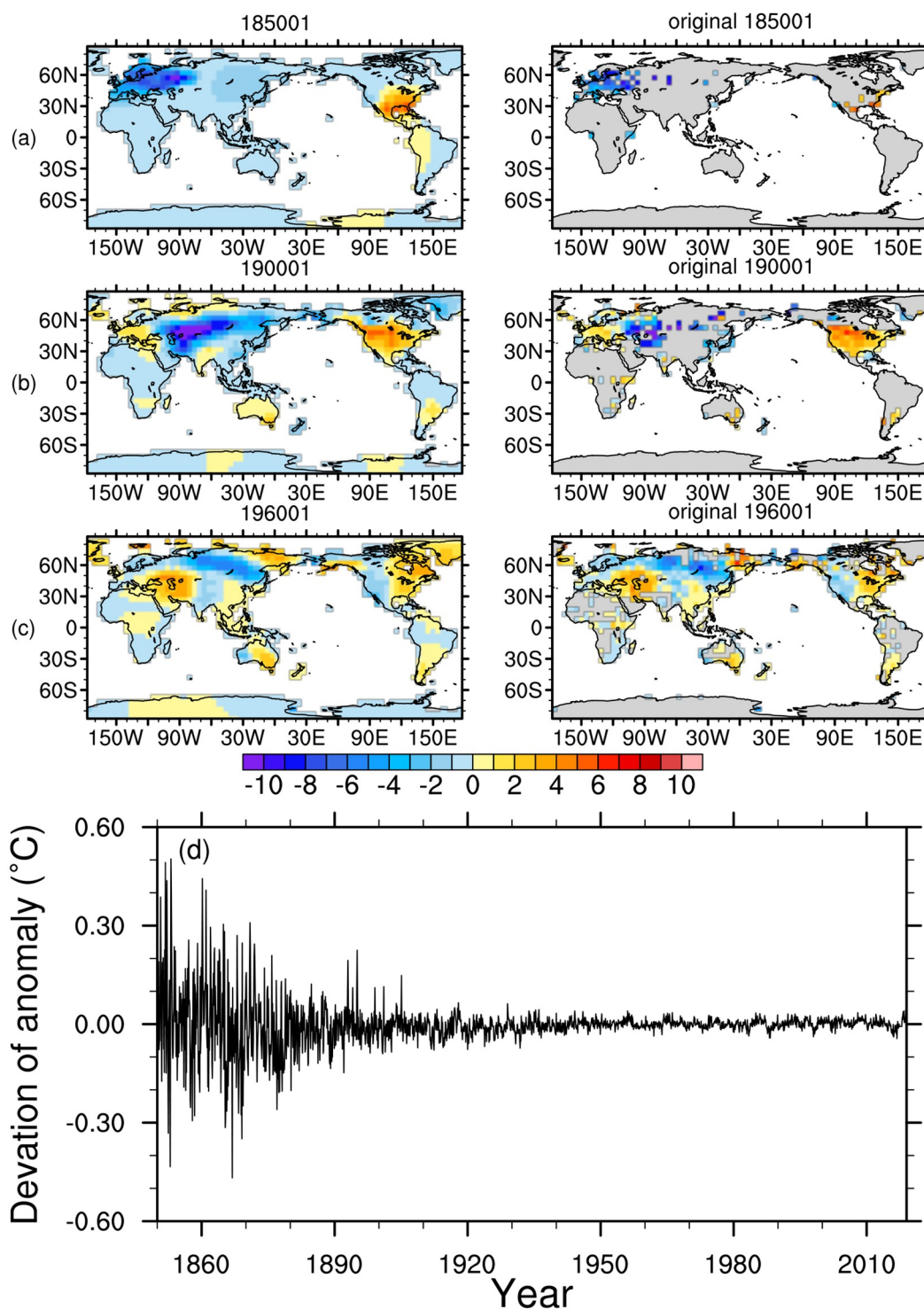
Fig. 3. (a)–(c) Comparison of global LSAT anomalies (units: °C) in the selected months and (d) monthly deviation reduced to the grids with observations before and after reconstruction.

given in Table 1 are still mainly concentrated before 1950, and the amplitude before the 1880s is the largest (from greater than 0.5°C to about 0.1°C), followed by 1880–1950s (from about 0.1°C to 0.05°C), and again after 1950 (below 0.025°C). It is worth mentioning that after the 2010s, the difference between ensemble members has slightly expanded (up to about 0.1°C). This is mainly due to the decrease in the number of data locations (or collection lag) in recent years. The uncertainty of the temperature anomalies in the last 10 years or so has increased, which partly explains why some studies have produced obviously different results in detection of the global ST trends during the "hiatus" period based on different datasets (Cowtan and Way, 2014; Karl et al., 2015; Lewandowsky et al., 2016;
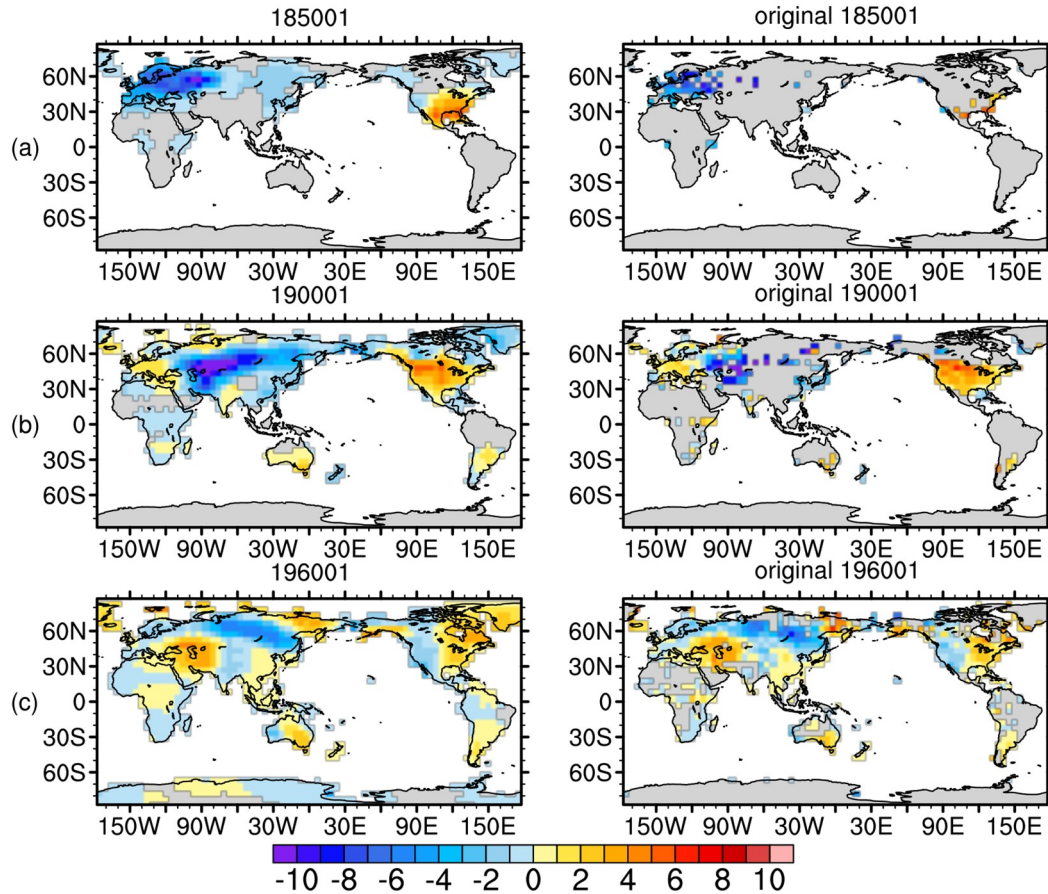
Fig. 4. The distribution of global LSAT anomalies (units: °C) before and after reconstruction for different periods (1850.1, 1900.1 and 1960.1) (same as Fig. 3a, the difference is that the reconstructed data are masked by using observation data constraints).

Huang et al., 2017b; Simmons et al., 2017; Li et al., 2021).

Similar to Li et al. (2020) and Huang et al. (2020), the uncertainty of LSAT is estimated (Fig. 7). Obviously, the parameter uncertainty is slightly lower than previous estimates (Li et al., 2020) (especially after the 1920s), but the inter-annual fluctuations have increased significantly more than the previous evaluation (Li et al., 2021). Overall, the uncertainty changes estimated in this study can better reflect inter-annual differences. It is worth pointing out that our uncertainty evaluation is slightly lower than that in Huang et al. (2020) while using a similar method, which is due to the lack of the reconstruction uncertainty evaluation for C-LSAT2.0 at this stage.

3.6. *The merge of CMST-Interim*

According to the practice of Yun et al. (2019), the reconstructed C-LSAT2.0 ensemble dataset and the ERSSTv5 ensemble dataset (provided by NCEI/NOAA) are merged together to obtain the reconstructed CMST-Interim dataset. Figure 8 shows a comparison of global data coverage rate before and after reconstruction (with the observation-constraint masking and sea-ice surface temperature included). Obviously, the reconstruction dataset greatly improves the global and regional data coverage rate, thereby increasing confidence in the data analysis. Figure 8a indicates the global

data coverage rate has changed from 78% in 1854 to about 93% in 1960 and has kept a coverage rate of more than 90% since then. The coverage rate for the reconstruction data of the Northern Hemisphere (including high and middle latitudes), low latitudes (30°N–30°S), and even the middle and high latitudes of the Southern Hemisphere (30°–60°S) has been close to 100% since the start of the second half of the 20th century. Some recent studies have used ice surface air temperature data instead of SST data, which has increased the coverage of global ST data to some extent (Rohde et al., 2013; Morice et al., 2020). However, since the area of sea ice often changes with time and is difficult to diagnose, it is hard to conclude whether using the ice-surface air temperature or the ice-surface temperature is more suitable. So, there is still a certain degree of uncertainty. Therefore, this study still adopts the most traditional method, and the ice-surface temperature is represented by a fixed value of –1.8°C.

4. The reconstructed GMST series and its uncertainty

4.1. *The differences between CMST and CMST-Interim*

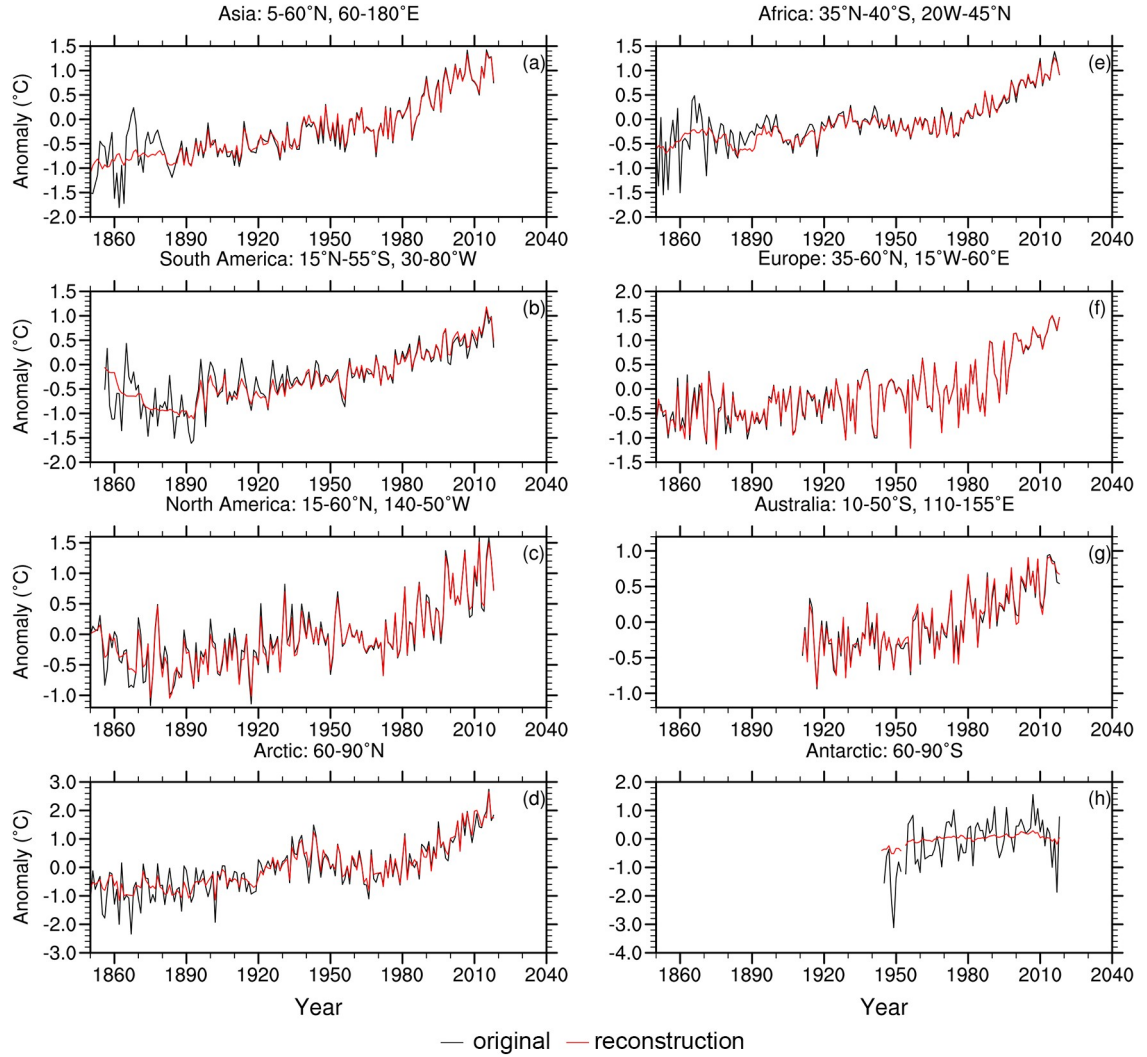Figure 9 shows the global and Southern and Northern

Fig. 5. Comparison of the regional LSAT series before and after reconstruction.

Table 1. Parameter settings used for reconstruction scenarios and the operational option.

Parameter	Operational options	Alternative options
Minimum number of months for annual average	2 months	1, 2, 3 months
LF filter periods	15 years	10, 15, 20 years
Min number of years for LF filter	2 years	1, 2, 3 years
EOTs training periods and spatial scales	1979−2018, Lx=4000, 3000, 2500, Ly=2500	1979−2018, Lx=3000, 2000, 1500, Ly=1500; 1979−2018, Lx=5000, 4000, 3500, Ly=3500; Lx=4000, 3000, 2500, Ly=2500; 1979−2008, Lx=4000, 3000, 2500, Ly=2500; 1989−2018, Lx=4000, 3000, 2500, Ly=2500; even year, Lx=4000, 3000, 2500, Ly=2500; odd year, Lx=4000, 3000, 2500, Ly=2500;
EOTs acceptance criterion	0.2	0.10, 0.15, 0.20, 0.25

Hemisphere average ST series and their comparison based on the reconstructed CMST-Interim and original CMST. From the global average series (Fig. 9a), the ST anomalies after reconstruction before the 19th century are slightly lower than those before reconstruction, while the ST anomalies after reconstruction from the 21st century are slightly higher than those of the unreconstructed CMST. In this way, the reconstructed global warming trend of the ST series is higher than that before reconstruction. This is mainly due to the fact that there are fewer grids of LSAT observations before reconstruction, leading to an increase in the proportion of sea temperature in the calculation of the
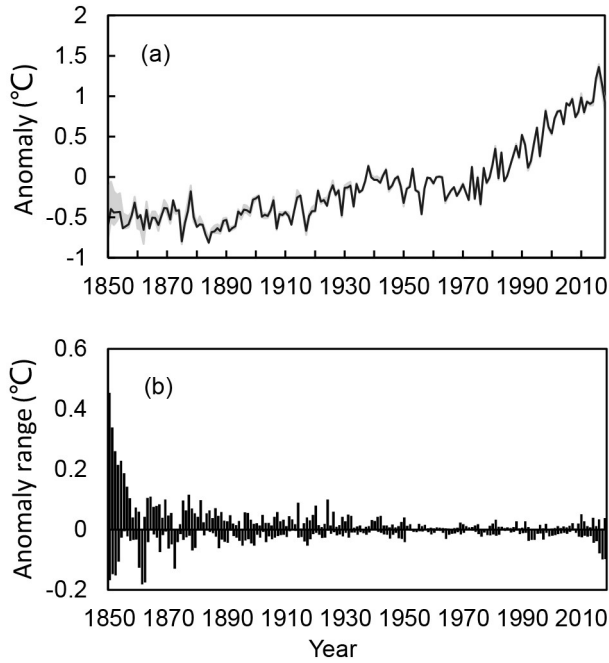
Fig. 6. (a) Spread for different reconstruction parameters (black: operational option; shadow: difference range of different parameters) and (b) range of difference between all series.
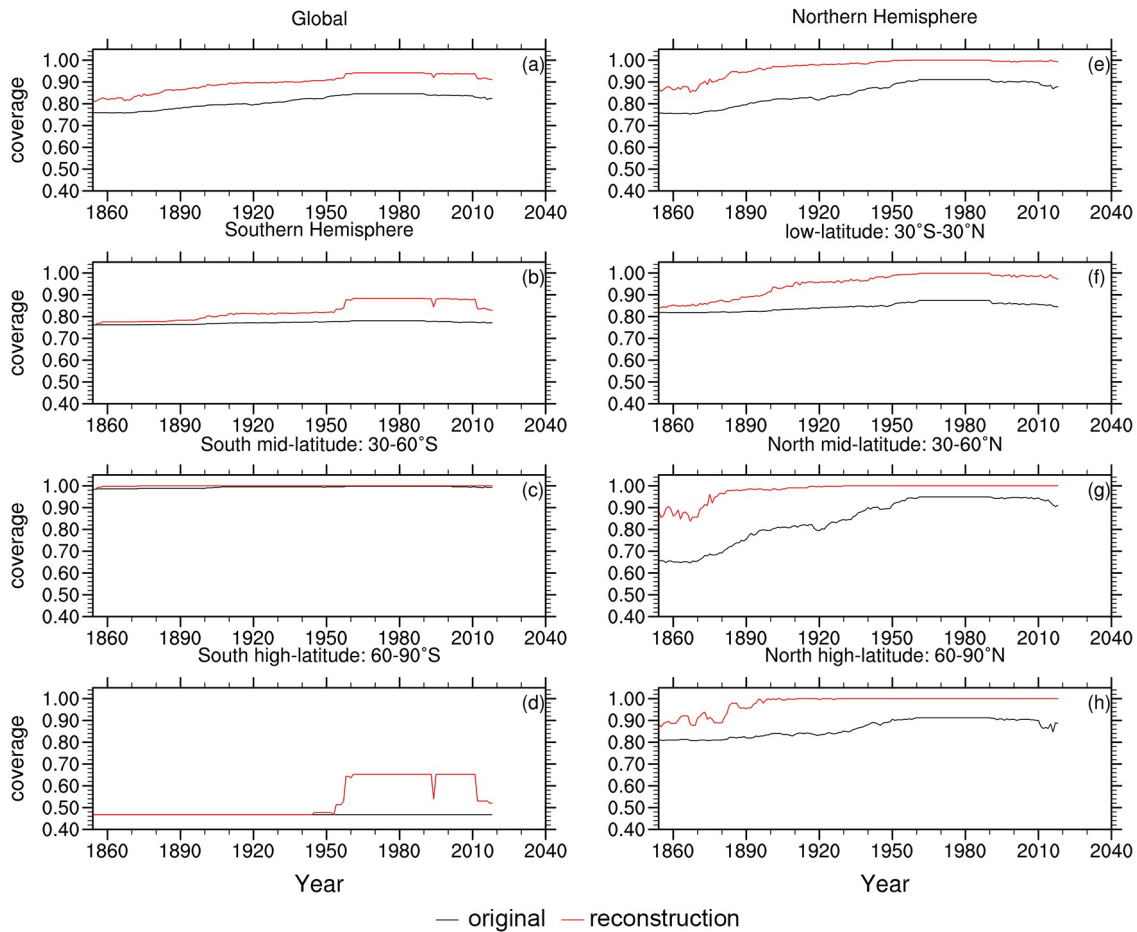
global average ST series in the 19th century. The trends of CMST-Interim and original CMST become similar after the land and sea area proportion has been correctly used. For the Northern Hemisphere series (Fig. 9b), the two series show similar characteristics, that is, the reconstructed ST anomalies before the 19th century slightly decreased and then slightly increased since the 21st century. For the Southern Hemisphere series (Fig. 9c), the series of ST anomalies before and after reconstruction is basically unchanged. That is to say, the reconstruction mainly changes the global ST anomalies by changing the average ST anomalies in the North-
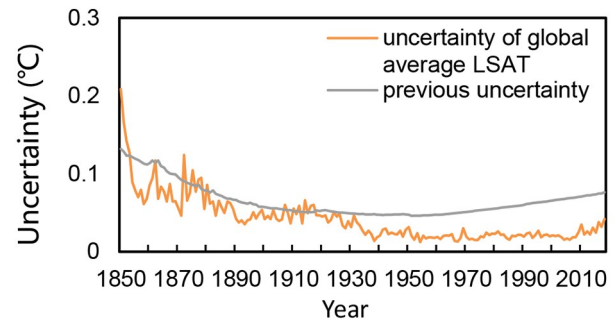

Fig. 7. Comparison of the 95% uncertainty evaluation between the reconstruction and original C-LSAT during the period of 1850–2018.

Fig. 8. Comparison of data coverage of global, hemisphere, and latitude zones before and after reconstruction in CMST-Interim.
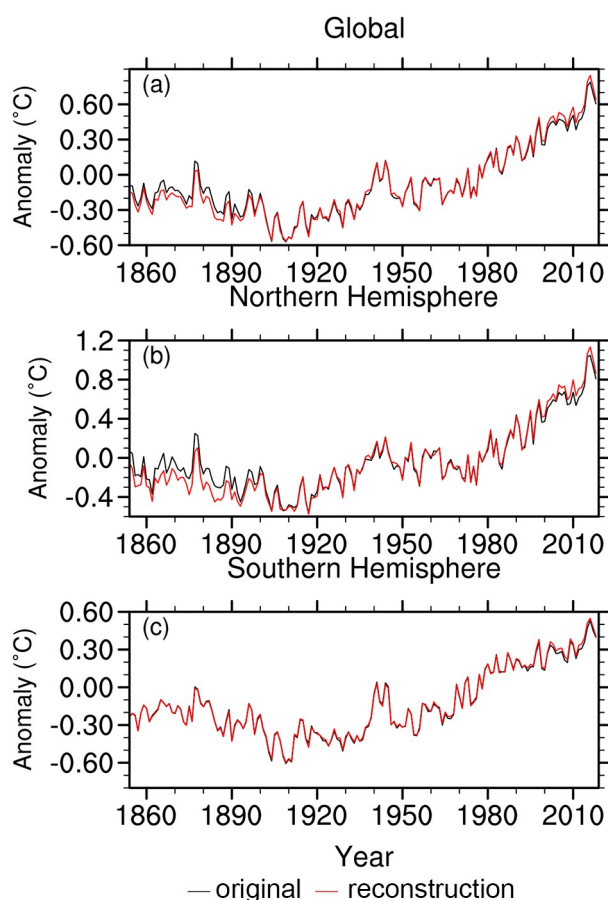
Global

Fig. 9. Comparison of the global (a), Northern Hemisphere (b), and Southern Hemisphere (c) surface temperature series.

ern Hemisphere.

In the ST series of each latitude zone (Fig. 10), the conclusions based on the original CMST and the reconstructed CMST-Interim also show quite good agreement. The same difference is shown in the Northern Hemisphere mid-latitudes (30°–60°N) and high latitudes (60°–90°N). That is, the ST anomalies decreased before 1900 and increased slightly after 2000. For low latitudes (30°S–30°N) and southern mid-latitudes (30°–60°S), there is almost no difference between the two. But, in the high latitudes of the Southern Hemisphere (60°–90°S), the CMST-Interim series starts from 1854 (the unreconstructed series can only start from 1950 due to lack of observation data). However, from the overlap period (1950–2018), the ST anomaly changes are more consistent.

4.2. *The GMST trends and its uncertainties*

The overall annual uncertainty of GMST is composed of uncertainties from the land and marine components (Huang et al., 2020; Li et al., 2020). The annual uncertainty of the land component (GLSAT series) (U_L) is based on C-LSAT2.0. According to section 3.6, the 5%–95% uncertainty levels are assessed using an ensemble approach for the parametric uncertainties (Fig. 7). Figure 11 is the GMST series based on the CMST-Interim from 1854 to 2018 along

with their uncertainty ranges at the 5%–95% level.

The differences of the warming trends between CMST and other global ST datasets have been comprehensively compared in Li et al (2021). Table 2 shows the warming trends for global and hemispheric ST change based on CMST-Interim from different time scales and the comparison with original CMST. For the global average ST change since 1854, the CMST-Interim gives linear trends of 0.048 ± 0.003, 0.070 ± 0.003, 0.089 ± 0.004, and 0.137 ± 0.007°C (10 yr)$^{-1}$, respectively, for the periods 1854–2018, 1880–2018, 1900–2018, and 1950–2018. Those are slightly higher than the CMST evaluation [0.042 ± 0.003, 0.064 ± 0.004, 0.085 ± 0.004, and 0.128 ± 0.006°C (10 yr)$^{-1}$, respectively, for the corresponding periods]. Similarly, the hemispheric mean ST trends also have varying degrees of enlargement. This shows that through the reconstruction of the C-LSAT2.0 dataset, the data of high latitudes and some high-altitude areas have been filled, and the coverage rate of the CMST-Interim has been greatly improved. Therefore, the global warming trend estimation has been amplified to a certain extent. According to Huang et al. (2020), the trends for period of 1880–2016 are 0.066, 0.070, 0.071, and 0.071°C (10 yr)$^{-1}$ for HadCRUT4, BE, GISTEMP, and NOAAGlobalTempv5 operational, respectively. Those are quite similar to our new evaluation in Table 2 (0.070 for CMST-Interim).

5. Summary

The method based on high- and low-frequency components reconstruction has effectively improved the coverage of the global LSAT dataset (C-LSAT2.0). Through the observation-constraint masking scheme, the reconstruction of C-LSAT2.0 is more in-line with the actual observations at both global and regional scales.

The comparison shows that the coverage rate of the data before 1950 after reconstruction (CMST-Interim) increased to 81%–89% from 78%–81% of the original CMST, and it reached about 93% after 1955. The coverage was greater than 98% in the Northern Hemisphere and between 81%–89% in the Southern Hemisphere. The lower coverage in the Southern Hemisphere is mainly due to insufficient observation coverage in the high latitudes.

The reconstructed dataset more accurately describes the characteristics of multi-scale changes in global and regional ST. Before the 19th century, the reconstructed global average ST anomalies based on CMST-Interim were slightly lower than the original CMST average anomalies. In contrast, during the early 21st century, the reconstructed global average ST anomalies were slightly higher than the original CMST average anomalies. This difference is mainly reflected in the reconstruction in the Northern Hemisphere, while the series of average ST anomalies before and after the reconstruction in the Southern Hemisphere is almost unchanged.

Through the reconstruction experiments with different parameters, a good foundation is provided for more systematic uncertainty evaluation of C-LSAT2.0 and CMST-
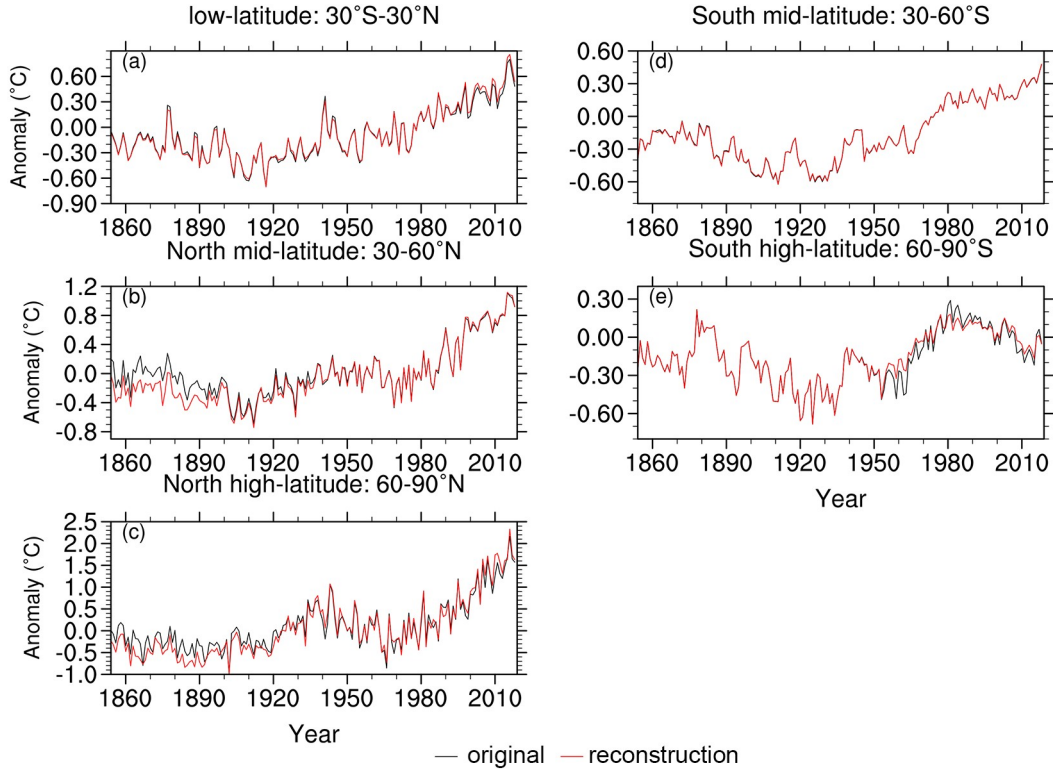
Fig. 10. The average ST series of different latitude zones before and after reconstruction (black: the unreconstructed CMST, red: reconstructed CMST-Interim).
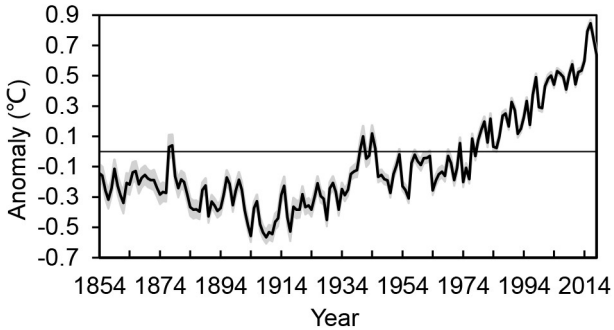

Fig. 11. the GMST based on CMST-Interim and its 5%−95% uncertainty range during the period of 1854−2019.

Interim (and future CMST2.0). Estimates show that the uncertainty of the ensemble reconstruction dataset for the global LSAT series is quite similar to previous estimations by combination of the observational, sampling, bias errors, and insufficient coverage (Brohan et al., 2006; Li et al., 2021). The uncertainty of the global and regional average ST series based on the reconstructed CMST-Interim is similar to previous estimates (Li et al., 2020).

It is also worth mentioning that there are still some deficiencies existing in the current studies: (1) For some regions such as Antarctica, due to the scarcity of actual observational data, reconstruction mainly relies on the influence of low-frequency components. This results in a certain deviation in the anomaly changes for the reconstruction data. Therefore, this study restricts the masking results based on the actual observation data, and only retains the grid data with both reasonable high- and low-frequency components. (2) This study has not considered the reconstruction uncertainty after observation constraints and masking, which may lead to a slightly lower level of uncertainty. (3) Since sea-ice area changes over time cannot be quantitatively described, in the current study the traditional fixed value of the sea-ice surface temperatures is used.

Table 2. Trends evaluation in global and hemispheric ST change from different scales [°C (10 yr)$^{-1}$].

Global/region	Datasets	1854–2018	1880–2018	1900–2018	1950–2018
Global	Original	0.042±0.003	0.064±0.004	0.085±0.004	0.128±0.006
	Reconstruction	0.048±0.003	0.070±0.003	0.089±0.004	0.137±0.007
North Hemisphere	Original	0.046±0.004	0.071±0.004	0.091±0.005	0.147±0.010
	Reconstruction	0.058±0.004	0.082±0.004	0.098±0.005	0.160±0.011
South Hemisphere	Original	0.037±0.003	0.057±0.003	0.078±0.003	0.108±0.005
	Reconstruction	0.038±0.003	0.058±0.003	0.079±0.003	0.113±0.005

Data access

The reconstructed C-LSAT2.0 and the upgraded version CMST-Interim datasets are available from http://atmos.sysu.edu.cn/ResearchDownload.

APPENDIX

Table A1. Abbreviations of datasets.

Acronym	Full name
C-LSAT	China-Land Surface Air Temperature
CRUTEM	Climatic Research Unit global surface air Temperature
GHCN	Global Historic Climate Network
ERSST	Extended Reconstruction Sea Surface Temperature
HadSST	Hadley Centre Sea Surface Temperature
BE	Berkeley Earth surface temperature
CMST	China Merged Surface Temperature
HadCRUT	Hadley Centre & Climatic Research Unit global surface Temperature
NOAAGlobalTemp	NOAA Global surface Temperature
ERA5	The 5th generation of ECMWF atmospheric reanalysis of the global climate

REFERENCES

Brohan, P., J. J. Kennedy, I. Harris, S. F. B. Tett, and P. D. Jones, 2006: Uncertainty estimates in regional and global observed temperature changes: A new data set from 1850. *J. Geophys. Res.*, **111**, D12106, https://doi.org/10.1029/2005JD006548.

Chao, L. Y., B. Y. Huang, Y. J. Yang, P. Jones, J. Y. Cheng, Y. Yang, and Q. X. Li, 2020: A new evaluation of the role of urbanization to warming at various spatial scales: Evidence from the Guangdong-Hong Kong-Macau Region, China. *Geophys. Res. Lett.*, **47**, e2020GL089152, https://doi.org/10.1029/2020GL089152.

Cheng, J. Y., Q. X. Li, L. Y. Chao, S. Maity, B. Y. Huang, and P. Jones, 2020: Development of high resolution and homogenized gridded land surface air temperature data: A case study over pan-east Asia. *Frontiers in Environmental Science*, **8**, 588570, https://doi.org/10.3389/fenvs.2020.588570.

Cowtan, K., and R. G. Way, 2014: Coverage bias in the HadCRUT4 temperature series and its impact on recent temperature trends. *Quart. J. Roy. Meteor. Soc.*, **140**, 1935−1944, https://doi.org/10.1002/qj.2297.

Hansen, J., R. Ruedy, M. Sato, and K. Lo, 2010: Global surface temperature change. *Rev. Geophys.*, **48**(4), RG404, https://doi.org/10.1029/2010RG000345.

Hersbach, H., and Coauthors, 2020: The ERA5 global reanalysis. *Quart. J. Roy. Meteor. Soc.*, **146**, 1999−2049, https://doi.org/10.1002/qj.3803.

Huang, B. Y., and Coauthors, 2015: Further exploring and quantifying uncertainties for Extended Reconstructed Sea Surface Temperature (ERSST) version 4 (v4). *J. Climate*, **29**, 3119−3142, https://doi.org/10.1175/JCLI-D-15-0430.1.

Huang, B. Y., and Coauthors, 2017a: Extended Reconstructed Sea Surface Temperature, version 5 (ERSSTv5): Upgrades, validations, and intercomparisons. *J. Climate*, **30**, 8179−8205, https://doi.org/10.1175/JCLI-D-16-0836.1.

Huang, B. Y., and Coauthors, 2020: Uncertainty estimates for sea surface temperature and land surface air temperature in NOAAGlobalTemp version 5. *J. Climate*, **33**, 1351−1379, https://doi.org/10.1175/JCLI-D-19-0395.1.

Huang, J. B., and Coauthors, 2017b: Recently amplified arctic warming has contributed to a continual global warming trend. *Nature Climate Change*, **7**, 875−879, https://doi.org/10.1038/s41558-017-0009-5.

Jones, P., 2016: The reliability of global and hemispheric surface temperature records. *Adv. Atmos. Sci.*, **33**(3), 269−282, https://doi.org/10.1007/s00376-015-5194-4.

Kadow, C., D. M. Hall, and U. Ulbrich, 2020: Artificial intelligence reconstructs missing climate information. *Nature Geoscience*, **13**, 408−413, https://doi.org/10.1038/s41561-020-0582-5.

Karl, T. R., and Coauthors, 2015: Possible artifacts of data biases in the recent global surface warming hiatus. *Science*, **348**, 1469−1472, https://doi.org/10.1126/science.aaa5632.

Kennedy, J. J., N. A. Rayner, R. O. Smith, D. E. Parker, and M. Saunby, 2011a: Reassessing biases and other uncertainties in sea surface temperature observations measured in situ since 1850: 2. Biases and homogenization. *J. Geophys. Res.*, **116**, D14104, https://doi.org/10.1029/2010JD015220.

Kennedy, J. J., N. A. Rayner, R. O. Smith, D. E. Parker, and M. Saunby, 2011b: Reassessing biases and other uncertainties in sea surface temperature observations measured in situ since 1850: 1. Measurement and sampling uncertainties. *J. Geophys. Res.*, **116**, D14103, https://doi.org/10.1029/2010JD015218.

Lenssen, N. J. L., G. A. Schmidt, J. E. Hansen, M. J. Menne, A. Persin, R. Ruedy, and D. Zyss, 2019: Improvements in the GISTEMP uncertainty model. *J. Geophys. Res.*, **124**(12), 6307–6326, https://doi.org/10.1029/2018JD029522.

Lewandowsky, S., J. S. Risbey, and N. Oreskes, 2016: The "pause" in global warming: Turning a routine fluctuation into a problem for science. *Bull. Amer. Meteor. Soc.*, **97**, 723–733, https://doi.org/10.1175/BAMS-D-14-00106.1.

Li, Q. X., and Q. P. Tu, 2000: Interpolation and primary analysis of northern hemisphere land surface precipitation in the past 100 years. *Journal of Nanjing Institute of Meteorology*, **23**(4), 528–535, https://doi.org/10.3969/j.issn.1674-7097.2000.04.009. (in Chinese with English abstract)

Li, Q. X., and Q. P. Tu, 2002: The regional characters of annual precipitation in northern hemisphere land surface and China and their correlation. *Journal of Nanjing Institute of Meteorology*, **25**(1), 92–99, https://doi.org/10.3969/j.issn.1674-7097.2002.01.012. (in Chinese with English abstract)

Li, Q. X., W. J. Dong, W. Li, X. R. Gao, P. Jones, J. Kennedy, and D. Parker, 2010: Assessment of the uncertainties in temperature change in China during the last century. *Chinese Science Bulletin*, **55**, 1974–1982, https://doi.org/10.1007/s11434-010-3209-1.

Li, Q. X., W. B. Sun, B. Y. Huang, W. J. Dong, X. L. Wang, P. M. Zhai, and P. Jones, 2020: Consistency of global warming trends strengthened since 1880s. *Science Bulletin*, **65**(20), 1709–1712, https://doi.org/10.1016/j.scib.2020.06.009.

Li, Q. X., and Coauthors, 2021: An updated evaluation of the global mean Land Surface Air Temperature and Surface Temperature trends based on CLSAT and CMST. *Climate Dyn.*, https://doi.org/10.1007/s00382-020-05502-0.

Medhaug, I., M. B. Stolpe, E. M. Fischer, and R. Knutti, 2017: Reconciling controversies about the 'global warming hiatus. *Nature*, **545**, 41–47, https://doi.org/10.1038/nature22315.

Morice, C. P., J. J. Kennedy, N. A. Rayner, and P. D. Jones, 2012: Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The HadCRUT4 data set. *J. Geophys. Res.*, **117**, D08101, https://doi.org/10.1029/2011JD017187.

Morice, C. P., and Coauthors, 2020: An updated assessment of near-surface temperature change from 1850: The HadCRUT5 dataset. *J. Geophys. Res.*, https://doi.org/10.1029/2019JD032361.

Osborn, T. J., P. D. Jones, D. H. Lister, C. P. Morice, I. R. Simpson, J. P. Winn, E. Hogan, and I. C. Harris, 2020: Land surface air temperature variations across the globe updated to 2019: The CRUTEM5 dataset. *J. Geophys. Res.*, **125**, e2019JD032352, https://doi.org/10.1029/2019JD032352.

Press, W. H., B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, 1992: LU decomposition and its applications. *Numerical Recipes in FORTRAN*: *The Art of Scientific Computing*, 2nd ed., Cambridge University Press, 34–42.

Rohde, R., and Coauthors, 2013: A new estimate of the average earth surface land temperature spanning 1753 to 2011. *Geoinformatics & Geostatistics: An Overview*, **1**, 1, https://doi.org/10.4172/2327-4581.1000101.

Simmons, A. J., P. Berrisford, D. P. Dee, H. Hersbach, S. Hirahara, and J. N. Thépaut, 2017: A reassessment of temperature variations and trends from global reanalyses and monthly surface climatological datasets. *Quart. J. Roy. Meteor. Soc.*, **143**, 101–119, https://doi.org/10.1002/qj.2949.

Smith, T. M., R. W. Reynolds, T. C. Peterson, and J. Lawrimore, 2008: Improvements to NOAA's historical merged land-ocean surface temperature analysis (1880-2006). *J. Climate*, **21**, 2283–2296, https://doi.org/10.1175/2007JCLI2100.1.

Thorne, P. W., and Coauthors, 2011: Guiding the creation of a comprehensive surface temperature resource for twenty-first-century climate science. *Bull. Amer. Meteor. Soc.*, **92**(11), ES40–ES47, https://doi.org/10.1175/2011BAMS3124.1.

van den Dool, H. M., S. Saha, and Å. Johansson, 2000: Empirical orthogonal teleconnections. *J. Climate*, **13**(8), 1421–1435, https://doi.org/10.1175/1520-0442(2000)013<1421:EOT>2.0.CO;2.

Vose, R. S., and Coauthors, 2012: NOAA's merged land-ocean surface temperature analysis. *Bull. Amer. Meteor. Soc.*, **93**(11), 1677–1685, https://doi.org/10.1175/BAMS-D-11-00241.1.

Xu, W. H., and Coauthors, 2018: A new integrated and homogenized global monthly land surface air temperature dataset for the period since 1900. *Climate Dyn.*, **50**, 2513–2536, https://doi.org/10.1007/s00382-017-3755-1.

Yun, X., B. Y. Huang, J. Y. Cheng, W. H. Xu, S. B. Qiao, and Q. X. Li, 2019: A new merge of global surface temperature datasets since the start of the 20th Century. *Earth System Science Data*, **11**, 1629–1643, https://doi.org/10.5194/essd-11-1629-2019.