An Integrated Uncertainty Framework for the China-MST 3.0 Global Surface Temperature Dataset

Zichen Li[1], Qingxiang Li[1,2#*], Boyang Jiao[3], Qiya Xu[1], Sihao Wei[1], Xutong Ru[1], Peng Si[4], Liya Chao[5], Hanyu Zhang[1], Jiaxue Lin[1], Longshi Liao[1], Huixian Zhang[1], Boyin Huang[6], Phil Jones[7]

[1] *School of Atmospheric Sciences and Guangdong Province Key Laboratory for Climate Change and Natural Disasters, SUN Yat-Sen University, Zhuhai, Guangdong, China*

[2] *Research Center for Ecology and Environment of Central Asia, Chinese Academy of Sciences, Wulumuqi, Xinjiang, China.*

[3] *Guangzhou University, Guangzhou, China*

[4] *Tianjin Meteorological Information Center, Tianjin Meteorological Bureau, Tianjin, China*

[5] *College of Geographical Science, Inner Mongolia Normal University, Hohhot, China.*

[6] *National Centers for Environmental Information, National Oceanic and Atmospheric Administration, Asheville, North Carolina, USA*

[7] *Climatic Research Unit, School of Environmental Sciences, University of East Anglia, Norwich, Norfolk, UK*

 # Southern Laboratory of Ocean Science and Engineering (Guangdong Zhuhai), Zhuhai, China

Corresponding author: Qingxiang Li (liqingx5@mail.sysu.edu.cn)

Key Points

A traceable uncertainty framework developed for China-MST 3.0 enables the separate quantification of individual uncertainty components.

Coverage uncertainty is the main source of land surface air temperature uncertainty but has declined significantly since the 19th century.

Global mean surface temperature uncertainty has decreased to below 0.03°C in recent decades, showing high confidence in modern estimates.

Abstract

Global Mean surface temperature (GMST) is among the most important indicators of climate change, and its associated uncertainties affect the assessment of historical warming and the formulation of mitigation and adaptation policies. China-MST 3.0 is a newly updated GST dataset that merges China-LSAT 2.1 for land surface air temperature (LSAT) and ERSST v6 for sea surface temperature (SST). In this study, we develop a systematic and traceable uncertainty analysis framework for the construction process of this dataset. Specifically, we comprehensively evaluate three components of LSAT uncertainty: observation, analysis, and coverage uncertainty, while describing SST uncertainty in terms of both parameter and reconstruction uncertainty. We also provide a quantitative assessment of the spatial and temporal evolution of these uncertainties. The results show that LSAT uncertainty is generally larger than that of SST and is mainly driven by coverage uncertainty. The overall uncertainty in global mean surface temperature (GMST) shows a significant downward trend, with the annual 1σ uncertainty falling below 0.03℃ in recent decades, indicating high data reliability. However, uncertainty was high during the second-half of the 19th century and remains large at high latitudes in the Southern Hemisphere (SH). Comparative analyses indicate that China-MST 3.0 is broadly consistent with other datasets in both the magnitude and temporal evolution of GMST uncertainty. These findings demonstrate the utility of China-MST 3.0 as a valuable tool for evaluating global warming since the 1850s.

Plain Language Summary

China-MST 3.0 is a newly updated dataset that merges land surface air temperature from China-LSAT 2.1 with sea surface temperature data from ERSST v6. This study introduces a clear and step-by-step method to decompose and measure different sources of uncertainty in the data, such as station records homogeneity, reconstruction methods,

and missing coverage. We found that uncertainties mainly come from incomplete data, especially over land, but these uncertainties have greatly decreased over time as observational quality and quantity improved. By fully accounting for these uncertainties, this framework not only strengthens confidence in the dataset but also guides future improvements in observations, especially in poorly monitored regions, and enables more accurate and transparent quantification of dataset uncertainty.

1. Introduction

Global Surface Temperature (GST) is one of the key climate variables for evaluating climate change. It not only captures the aggregate response of the climate system to external forcings such as greenhouse gases and aerosols but also serves as a fundamental basis for tracking global warming since the pre-industrial era and informing climate policy decisions (Hawkins et al., 2017; IPCC, 2021). The Paris Agreement set a long-term temperature objective to limit global temperature rise to well below 2℃ above pre-industrial levels, with efforts to constrain warming to 1.5℃ in order to mitigate the most severe climate impacts. Accordingly, accurately quantifying historical temperature change is of critical scientific and policy relevance for understanding the current climate state and supporting informed mitigation and adaptation strategies.

However, early temperature observations suffered from uneven spatial coverage, incomplete records, and inconsistencies in measurement techniques, leading to considerable uncertainty in long-term GST series, particularly during the early industrial period (Brohan et al., 2006; Folland et al., 2001; Huang et al., 2020; Lenssen et al., 2019, 2024; Morice et al., 2012, 2021; Rohde et al., 2013; Rohde & Hausfather, 2020). Therefore, major GST datasets, such as HadCRUT5, Goddard Institute for Space Studies (GISS) Surface Temperature (GISTEMP) v4, NOAAGlobalTemp v5, and Berkeley Earth, have developed increasingly sophisticated methods to quantify uncertainties, using ensemble-based methods and statistical techniques (Huang et al.,

2020; Lenssen et al., 2024; Morice et al., 2021; Rohde & Hausfather, 2020).

HadCRUT5, developed by the UK Met Office Hadley Centre and the Climatic Research Unit (CRU) at the University of East Anglia, integrates the CRUTEM5 (land) and HadSST4 (ocean) data (Morice et al., 2021). It offers 200 ensemble members in two versions: (1) the HadCRUT5 non-infilled data set, using a gridding and ensemble generation approach consistent with HadCRUT4 (Morice et al., 2012), which includes multiple uncertainty components such as bias (large-scale systematic bias), partially correlated error (localized structural uncertainty), uncorrelated error, and coverage uncertainty from spatial sampling gaps; (2) the HadCRUT5 Analysis, a spatially infilled version based on Gaussian Process analysis, which improves field completeness. The latter incorporates analysis uncertainty, reflecting reconstruction and observational uncertainty, and a coverage uncertainty term associated with incomplete spatial averaging.

GISTEMP v4, produced by NASA's Goddard Institute for Space Studies, combines the Global Historical Climatology Network (GHCN) monthly version 4 (GHCNm v4) with Extended Reconstructed Sea Surface Temperature version 5 (ERSSTv5) (Lenssen et al., 2019). Its uncertainty framework comprises: (1) a 100-member GHCN-ERSST-GISTEMP ensemble representing uncertainties excluding LSAT sampling error; (2) two realizations of LSAT sampling uncertainty derived from the European Centre for Medium-Range Weather Forecasts (ECMWF) Reanalysis version 5 (ERA5) reanalysis, which are added to the 100-member ensemble to form a 200-member ensemble; (3) an additional 200-member full-coverage ensemble to assess uncertainty in global and regional means (Lenssen et al., 2024).

The latest version of NOAAGlobalTemp (v6), produced by NOAA, also uses GHCNv4 (land) and ERSSTv6 (ocean) and utilizes artificial intelligence techniques to reconstruct land and Arctic temperatures (Yin et al., 2024). Since v6 does not provide its uncertainty estimates, the uncertainty framework described for v5 is therefore

summarized here, following the methodology of Huang et al. (2020). Specifically, v5 generates 1000-member ensembles for LSAT and SST to assess parameter and reconstruction uncertainty. Parameter uncertainty reflects sensitivity to reconstruction settings, while reconstruction uncertainty captures residual errors from approximating the temperature field using a limited set of Empirical Orthogonal Teleconnection (EOT) modes (van den Dool et al., 2000).

Berkeley Earth provides a global gridded surface temperature dataset that merges its land product with HadSST4 (Rohde & Hausfather, 2020). Unlike the ensemble-based approaches of other datasets, it evaluates LSAT and SST uncertainties independently. LSAT uncertainty is divided into statistical uncertainty, reflecting errors in model parameter estimates due to imperfect raw observations, and spatial uncertainty, resulting from incomplete observational coverage (Rohde et al., 2013). SST uncertainty includes measurement and sampling uncertainty, bias uncertainty associated with evolving measurement practices, and coverage uncertainty due to gaps in ocean observations (Rohde & Hausfather, 2020).

The China global Merged Surface Temperature (China-MST/C-MST) dataset is a new GST dataset developed by the team at Sun Yat-sen University. The latest version, China-MST 3.0, integrates the LSAT dataset China-Land Surface Air Temperature 2.1 (China-LSAT 2.1 or C-LSAT 2.1) (Wei et al., 2025; Xu et al., 2025) with the Extended Reconstructed Sea Surface Temperature version 6 (ERSST v6) (Huang et al., 2025) as the ocean component. In contrast to datasets such as HadCRUT5, GISTEMP, and NOAAGlobalTemp v5, which derive GMST uncertainty from ensemble-based reconstructions, the GMST uncertainty in China-MST is estimated through the integration of uncertainty components from the LSAT and SST datasets, following an approach similar to that used by Berkeley Earth. For the ocean component, the uncertainty is derived from the combination of parameter and reconstruction uncertainties (Huang et al., 2020). For land component, while previous versions provided an initial uncertainty assessment (Sun et al., 2021, 2022), the quantification

Commented [PJ1]: Maybe here, Something like

In this context, our approach is complementary to the ensemble-based methods used elsewhere.

of uncertainties remained incomplete. For instance, while the reconstruction used in China-LSAT 2.0 effectively captured low-frequency components in certain regions, it failed to reproduce corresponding high-frequency variability during earlier periods (e.g., before 1900). This limitation is especially critical in regions with sparse observational coverage, where short-term fluctuations cannot be reliably given. Retaining such incomplete time series may therefore distort the representation of local interannual climate change. To mitigate this issue, China-LSAT 2.1 adopts a new observational constraint approach (Li & Li, 2024), which treats the grid boxes with poorly reconstructed series as missing values (Appendix D). While this approach improves data reliability, the resulting spatial gaps introduce additional coverage-related uncertainty, which should be considered in the LSAT uncertainties.

In this study, we present an integrated uncertainty framework for China-MST 3.0. Our emphasis remains on the land component (China-LSAT 2.1), in which uncertainties arising from observational limitations, reconstruction procedures, and spatial averaging processes (Section 3). The uncertainties of the land and ocean components are then combined through weighted averaging to estimate the GMST uncertainty, consistent with previous studies (Huang et al., 2020; Lenssen et al., 2019; Rohde & Hausfather, 2020). This framework enables a more robust and spatially consistent assessment of global and regional surface temperature uncertainties, thereby enhancing our understanding of historical warming trends. Thus, the remainder of this paper is structured as follows. Section 2 provides a concise overview of the development of the China-MST dataset. Section 3 presents a detailed description of the procedures for evaluating LSAT uncertainty. Sections 4 and 5 describe the methods for assessing SST and Global Mean Surface Temperature (GMST) uncertainty, respectively. Section 6 presents the main results, including global and regional analyses of LSAT uncertainty, estimates of GMST uncertainty, and comparisons with other datasets. A summary is given in Section 7.

2. Development of the China-MST dataset

The first version, China-MST 1.0, was constructed by merging the LSAT dataset China-LSAT 1.3 (also known as C-LSAT 1.3) with the SST dataset ERSST v5 (Yun et al., 2019). According to Xu et al. (2018), China-LSAT 1.3 integrated 14 data sources, including three global datasets (CRUTEM4, GHCN, and BEST), three regional datasets, and eight national-level collections, substantially improving station coverage. Homogenization was performed using the Penalized Maximal T-test (PMT) method (Wang, 2008a, 2008b). The ocean component, ERSST v5, was reconstructed using a combination of high- and low-frequency reconstruction methods and the EOT technique, with the start year of 1854 (Huang et al., 2017). The merging procedure involved: (1) regriding China-LSAT 1.3 (at 5°×5°) and ERSST v5 (at 2°×2°) to a common 1°×1° resolution; (2) applying a land-sea mask (available at http://www.ncl.ucar.edu/Applications/Data/cdf/landsea.nc), and (3) averaging the merged product to a 5° × 5° resolution (Yun et al., 2019). China-MST 1.0 provides monthly temperature anomalies on a global grid beginning in January 1900.

Subsequent versions, China-MST-Interim and China-MST 2.0 updated the land component using China-LSAT 2.0 (or C-LSAT 2.0), an enhanced version of the earlier China-LSAT 1.3 (Sun et al., 2021, 2022). Compared to China-LSAT 1.3, which covered the period from January 1900 to December 2017, China-LSAT 2.0 extended temporal coverage back to January 1850 and forward to December 2022. The number of stations also increased, particularly during the period 2013-2018, with additional observations sourced from CLIMAT reports and the Global Surface Summary of the Day (GSOD). Furthermore, temperature fields over land and in high-latitude regions were reconstructed using a combination of low- and high-frequency reconstruction methods and the EOT technique. The ocean component remained ERSST v5 (Huang et al., 2017), but its time coverage was extended back to 1850 by incorporating SST anomalies from ICOADS Release 3.0 (Freeman et al., 2017) for the 1850-1853 period. In addition, the

Arctic temperature in China-MST 2.0 was reconstructed based on the maximum and minimum sea ice extents, resulting in two dataset variants: Imax, in which the 65°N - 90°N region is infilled with air temperature data instead of SST; and Imin, where only the 80°N-90°N band is infilled (Sun et al., 2022). Uncertainty in China-MST-Interim and China-MST 2.0 was assessed using an ensemble reconstruction approach, with total uncertainty comprising both land and ocean components. Ocean uncertainty was derived from the ERSST v5 product, while land uncertainty was based on the ensemble of the reconstructed China-LSAT 2.0 dataset. The latter was further separated into two components: parameter uncertainty and reconstruction uncertainty.

China-MST 3.0, released in 2025, includes major updates to its data sources and introduces a newly designed uncertainty analysis framework. The land component has been replaced by China-LSAT 2.1 (Wei et al., 2025; Xu et al., 2025). Building on the foundation of China-LSAT 2.0, China-LSAT 2.1 substantially expanded station network after 2015, increasing the number of stations providing temperature observations to 25,085 for Tavg, 25,086 for Tmax and 25,083 for Tmin. The updated station network exhibits markedly improved spatial coverage, particularly in East Asia. Temporally, most stations provide records spanning 50-80 years, with a few extending to 80-100 years (Wei et al., 2025; Xu et al., 2025). The ocean component has been updated to ERSST v6 (Huang et al., 2025). ERSST v6 improved high-frequency SST reconstruction using Artificial Neural Network (ANN) interpolation, which enhanced consistency with observed SST, as indicated by an approximately 5% increase in spatial correlation and a reduction in Root Mean Square Error (RMSE) by 0.03℃ (Huang et al., 2025). The methodologies for quantifying uncertainty in LSAT, SST, and the resulting global mean are described in Sections 3, 4, and 5, respectively. As in previous releases, China-MST 3.0 provides both Imax and Imin variants, with Imax recommended for general use. The dataset is available at: http://www.gwpu.net/en/ (English access).

3. LSAT Uncertainty

The uncertainty in China-LSAT 2.1 originates from multiple stages in the dataset construction process, with distinct types of uncertainties introduced at each stage. The LSAT dataset construction can be broadly divided into three main steps: (1) station data processing, (2) grid data reconstruction, and (3) observation constraints. Specifically, during station data processing, observation uncertainty arises, including station uncertainty, sampling uncertainty, and bias uncertainty. During the grid data reconstruction, analysis uncertainty occurs, primarily comprising parameter uncertainty and reconstruction uncertainty. Finally, after the observation constraint step, coverage uncertainty emerges when estimating regional averages due to missing data. A schematic flowchart is presented in Figure 1. Because different types of uncertainty are derived from distinct and independent processes, they are treated as statistically independent. The total LSAT uncertainty is therefore obtained by combining their variances in quadrature (Appendix C). The detailed methodology for estimating LSAT uncertainty is described below.

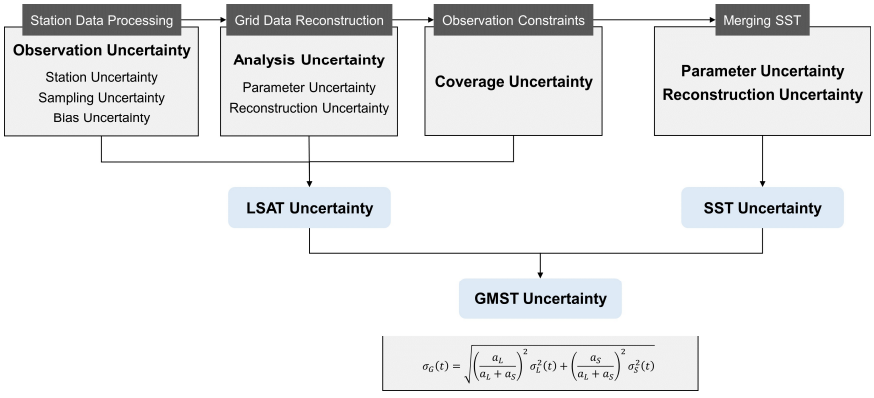


Figure 1. Schematic diagram of the China-MST 3.0 construction framework, showing the main processing stages and the corresponding sources of uncertainty introduced at each stage.

3.1. Observation Uncertainty

During the collection and processing of station data, observation uncertainty is categorized into three components: station uncertainty, sampling uncertainty, and bias uncertainty (Brohan et al., 2006; Li et al., 2010; Morice et al., 2012). Station uncertainty represents the uncertainty within individual station records, arising from measurement errors, uncertainties in homogenization adjustments, and errors due to the finite number of years used to calculate station normals. Sampling uncertainty occurs because the mean temperature anomaly of a grid box, even if based on error-free station data, may not equal the true spatial mean when estimated from a limited and uneven set of station anomalies. Bias uncertainty results from systematic, non-climatic influences, including changes in temperature measurement methods and the local warming effects of urbanization.

3.1.1. Station Uncertainty

The estimation of station uncertainty follows the framework proposed by Brohan et al. (2006), comprising three elements: measurement uncertainty, homogenization uncertainty, and normals uncertainty.

According to Folland et al. (2001), the random error associated with a single thermometer reading is approximately 0.2℃. Assuming 60 observations per month, the maximum measurement uncertainty for monthly mean temperature is estimated as $\frac{0.2}{\sqrt{60}} = 0.03$℃.

Homogenization uncertainty arises from residual errors associated with adjustments made during the homogenization process. China-LSAT 2.1 employs several procedures to detect and correct inhomogeneities (Xu et al., 2025) (see Text S1 in the Supporting Information). Based on the distribution of applied adjustments for all modified Tmax and Tmin series (Figure S1 in the Supporting Information), all

adjustment magnitudes applied to China-LSAT 2.1 are summarized in Figure 2 as the pink histogram. The distribution is approximately bimodal, with a greater number of negative adjustments than positive ones. As noted in previous studies, such bimodality is expected because large inhomogeneities are more readily detected and corrected, whereas small shifts are more likely to remain unresolved(Brohan et al., 2006; Menne et al., 2009, 2018; Menne & Williams, 2009). Therefore, following Brohan et al. (2006), the underlying population of required adjustments can be treated as approximately Gaussian as a first-order statistical representation. We therefore assume that the required adjustments follow a normal distribution N (0, 0.85) (dashed line). The difference between this ideal distribution and the observed adjustment distribution is indicated by blue histogram, which is further approximated by a Gaussian distribution to quantify its spread, with a standard deviation of 0.5 ℃ (solid line). Although the empirical adjustment distribution exhibits some asymmetry (Fig. 1; Fig. S1), this effect is small relative to the overall adjustment magnitude and is difficult to robustly quantify. Therefore, consistent with previous homogenization studies (Brohan et al., 2006), the residual uncertainty is treated as symmetric about zero. Hence, the homogenization uncertainty for a single station is estimated as 0.5℃, and for N stations, it scales as $\frac{0.5}{\sqrt{N}}$ ℃.

The selection of different climatological reference periods introduces normals uncertainty. The normals period 1961–1990 is adopted as the baseline in the China-LSAT 2.1 (see discussion of development of different normals periods in World Meteorological Organization (WMO) (2007, 2017). WMO (2017) recommend keeping 1961-1990 as the normals period for climate any climate change contexts. Each station is required to have at least 10 years of complete monthly data within this period. Stations are classified into two categories: (1) for stations with complete monthly records for the entire 1961–1990 period, the uncertainty is computed as $\frac{\sigma}{\sqrt{30}}$, where σ denotes the standard deviation of the 30-year monthly values; (2) for stations with N

years of available data (N > 10) within the baseline period, the uncertainty is computed as $\frac{\sigma}{\sqrt{N}}$. In contrast to the method used by Brohan et al. (2006), we do not substitute a station's climatology with the World Meteorological Organization (WMO) normals, even when the amount of available data during the baseline period is limited.

Since the three components of station uncertainty: measurement, homogenization, and normals uncertainty, are assumed to be statistically independent, the total station uncertainty is obtained by combining their variances in quadrature (see Appendix A). Figure 3a presents the spatial distribution of total station uncertainty for January 2020. Higher uncertainty is observed in high-latitude regions of the Northern Hemisphere (NH).

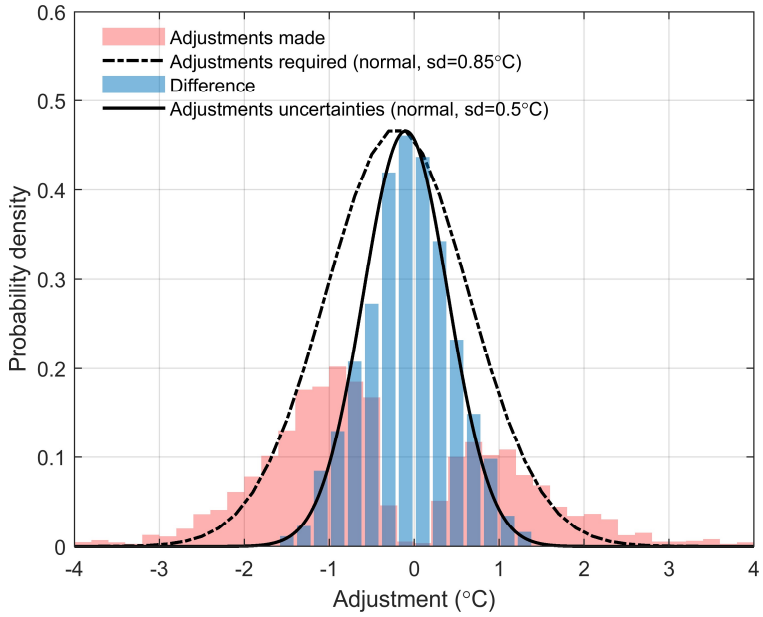


Figure 2. Distribution of homogenization adjustments applied to individual stations. The pink histogram represents the distribution of all detected breakpoints; the dashed line indicates the assumed distribution of required adjustments; the blue histogram shows the difference between the two, and the solid line represents the distribution of homogenization uncertainty.

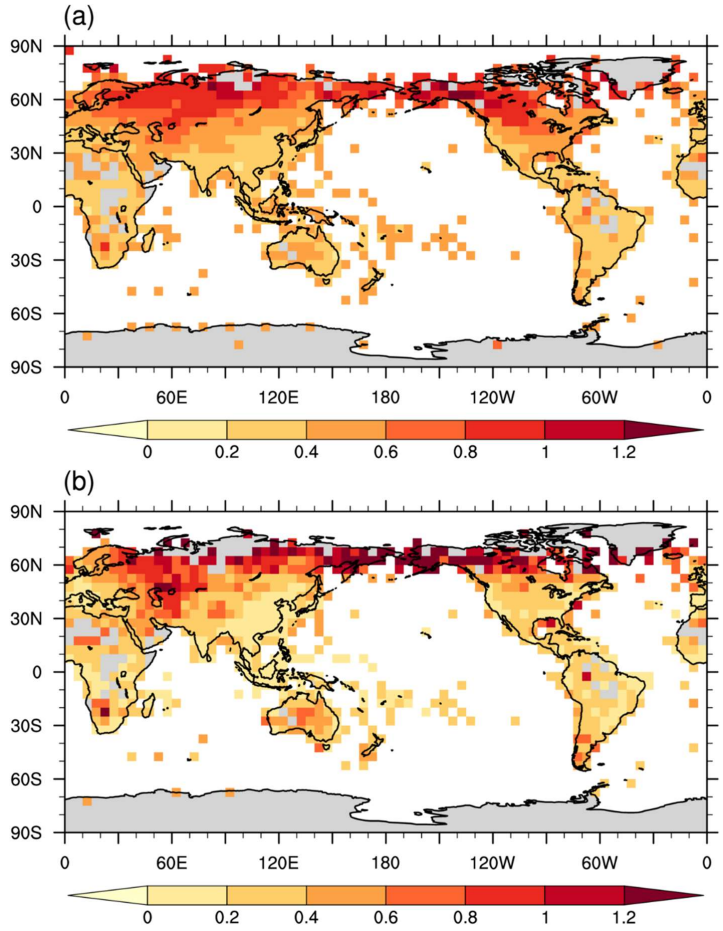


Figure 3. Spatial distribution of (a) station uncertainty and (b) sampling uncertainty (1σ) for January 2020.

3.1.2. Sampling uncertainty

When interpolating station data onto a regular grid, we adopt the Climate Anomaly Method (CAM) (Jones, 1994), in which the temperature anomaly of each grid box is calculated as the average of all station anomalies within the corresponding 5°×5° box. However, the average of station anomalies may not accurately represent the true spatial mean of temperature anomalies within the grid box, thereby introducing sampling

uncertainty. This uncertainty depends on the number of stations, their spatial distribution, and the underlying climate variability within the grid cell. Sampling uncertainty is calculated following the method proposed by Jones et al. (1997) (see Appendix A).

Figure 3b shows the spatial distribution of sampling uncertainty for January 2020. Similar to station uncertainty, sampling uncertainty tends to be higher in high-latitude regions of the NH.

3.1.3. Bias Uncertainty

Bias uncertainty primarily arises from two large-scale systematic sources: the urbanization effects and thermometer exposure bias.

Jones et al. (1990) estimated that urbanization contributed approximately 0.05℃ to LSAT increases between 1900 and 1990. Based on this, Folland et al. (2001) proposed an uncertainty model wherein the 2σ uncertainty associated with urbanization bias increases from 0℃ in 1900 to 0.12℃ by 2000. In this study, we adopt a similar framework to assess the influence of urbanization on the temperature within the China-LSAT2.1 dataset. Stations are first classified as urban or rural, and the contribution of urbanization is then estimated using the All-minus-Rural (AMR) method (see Text S2 in the Supporting Information). The result is shown in Figure 4, where the solid red line represents the temperature difference between the All and Rural GMST series (relative to 1961-1990), and the dashed black line indicates the least-squares linear trend, with the regression coefficient displayed in the upper-right corner. The inferred urbanization-induced warming trend since 1900 is approximately 0.0052℃/10yr. This contribution is an order of magnitude smaller than the overall climate warming trend, indicating that the urbanization effect on large-scale temperature averages is secondary. Following the treatment of Brohan et al. (2006) and subsequent HadCRUT uncertainty assessments (Morice et al., 2012; 2020), urbanization is physically regarded as a potential one-sided

warm bias; however, its magnitude is represented statistically using a standard-deviation-based uncertainty parameter. Accordingly, based on the results of China-LSAT 2.1, we update the uncertainty model for urbanization bias, assigning a 1σ uncertainty that increases linearly from zero starting in 1900 at a rate of 0.0052℃ per decade at the grid-box level. Thermometer exposure bias arises from historical changes in the shielding and sheltering of thermometers at regional- to global-scales, which affect their exposure to sunlight, wind, and precipitation. Such changes are difficult to correct using traditional homogenization methods, as exposure biases tend to affect many stations systematically rather than individually (Wallis et al., 2024). Parker (1994) estimated that exposure changes could result in a systematic difference of approximately 0.2℃ between early 20th-century and modern records. Subsequently, Folland et al. (2001) and Brohan et al., (2006) proposed an uncertainty model that defines grid-box level exposure bias uncertainty: for stations located in the 20°S–20°N region, a 1σ uncertainty of 0.2℃ is assumed before 1930, decreasing linearly to zero by 1950; for stations outside this latitude range, the 1σ uncertainty is 0.1℃ before 1900, also tapering to zero by 1930. This empirical model is also adopted in our assessment of exposure bias uncertainty. Recently, Wallis et al. (2024) proposed a new regression-based method incorporating exposure metadata to estimate seasonal bias, arguing that previous models did not consider seasonal variations. However, as Morice et al. (2024) noted, these adjustments fail to cover many stations in the tropics and high latitudes and do not fully account for early-period metadata. Consequently, we ignore the seasonal cycle of exposure bias and retain the original uncertainty model.

Figure 5 summarizes the temporal evolution of the three types of observation uncertainty: station uncertainty (red), sampling uncertainty (blue), and bias uncertainty (green) (calculation method is described in Appendix B). Station uncertainty consistently dominates the total uncertainty over the entire period. It is relatively high in the mid-19th century (about 0.07℃), then gradually declines with improvements in observational network density and data quality, stabilizing by the mid-20th century.

Sampling uncertainty remains low throughout, exhibiting a slight decreasing trend. Bias uncertainty contributes the least and remains nearly stable over time, with only a modest reduction during the mid-20th century, primarily attributable to thermometer exposure bias. In summary, limited station quantity and quality largely drove uncertainty in the 19th and early 20th centuries, while differences among uncertainty components have diminished notably in recent decades.

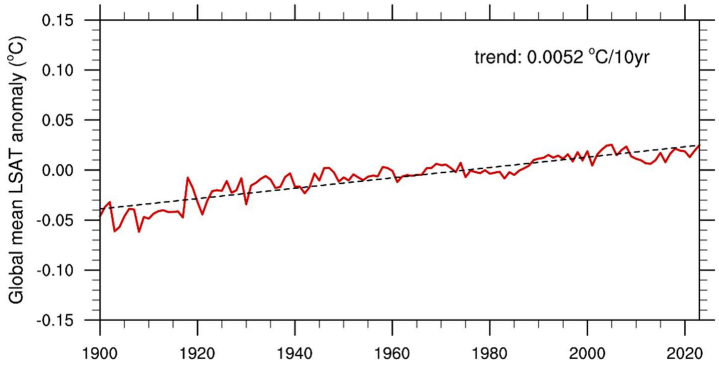


Figure 4. Contribution of urbanization to global temperature. The solid red line shows the time series (relative to 1961–1990), and the dashed black line indicates the least-squares linear trend

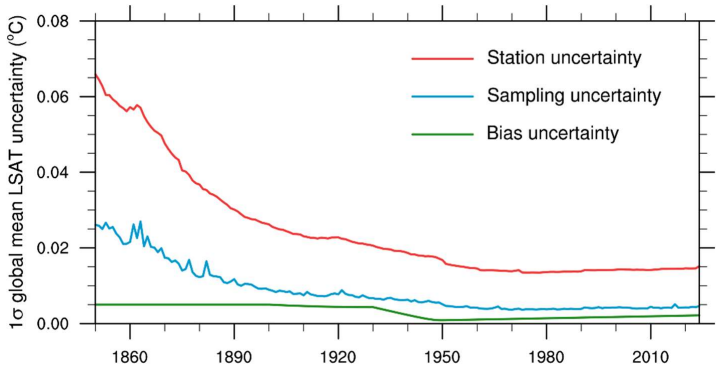


Figure 5. Temporal Evolution of Observational Uncertainty from 1850 to 2024: Station Uncertainty (red), Sampling Uncertainty (blue), and Bias Uncertainty (green).

3.2. Analysis Uncertainty

Due to substantial data gaps in early observations, the estimation of global mean temperature is subject to considerable uncertainty. To address this issue, we reconstruct the LSAT by decomposing it into high- and low-frequency components and applying the EOT method to infill missing land grid cells (Sun et al., 2021, 2022). Given the pronounced polar amplification in the Arctic, temperature changes in this region have an increasing influence on global averages (Li et al., 2023). Considering the comparable physical characteristics between ice-covered surfaces and land, regions north of 65°N were treated as land surfaces in the reconstruction process (Li et al., 2023; Sun et al., 2022).

To quantify associated uncertainties, we employed a multi-parameter ensemble approach (Table S1), accounting for both parameter uncertainty and reconstruction uncertainty, collectively referred to as analysis uncertainty. This approach is conceptually similar to that of Sun et al. (2021) but with a simplified parameterization scheme. Specifically, the parameter "minimum number of months for annual average", is excluded here, as it is irrelevant to LSAT: an annual mean is computed only when all twelve months of data are available. As a result, varying this parameter yields no actual difference in outcome, thereby reducing the number of distinct, effective parameter combinations. Consequently, the ensemble in this study includes 252 unique configurations, fewer than those used in previous studies.

3.2.1. Parameter Uncertainty

Parameter uncertainty represents the variability introduced by alternative methodological choices in the reconstruction (e.g., the length of the low-frequency (LF) filter or the training parameters of the EOTs; see Table S1). We test 252 different parameter combinations to generate an ensemble of temperature datasets. Parameter uncertainty thus reflects the variability among these ensemble members. It is quantified

401 at each grid point following Huang et al. (2020) (see Appendices A and B).

402 ### 3.2.2. Reconstruction Uncertainty

403 The EOT method can smooth local temperature variations, potentially leading to the
404 loss of fine-scale information and introducing bias from observations. Reconstruction
405 uncertainty reflects the intrinsic uncertainty of the EOT approach. Following Huang et
406 al. (2020), we quantify it using pseudo-observations with full spatial coverage from the
407 CMIP6 BCC-CSM2-MR model, using historical simulations for 1850-2014 and the
408 SSP585 scenario thereafter, to ensure consistency in the temporal coverage between
409 model and observation data. The same reconstruction procedures and parameter settings
410 are applied to generate an ensemble, and reconstruction uncertainty is calculated as the
411 difference between the original and reconstructed fields (see Appendices A and B).

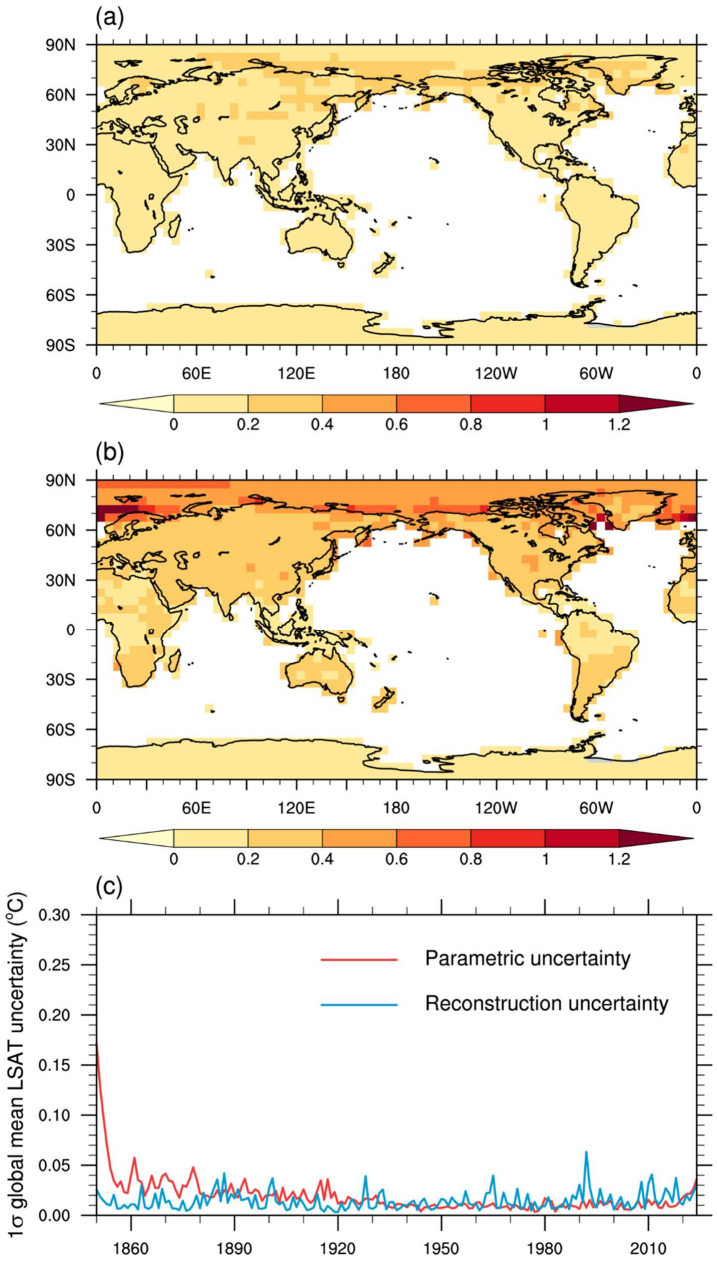
412 Figures 6a and 6b show the spatial distributions of parameter and reconstruction
413 uncertainty, respectively, averaged over the period 1850-2024. Parameter uncertainty is
414 generally low across most land regions, typically below 0.4℃. However, it is notably
415 higher over the high-latitude NH, where sparse observations and greater climate
416 variability contribute to larger parameter uncertainty. In contrast, reconstruction
417 uncertainty exhibits more spatial heterogeneity, with particularly high values in the
418 Arctic.

419 Figure 6c illustrates the temporal evolution of two types of analysis uncertainties.
420 Parameter uncertainty (red line) shows a marked decline over time, driven not only by
421 the progressive increase in observational coverage but also significantly influenced by
422 boundary effects caused by low-frequency filtering. The choice of low-frequency
423 filtering parameters introduces boundary effects, manifested as insufficient filter
424 response at the beginnings and ends of the time series, which leads to biased signal
425 estimation and larger fluctuations in the filtered signal at these edges. Because the high-
426 frequency data used for reconstruction are obtained by subtracting the low-frequency

component from the original data, differences in the low-frequency estimates amplify discrepancies in the high-frequency data, resulting in greater variability in the reconstructed fields. This effect is especially pronounced during the sparse observational period starting around 1850, where parameter uncertainty peaks at approximately 0.17℃ and then declines nearly linearly over about five years to around 0.05℃ by 1854. A similar increase in parameter uncertainty is observed after 2010; however, due to the larger number of effective grid points and higher data completeness in this period, the fluctuations in uncertainty are comparatively smaller. As observational coverage improves and the time series lengthens, the influence of boundary effects diminishes, contributing to the rapid decline in parameter uncertainty during the mid-to-late 19th century. After 1880, with reduced boundary effects and increased data density, parameter uncertainty continues to decrease at a slower rate, stabilizing around 0.01-0.02℃ by approximately 1930. In contrast, reconstruction uncertainty (blue line) remains relatively stable over time, hovering around 0.02℃, as it is primarily determined by the structural limitations of the reconstruction method rather than by changes in data coverage. Before 1880, parameter uncertainty is the dominant source of total uncertainty. Between 1880 and 1950, parameter and reconstruction uncertainties are of comparable magnitude and jointly determine the overall uncertainty. After 1950, parameter uncertainty declines below reconstruction uncertainty, which subsequently becomes the primary source of both the magnitude and temporal variability of total uncertainty.

It is also worth noting that Huang et al. (2020) adopted a fundamentally different uncertainty framework by generating 100 ensemble members based on the GHCNm v4 dataset (Menne et al., 2018), which incorporates a wide range of uncertainties, including homogenization uncertainty, station uncertainty, exposure bias uncertainty, and sampling uncertainty. In this framework, observation uncertainty is directly propagated into the spatial reconstruction process, resulting in large uncertainty over regions with sparse observations, such as Antarctica. In contrast, our study produces a

single operational reconstruction rather than an ensemble. We apply a stepwise uncertainty decomposition framework in which observational uncertainty is estimated independently during the station data processing stage and is not propagated into the analysis uncertainty. The analysis uncertainty presented here solely reflects the sensitivity to parameter selection and the structural limitations of the EOT-based reconstruction method, without including the propagation of observational errors. This methodological distinction largely explains the smaller local uncertainties (e.g., over Antarctica ; Figure 6 a and b) and the lower global mean parameter and reconstruction uncertainty (Figure 6c) compared to Huang et al. (2020). In addition, the incorporation of more numerous and higher-quality station observations in these regions may have further improved the training dataset and contributed to the reduced uncertainties.



467

468 Figure 6. Spatial distributions of (a) local parameter uncertainty (1σ) and (b) local reconstruction
469 uncertainty (1σ) of LSAT averaged over the period 1850–2024; (c) Time series of global mean LSAT
470 parameter uncertainty (red) and reconstruction uncertainty (blue) from 1850 to 2024.

3.3. Coverage Uncertainty

When reconstructing LSAT fields using the EOT method, the quality of the results largely depends on the availability of relatively sufficient and well-distributed observations. The method may struggle in observation-sparse regions. which can further affect local temperature estimates. To address this issue, we apply the variance-based observational constraint method (see Appendix D). A sliding F-test is employed to ensure the significance of the retained series, filtering out statistically unreliable components. Physically, the high-frequency component of temperature should exhibit sufficient variance to represent realistic interannual fluctuations; grid points showing unrealistically low variance, which are often caused by sparse observations, are thus masked as missing. The procedure enhances the reliability and robustness of the reconstructed LSAT fields.

However, this procedure also leads to coverage uncertainty. This term refers to the uncertainty in regional or global mean estimates that arises from incomplete spatial and temporal coverage of observations. Following the approach of Morice et al. (2021), we estimate this coverage uncertainty using sub-sampling experiments based on a complete reanalysis dataset. Specifically, we employ European Centre for Medium-Range Weather Forecasts (ECMWF) Reanalysis version 5 (Hersbach et al., 2020) from 1961 to 1990 as a spatially complete reference field. For each same calendar month, the reanalysis field is resampled according to the spatial mask of available observations, producing a masked field with the same coverage as the constrained dataset. Regional or global mean temperature anomalies are then calculated for both the full and masked fields. The difference between these two averages represents the error associated with incomplete coverage. The Root Mean Square (RMS) of these differences across the 30 years is taken as the estimate of the monthly coverage uncertainty (see Appendix B). For annual uncertainties, both datasets are first averaged to annual resolution before applying the same sub-sampling procedure. Figure 7 shows the temporal evolution in

coverage uncertainty (red line) and land coverage rate (the ratio of the number of land grid boxes with valid observations to the total number of land grid boxes, blue line) from 1850 to 2024. After applying the variance-based observation constraints, the land coverage (including regions north of 65°N) is below 50% before 1880s, with corresponding coverage uncertainty reaching up to 0.24℃. As the observational network gradually improved, both coverage and accuracy increased. After the 1950s, the coverage rate stabilized above 80%, and coverage uncertainty declined to around 0.05℃. During this period, most of the residual coverage uncertainty was attributable to the Antarctic region, where observations remain sparse and reconstruction remains uncertain.

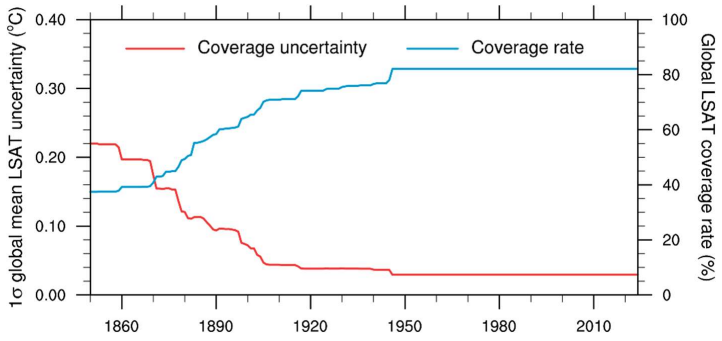


Figure 7. Temporal evolution of LSAT coverage uncertainty (red line) and land coverage rate (blue line) from 1850 to 2024.

4. SST Uncertainty

For the ocean component, we adopt NOAA's ERSST v6 dataset, which provides globally gridded monthly SST at a 2° × 2° resolution from 1850 to 2024. ERSST v6 integrates in situ observations from ships, drifting buoys, moored buoys, and Argo floats (Huang et al., 2016, 2017, 2025), drawing upon ICOADS 3.0.0, ICOADS 3.0.2, and GDAC archives (Freeman et al., 2017; Liu et al., 2022). The latest version applies an ANN interpolation method to improve SST reconstruction (Huang et al., 2025). However, since ERSST v6 does not yet provide a complete uncertainty estimate, we

still rely on the parameter and reconstruction uncertainty diagnostics developed for ERSST v5 to represent SST uncertainty. The evaluation method is similar to that described in Section 3.2, but differs in the parameter settings and the choice of model data used as pseudo-observations (Huang et al., 2020). The ERSSTv5 uncertainty is publicly available at https://www.ncei.noaa.gov/pub/data/cmb/ersst/v5/tmp/unc.ersst.v5.1850-2016/ and https://www.ncei.noaa.gov/pub/data/cmb/ersst/v5/tmp/unc.ersst.v5.2001-present/.

5. GMST Uncertainty

To estimate the overall uncertainty of the China-MST 3.0 dataset, we separately calculated the uncertainties of LSAT and SST. As these two components are derived from distinct observational systems and reconstruction methods, their uncertainties are treated as statistically independent. Accordingly, the total uncertainty of GMST is computed as a weighted combination of the two components, following the method proposed by Ku et al. (1966):

$$\sigma_G(t) = \sqrt{\left(\frac{a_L}{a_L + a_S}\right)^2 \sigma_L^2(t) + \left(\frac{a_S}{a_L + a_S}\right)^2 \sigma_S^2(t)}$$

where σ_G, σ_L, and σ_S represent the uncertainties of GMST, LSAT, and SST, respectively; and $a_L = 0.66$ and $a_S = 0.34$ are the fractional area weights for land and ocean (note that, in China-MST 3.0, the region north of 65°N is treated as land).

6. Result

6.1. Evolution Characteristics of Global Mean LSAT Uncertainty

Figure 8 presents the monthly and annual global mean LSAT anomaly, along with the corresponding time series of total uncertainty. This total uncertainty consists of three components: observation uncertainty (including station, sampling, and bias uncertainty),

analysis uncertainty (including parameter and reconstruction uncertainty), and coverage uncertainty. Coverage uncertainty (blue line) peaks in the 1850s, with a 1σ uncertainty exceeding 0.22℃. It then rapidly declines as observational data coverage improves and stabilizes around 0.05℃ by the mid-20th century (Figure 8d). This trend highlights the significant increase in the number of reliable grid points available for temperature reconstruction. Observation uncertainty (green line) and analysis uncertainty (red line) remain relatively low overall. Except for the 1850s when analysis uncertainty exceeds 0.1℃, both remain below 0.07℃ during other periods (Figure 8d), contributing minimally to the total uncertainty. Compared to coverage uncertainty, the temporal variations of observation and analysis uncertainties are more stable, although analysis uncertainty displays more fluctuation, with noticeable interannual variability in the uncertainty of annual mean temperatures. Overall, coverage uncertainty remains the dominant contributor to the global LSAT uncertainty time series, especially during the early stages of the record.

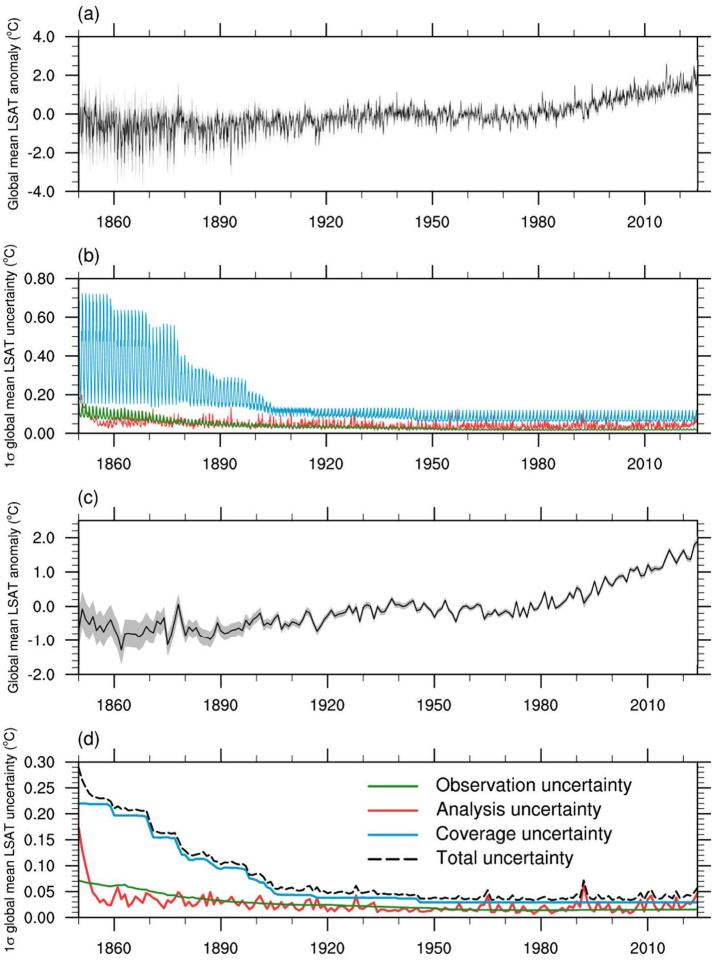


Figure 8. Global mean LSAT temperature series and associated uncertainties. (a, b) show the monthly time series and uncertainty components in monthly means, while (c, d) show the annual means. The black line represents the best estimate, and the gray shaded areas indicate the 95% confidence intervals in (a) and (c). The green, red, and blue lines represent observation, analysis, and coverage uncertainties, respectively, in (b) and (d). The black dotted line in (d) denotes the total uncertainty, obtained by combining the variances of the three annual uncertainty components in quadrature.

6.2. Evolution Characteristics of LSAT Uncertainty by Hemisphere and Latitude Bands

6.2.1. Hemisphere Means

Figures 9 and 10 present the hemispheric mean LSAT anomaly time series and their associated uncertainty.

In the NH, the overall uncertainty trend closely follows the global pattern but at a consistently lower level (Figures 9b, d), reflecting more extensive and reliable observational coverage. Coverage uncertainty decreases rapidly from the mid-19th century to the early 20th century and continues to decline in the latter half of the 20th century, falling below 0.01℃ in recent decades (Figure 9d). Both observation and analysis uncertainty remain at low levels. The uncertainty in the annual mean series further reduces, indicating a high level of confidence in NH temperature trends (Figure 9c).

In the Southern Hemisphere (SH), coverage uncertainty is higher than in the NH (Figures 10b, d), especially from the 19th to early 20th centuries, due to sparse station distribution and insufficient spatial coverage in early records. Although some improvement is evident after the late 20th century, coverage uncertainty remains the dominant contributor to LSAT uncertainty. Even in recent decades, annual mean uncertainty stays above 0.10℃ (Figure 10d), mainly due to the continued scarcity of observations in the Antarctic region. Observation and analysis uncertainties overall remain generally low and stable. However, unlike the global and NH results, observation uncertainty in the SH surpasses analysis uncertainty during the early period (19th century), mainly driven by station uncertainty reaching up to 0.2℃ (not shown), underscoring the poor quantity of early measurements. Note that in the earliest years of the SH annual mean temperatures and coverage uncertainty are occasionally unavailable due to incomplete monthly observations, i.e., fewer than 12 months per year.

In general, LSAT uncertainty in the SH remains substantially higher, highlighting the need for further methodological refinement in China-MST 3.0 to improve reconstruction quality and spatial coverage, particularly over Antarctica and its surrounding regions.

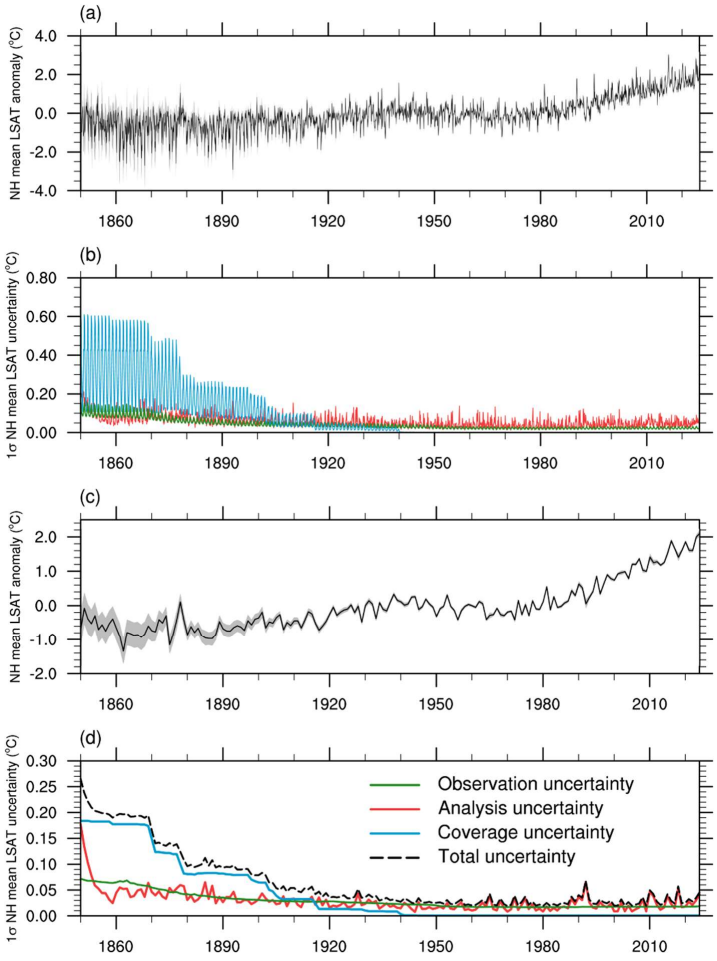


Figure 9. Same as Figure 8, but for NH mean LSAT.

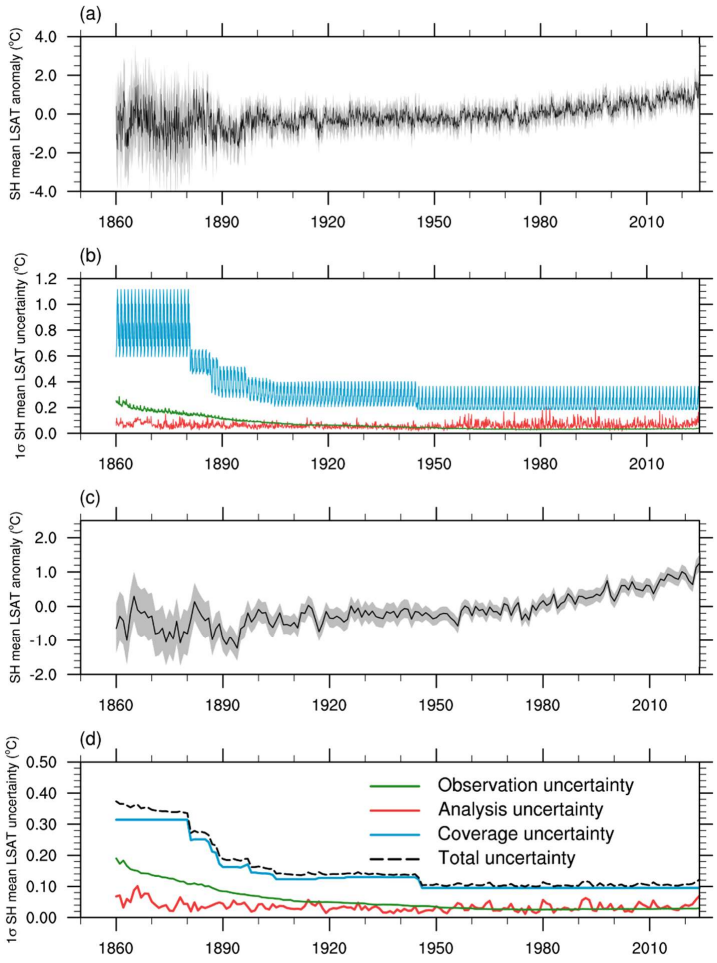


Figure 10. Same as Figure 8, but for SH mean LSAT.

6.2.2. Continental Means

Figure 11 presents the annual mean LSAT time series, associated uncertainty (represented by 95% confidence intervals), and spatial coverage rate for seven continental regions: Asia, Europe, Africa, Oceania, North America, South America, and Antarctica. Overall, LSAT trends and uncertainties exhibit pronounced regional contrasts across continents.

Asia and Europe show relatively early and dense station coverage (Figures 11a, b).

Commented [PJ2]: In Fig 11 you have 7 continents. So the Figure has one spare space on the bottom right. I realize that the Arctic isn't a continent, but putting this in the spare space might be an idea.
It would be next to the Antarctic. Arctic could be all north 65N. Again this is just a thought!

Commented [PJ2R2]: We did sort of have this with the 90-60N plot before. So you could just use that data, but the Arctic would be better from 65-90.
The pair of plots within Figure 11 highlights how much more certain the trends are in the Arctic than the Antarctic.

Both regions exhibit clear long-term warming trends. In Asia, the coverage rate increased rapidly from the late 19th century and has remained near complete since the mid-20th century, leading to progressively narrower uncertainty ranges. Europe demonstrates similarly high coverage throughout most of the record, resulting in consistently small uncertainties and a highly robust warming signal.

Africa and South America display moderate but gradually improving observational coverage (Figures 11c, f). In Africa, early coverage was sparse, leading to relatively wide confidence intervals prior to the mid-20th century. However, the steady expansion of station networks has reduced uncertainties in recent decades. South America shows a similar evolution, with uncertainty decreasing as coverage improves, although variability remains somewhat larger than in the Northern Hemisphere continents.

Oceania exhibits substantial early-period uncertainty due to limited station availability (Figure 11d). Although coverage improves after the early 20th century, it remains lower than in Asia and Europe, and the uncertainty envelope correspondingly remains wider. Nevertheless, a clear warming tendency is evident in the later record.

North America benefits from dense and well-distributed observations throughout most of the study period (Figure 11e). The coverage rate exceeds 80% relatively early and approaches completeness after the early 20th century. As a result, uncertainty remains comparatively small, and the long-term warming trend is highly credible.

Antarctica stands in sharp contrast to the other regions (Figure 11g). Reliable LSAT data only become available after the mid-20th century, and the spatial coverage remains persistently low. Consequently, the temperature series exhibits large uncertainty ranges, highlighting the continued difficulty of robust large-scale temperature assessment in this region.

In summary, coverage uncertainty remains the dominant factor controlling the magnitude of continental LSAT uncertainties. Continents with dense and early established observing networks (e.g., Europe, North America, and Asia) show relatively

small and stable uncertainties, whereas regions with historically sparse observations, exhibit substantially larger uncertainty envelopes. These results underscore the importance of station distribution in determining the reliability of regional temperature estimates.

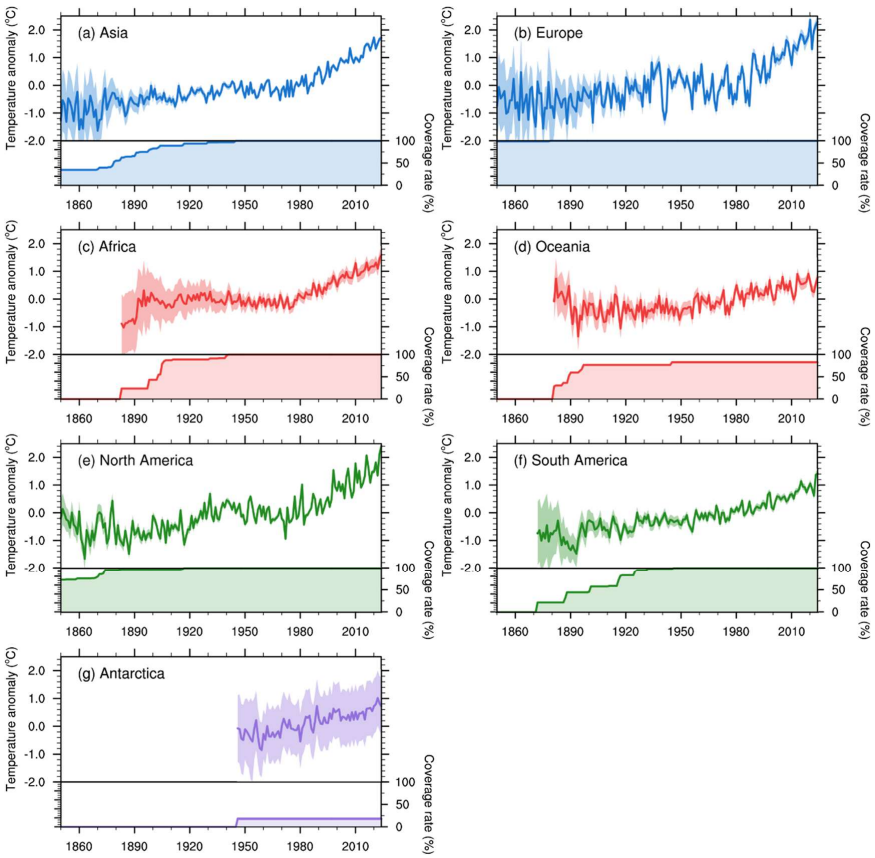


Figure 11. Annual mean LSAT time series and associated uncertainties (top panels) and coverage rate (bottom panels) for seven continents: (a) Asia, (b) Europe, (c) Africa, (d) Oceania, (e) North America, (f) South America, and (g) Antarctica. Shaded areas in the top panels represent the 95% confidence intervals of the uncertainties.

6.3. Evolution of GMST Uncertainty

Figure 12 presents the Global mean LSAT (GLSAT, red line), global mean SST (GSST,

643 blue line), and GMST (black line) series, along with their respective uncertainties. The
644 GMST uncertainty is calculated as a combined uncertainty from GLSAT and GSST.

645 The monthly mean uncertainties (Figure 12b) indicate that from the mid-19th
646 century to the early 20th century, GLSAT uncertainty was relatively high (greater than
647 0.2℃), reflecting the limited spatial representativeness. During the same period, GSST
648 uncertainty was also elevated but generally lower, staying below 0.1℃. After the mid-
649 20th century, uncertainties for both components decreased significantly and became
650 more stable, with GLSAT uncertainty remaining slightly higher than SST. Notably,
651 when applying a 12-month running mean, consistent with the filtering used in Huang
652 et al. (2020), the GLSAT monthly uncertainty in 1880 reduces to approximately 0.3℃,
653 which is broadly consistent with the combined LSAT (0.18℃) and coverage uncertainty
654 (0.12℃) reported in their study for the same year. While the direct comparison is made
655 for 1880, the subsequent temporal evolution shows broadly consistent patterns between
656 the two datasets, indicating that the overall discrepancy is minor.

657 The annual mean uncertainties (Figure 12d) exhibit a similar temporal pattern.
658 GLSAT uncertainty peaked at ~0.3℃ in the mid-19th century and gradually declined
659 to ~0.1℃ by the early 20st century. GSST uncertainty remained below 0.1℃ in early
660 decades and dropped below 0.06℃ in recent years, reflecting improved spatial
661 coverage and data quality.

662 GMST uncertainty is jointly determined by GLSAT and GSST uncertainties,
663 weighted by their respective area fractions. Since the ocean covers approximately twice
664 the area of land, the overall GMST uncertainty tends to follow the trend of GSST more
665 closely. As shown in Figure 12d, despite the relatively large uncertainties in GLSAT
666 during early periods, the temporal evolution of GMST uncertainty largely tracks that of
667 GSST. However, in the early period (1850–1860), elevated GLSAT uncertainty had a
668 notable influence, contributing to GMST uncertainty levels of approximately 0.08℃ to
669 0.12℃. Thereafter, GMST uncertainty decreased gradually. Notable peaks in GSST

uncertainty are observed in the late 1910s and early 1940s, corresponding to the two World Wars, during which SST observations were spatially limited, leading to increased uncertainty (Huang et al., 2020). Since the 1950s, with the expansion of the global observation network and advancements in data processing techniques, GMST uncertainty has declined to below 0.03°C, indicating a high level of reliability in current global temperature estimates.

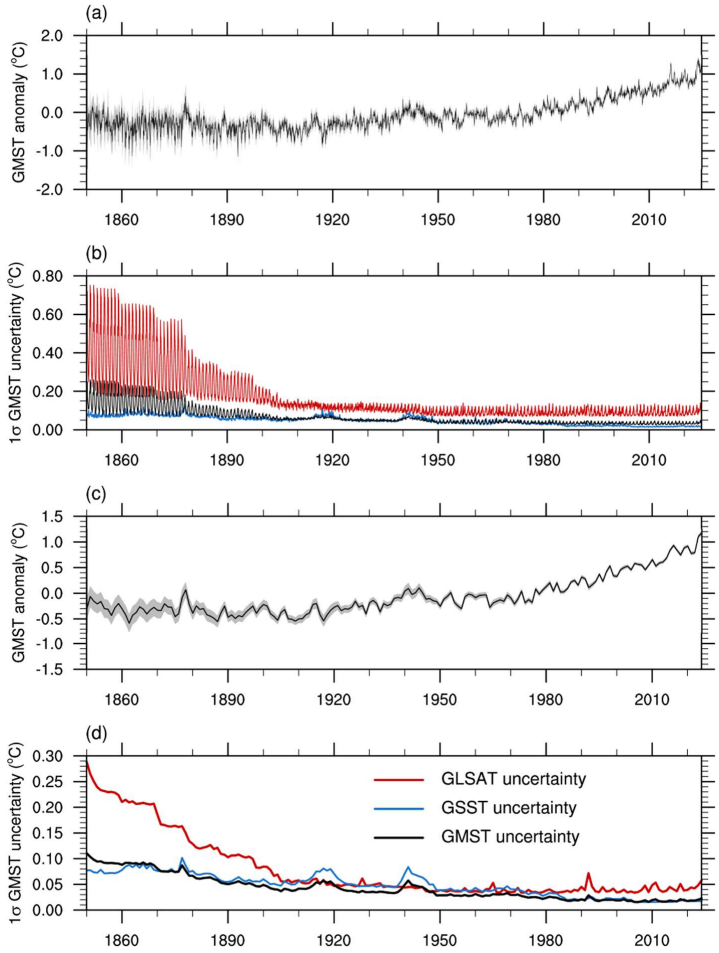


Figure 12. GMST temperature series and associated uncertainties. (a, b) show the monthly time series and uncertainty components in monthly means, while (c, d) show the annual means. The black line represents the best estimate, and the gray shaded areas indicate the 95% confidence intervals in

(a) and (c). The red, blue, and black lines represent GLSAT, GSST, and GMST uncertainties, respectively, in (b) and (d).

6.4. Comparison of GMST Uncertainty with Other Datasets

Figure 13 compares the GMST uncertainty from China-MST 3.0 (red line) with those from four GST datasets: HadCRUT5 (green line, data available at https://www.metoffice.gov.uk/hadobs/hadcrut5/) (Morice et al., 2021), Berkeley Earth (black line, data available at https://berkeley-earth-temperature.s3.us-west-1.amazonaws.com/Global/Land_and_Ocean_summary.txt) (Rohde & Hausfather, 2020), GISTEMPv4 (blue line, data available at https://data.giss.nasa.gov/gistemp/graphs_v4/) (Lenssen et al., 2024), and a recent developed Dynamically Consistent ENsemble of Temperature (DCENT) dataset (orange line, data available at https://www.dropbox.com/scl/fi/wcwwslemqigcohmyjj2sb/DCENT_GMST_statistics.txt?rlkey=wtfqtznicbix5b555xq5ehhky&e=1&dl=0) (Chan et al., 2024). These datasets adopt different approaches to estimating uncertainty: HadCRUT5, GISTEMPv4, and DCENT employ ensemble-based methods to jointly assess land and ocean uncertainties, while Berkeley Earth and China-MST 3.0 compute uncertainties for land and ocean components separately.

In general, the magnitude and temporal evolution of GMST uncertainty in China-MST 3.0 are broadly consistent with those in the other datasets, and the GMST estimates from different datasets show good agreement (Li et al., 2021), reflecting increasing growing robustness in global temperature assessments. A distinct spike in uncertainty appears in the late 1940s for both HadCRUT5 and Berkeley Earth, due to their use of the Hadley Centre's HadSST product for the ocean component. In contrast, China-MST 3.0 and GISTEMPv4, which adopt NOAA's ERSST dataset for SST, do not exhibit this feature.

Furthermore, China-MST 3.0 shows slightly higher uncertainty prior to the 1880s, similar to other major datasets. This is likely attributable to limited data coverage and greater uncertainty in early LSAT records (Figure 12d). As demonstrated by Morice et al. (2024), the most effective way to reduce such uncertainty is to recover and incorporate additional early observations, since methodological improvements alone cannot compensate for data sparsity.

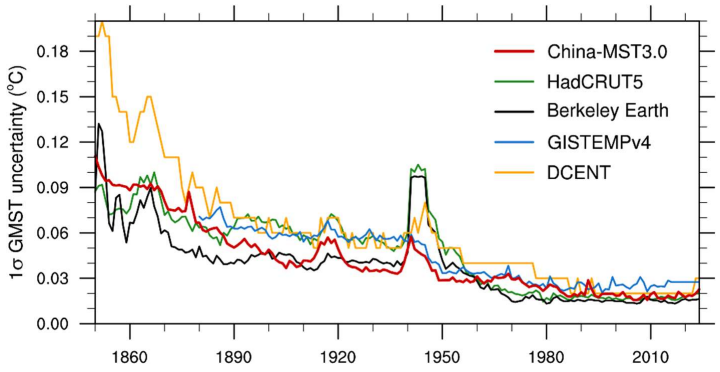


Figure 13. Comparison of GMST uncertainty (1σ) from China-MST 3.0 (red line) and other GST datasets, including HadCRUT5 (green line), Berkeley Earth (black line), GISTEMPv4 (blue line), and DCENT (orange line).

7. Conclusion

This study established an uncertainty assessment framework for the China-MST 3.0 dataset by separately quantifying uncertainties in LSAT (based on China-LSAT 2.1) and SST (from ERSST v5), following its construction procedures. Based on the several well-established approaches to assess different type of uncertainty, China-MST 3.0 incorporates several refinements. Specifically, the homogenization uncertainty was re-evaluated using an updated and expanded station dataset (China-LSAT 2.1); the urbanization effect was reassessed using an enhanced multi-source classification and AMR estimation scheme; and a new observation constraint procedure was introduced to improve the physical consistency of reconstructed temperature fields. These

methodological advancements ensure that the resulting uncertainty estimates are more representative of the actual data characteristics and better suited for regional climate assessment, since the relative magnitudes of different uncertainty components are explicitly quantified and separated.

The results reveal distinct temporal evolution in LSAT uncertainty at both global and regional scales, with a pronounced decline over time, reflecting progressive improvements in data reliability. Nevertheless, a certain spatial heterogeneity remains: the mid- and low-latitude regions of the NH exhibit the highest confidence, while the SH, particularly high-latitude areas, continue to experience substantial uncertainty due to sparse observational coverage.

Further decomposition indicates that LSAT uncertainty is primarily dominated by coverage uncertainty. Globally, 1σ coverage uncertainty decreased from approximately 0.3℃ in the mid-19th century to about 0.05℃ in recent decades. Regionally, annual coverage uncertainty remains above 0.10℃ in the SH, especially over Antarctica, whereas in the NH, it has dropped below 0.01℃. In contrast, observation and analysis uncertainties remain minor and play a less dominant role in shaping total LSAT uncertainty.

For GMST, the total uncertainty is largely governed by SST uncertainty, with 1σ uncertainty decreasing from around 0.12℃ in the 19th century to below 0.03℃ in recent decades, indicating a significant improvement in the robustness of global temperature estimates.

Beyond the numerical findings, this study underscores the essential role of rigorous uncertainty assessment in climate science. By clearly identifying the sources, magnitudes, and spatial-temporal patterns of uncertainties, the China-MST 3.0 framework not only facilitates more accurate interpretation of historical temperature records, but also lays a solid scientific foundation for detection and attribution studies, trend evaluations, and climate policy formulation.

Distinct from ensemble-based products that propagate observation uncertainty into the reconstruction process, China-MST 3.0 is constructed as a single-member best estimate, with uncertainty systematically assessed through a traceable, stepwise decomposition framework. This approach ensures that different types of uncertainty, whether from station data, methodological limitations, or coverage gaps, are independently quantified rather than conflated. This design not only improves transparency but also avoids overestimating uncertainty in poorly observed regions, such as Antarctica.

Moreover, this framework provides a methodological reference for improving future observational systems and data assimilation efforts, particularly in under-observed regions. In a broader sense, advancing the quantification and communication of uncertainty is pivotal to enhancing the credibility and usability of climate datasets in both scientific and decision-making communities.

AI Tools Disclosure

We have used generative AI tools (ChatGPT) only for language polishing and grammar refinement during the manuscript preparation process. All AI-generated content has been reviewed and edited, and we take full responsibility for the final content of this publication.

Acknowledgments

This study is funded by the Natural Science Foundation of China (Grant 41975105). The authors sincerely appreciate the three reviewers for their valuable and constructive feedback, which significantly contributed to the improvement of this manuscript. Conceptualization: Q.L.; Methodology: Z.L., Q.L., B.J., L.C.; Data Curation: Q.X., S.W., X.R., P.S.; Investigation: Z.L., Q.L.; Visualization: Z.L.; Funding acquisition: Q.L.; Supervision: Q.L., B.H., P.J.; Writing – original draft:

Z.L.; Writing – review & editing: Q.L., B.J., Q.X., S.W., X.R., P.S., L.C., H.Y.Z., J.X., L.S., H.X.Z., B.H., P.J.

Open Research

Data Availability Statement

China-MST 3.0 dataset provides gridded global surface temperature anomalies, global and hemispheric mean temperature time series and GMST uncertainty estimates. These data are publicly available at http://www.gwpu.net/en/ (English homepage). The dataset is archived for long-term preservation and open access to support future research and applications.

Conflict of Interest

The authors declare no conflict of interest relevant to this study.

Appendix A. Grid-scale uncertainty estimation

This appendix describes the estimation of individual uncertainty components at the grid scale, including station, sampling, parameter, and reconstruction uncertainties. Each component represents a distinct source of uncertainty introduced during different stages of the dataset construction process.

Station uncertainty

Station-level uncertainty quantifies the combined effects of measurement uncertainty, homogenization uncertainty, and normals uncertainty. The uncertainty of individual stations is obtained by combining these variances in quadrature:

$$\sigma_{sta}(i,x,y,t) = \sqrt{\sigma_{mea}^2(i,x,y,t) + \sigma_{homo}^2(i,x,y,t) + \sigma_{nor}^2(i,x,y,t)} \qquad (A1)$$

where i, x, y, and t denote station index, longitude, latitude, and time, respectively; σ_{sta}

is the station uncertainty, σ_{mea} is the measurement uncertainty, σ_{homo} is the homogenization uncertainty, and σ_{nor} is the normals uncertainty.

At the grid scale, the uncertainty for each grid cell is estimated as the root mean square (RMS) of the uncertainties from all N stations within the grid cell:

$$\sigma_{sta}(x,y,t) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\sigma_{stat}^2(i,x,y,t)\right)} \quad (A2)$$

Sampling uncertainty

Sampling uncertainty reflects the representativeness of the available stations within a grid box. The Sampling uncertainty at the grid-box level is calculated following the method proposed by Jones et al. (1997):

$$\sigma_{sam}(x,y,t) = \sqrt{\frac{\overline{s_i^2}\bar{r}(1-\bar{r})}{1+(n-1)\bar{r}}} \quad (A3)$$

where $\overline{s_i^2}$ is the average variance of station anomalies, $\bar{r}$ is the average interstation correlation, and n is the number of stations within the grid box.

Although $\bar{r}$ can, in theory, be estimated directly from station data in grid boxes with multiple stations, it is generally derived using the correlation decay length theory (Briffa & Jones, 1993; Jones & Briffa, 1996) due to sparse and uneven station coverage:

$$r = e^{-\frac{x}{x_0}} \quad (A4)$$

$$\bar{r} = \frac{x_0}{X}\left(1 - e^{-\frac{X}{x_0}}\right) \quad (A5)$$

where r is the correlation between neighboring boxes, x is the distance between the box centers, x_0 is the characteristic correlation decay length, and X is the distance between opposite corners of a grid box.

Parameter uncertainty

In the reconstruction process, parameter uncertainty arises from different

parameterization settings. Following Huang et al. (2020), parameter uncertainty at each grid point is quantified as the standard deviation of the reconstructed temperature across all parameter sets:

$$\sigma_p(x,y,t) = \sqrt{\frac{1}{M}\sum_{m=1}^{M}[A_m(x,y,t) - \bar{A}(x,y,t)]^2} \quad (A6)$$

$$\bar{A} = \frac{1}{M}\sum_{m=1}^{M} A_m(x,y,t) \quad (A7)$$

where σ_p is the parameter uncertainty, M=252 is the total number of ensemble members, A_m is the reconstructed temperature of the m-th member, and $\bar{A}$ denotes the ensemble mean. The x, y, and t represent longitude, latitude, and time, respectively.

Reconstruction uncertainty

Reconstruction uncertainty represents the intrinsic uncertainty associated with the reconstruction method itself. It is calculated as the difference between the original and reconstructed values. Following Huang et al. (2020), reconstruction uncertainty is expressed as:

$$\sigma_r(x,y,t) = \sqrt{\frac{1}{M}\sum_{m=1}^{M}[R_m(x,y,t) - D(x,y,t)]^2} \quad (A8)$$

where σ_r is the reconstruction uncertainty, D represents the pseudo-observations, and R_m denotes the reconstructed data from D using the high- and low-frequency reconstruction methods.

Appendix B. Uncertainty in spatially averaged time series

This appendix describes how grid-level uncertainties propagated to regionally averaged time series. Four components are considered: observation uncertainty, parameter uncertainty, reconstruction uncertainty, and coverage uncertainty.

Observation uncertainty

When each type of observation uncertainty (including station, sampling, and bias uncertainties) is propagated to the regional mean, the uncertainty for any component is calculated as:

$$\sigma(t) = \sqrt{\sum_{x,y} (\sigma(x,y,t)\omega_{x,y})^2} \quad (B1)$$

Each grid-cell weight $\omega_{x,y}$ is defined as the area of the respective grid cell, normalized by the sum of all non-missing grid-cell areas within the region.

Parameter uncertainty

For regional mean temperatures, parameter uncertainty is expressed as:

$$\sigma_p(t) = \sqrt{\frac{1}{M}\sum_{m=1}^{M} [A_m^r(t) - \overline{A^r}(t)]^2} \quad (B2)$$

$$\overline{A^r} = \frac{1}{M}\sum_{m=1}^{M} A_m^r(t) \quad (B3)$$

where the superscript r indicates a regional average.

Reconstruction uncertainty

The regional mean reconstruction uncertainty is similarly defined:

$$\sigma_r(t) = \sqrt{\frac{1}{M}\sum_{m=1}^{M} [R_m^r(t) - D^r(t)]^2} \quad (B4)$$

Coverage uncertainty

To quantify the uncertainty introduced by an incomplete spatial field, we adopt a sub-sampling approach similar to that of Morice et al. (2021). A complete reanalysis dataset ERA5 reanalysis (Hersbach et al., 2020) for the period 1961-1990 is used as the reference field. For each month (or year), we apply the observational coverage mask

corresponding to the constrained dataset to generate a masked field. Regional means are then calculated for both the complete and masked fields, yielding $\tilde{B}_p(t)$ and $B_p(t)$, respectively. Here, p=30 is the number of years used in the resampling experiment.

The coverage-induced error for each realization is given by:

$$\epsilon_p(t) = \tilde{B}_p(t) - B_p(t) \quad (B5)$$

The coverage uncertainty σ_c at time t is then estimated as the root mean square of these errors:

$$\sigma_c(t) = \sqrt{\frac{1}{P}\sum_{p=1}^{P}\left[\epsilon_p(t)\right]^2} \quad (B6)$$

Appendix C. Combination of multiple uncertainty components

This appendix presents how different uncertainty sources are combined to obtain the total LSAT uncertainty.

Observation uncertainty

Observation uncertainty represents the combined effect of three statistically independent components: station sampling uncertainty, sampling uncertainty, and bias uncertainty. Their total variance is obtained by summing in quadrature:

$$\sigma_{obs}(t) = \sqrt{\sigma_{sta}^2(t) + \sigma_{sam}^2(t) + \sigma_{bias}^2(t)} \quad (C1)$$

Here, σ_{obs} represents the total observational uncertainty. σ_{bias} is the bias uncertainty, calculated from two independent large-scale systematic components (urbanization effects and thermometer exposure bias) after propagating their grid-level uncertainties to the regional mean.

Analysis uncertainty

Analysis uncertainty arises from the reconstruction process. It is obtained by combining

parametric and reconstruction uncertainty in quadrature:

$$\sigma_{ana}(t) = \sqrt{\overline{\sigma_p^2(t) + \sigma_r^2(t)}} \quad (C2)$$

where σ_{ana} denotes the analysis uncertainty.

Total LSAT uncertainty

Finally, the overall LSAT uncertainty incorporates observation, analysis, and coverage uncertainties, assuming mutual independence:

$$\sigma_{LSAT}(t) = \sqrt{\overline{\sigma_{obs}^2(t) + \sigma_{ana}^2(t) + \sigma_{cov}^2(t)}} \quad (C3)$$

where σ_{LSAT} denotes the total LSAT uncertainty.

Appendix D. Observational constrain method

When reconstructing LSAT, although the EOT method effectively captures large-scale trends and spatial patterns, its accuracy may decline during early periods (e.g., the 19th century) and in regions with sparse observations. Therefore, it is necessary to evaluate and refine the reliability of the reconstructed data.

In previous versions, a distance-based constraint method was used (Sun et al., 2021, 2022). Specifically, if the distance between a reconstructed grid point and the nearest observation data was within a specified threshold, the reconstructed data were considered reliable; otherwise, they were flagged as missing. However, this approach has several limitations. First, in early periods, sparse observations often resulted in reconstructions that retained only low-frequency variability (i.e., smaller amplitude fluctuations), while high-frequency components (i.e., larger short-term variations) are poorly reconstructed or entirely absent. This outcome arises because sparse observations inhibit the accurate reconstruction of high-frequency variability. In such cases, it is more appropriate to treat the affected data as missing. Second, the application of a uniform distance threshold globally ignored the varying length of latitude circles, resulting in uneven spatial constraints: grid points at low latitudes are constrained over

larger areas than those at high latitudes, leading to latitudinal inconsistencies in the weighting of constraint criteria.

To address these issues, a new variance-based observational constraint method was developed, based on the variance characteristics of the high-frequency component at each grid point (Li & Li, 2024). The procedure consists of three main steps:

(a) Sliding F-test:

For each grid point, a sliding F-test is applied to the monthly time series starting from January 1850 (Chen & Xie, 2006). Each time point τ in the series is treated as a candidate point that divides the record into pre-τ and post-τ segments. The F-test then evaluates whether the difference in variance between these two segments is statistically significant (Figure D1).

Before performing the test, two conditions must be satisfied:

(1) the variance of the pre-τ segment must be smaller than that of the post-τ segment;

(2) if the variance of the pre-τ segment is zero, the starting point of the time series is shifted to the first nonzero-variance month, after which the test is reinitiated.

The F statistic and corresponding p-value are computed as:

$$F = \frac{S_{max}^2}{S_{min}^2} \quad (D1)$$

$$p = 1 - I_x\left(\frac{d_1}{2}, \frac{d_2}{2}\right) \quad (D2)$$

where $x = \frac{d_1 F}{d_1 F + d_2}$, $d_1 = n_{max} - 1$, $d_2 = n_{min} - 1$, I_x denotes the regularized incomplete Beta function.

However, when F statistic is large, the p-value can be extremely small, potentially causing numerical underflow (recorded as zero). To avoid this, a tail approximation of the Beta function is adopted to compute very small p-values more accurately:

$$p \approx \frac{\left(\frac{d_1 F}{d_1 F + d_2}\right)^{\frac{d_2}{2}}}{B\left(\frac{d_1}{2}, \frac{d_2}{2}\right)} \quad (D3)$$

where B denotes the Beta function.

Finally, the results are expressed as $-\log_{10}(p)$ for easier comparison.

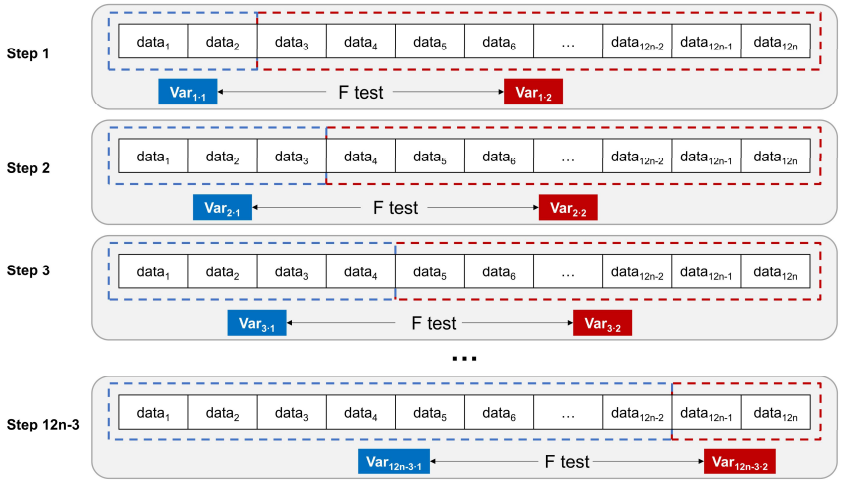


Figure D1. Conceptual illustration of the sliding F-test. Each candidate point τ divides the monthly time series into pre-τ and post-τ segments for variance comparison. n denotes the number of years of data, with 12n monthly samples used in total.

(b) Identification of the variance change point τ_0:

Among all candidate points τ, only those satisfying the following three conditions are retained as potential variance change points τ^*:

(1) $p < 0.01$ (equivalently, $-\log_{10}(p) > 2$);

(2) the variance of the monthly data before the change point is smaller than the annual mean variance after it.

If multiple qualified points exist for a given grid cell, the point with the maximum $-\log_{10}(p)$ is selected as the final variance change point τ_0, ensuring that the identified

change represents the most statistically significant variance shift.

(c) Application of the constraint

For grid points with an identified variance change point τ_0, the data prior to τ_0 are considered less reliable. These earlier values are therefore masked as missing, thereby imposing the observational constraint.

Reference

Briffa, K. R., & Jones, P. D. (1993). Global surface air temperature variations during the twentieth century: Part 2 , implications for large-scale high-frequency palaeoclimatic studies. *The Holocene*, *3*(1), 77–88. https://doi.org/10.1177/095968369300300109

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

Chan, D., Gebbie, G., Huybers, P., & Kent, E. C. (2024). A Dynamically Consistent ENsemble of Temperature at the Earth surface since 1850 from the DCENT dataset. *Scientific Data*, *11*(1), 953. https://doi.org/10.1038/s41597-024-03742-x

Chen, G., & Xie, P. (2006). Slide F test of Change-Point Analysis. *Journal of China Hydrology*, (02), 57–60. https://doi.org/10.3969/j.issn.1000-0852.2006.02.014

van den Dool, H. M., Saha, S., & Johansson, Å. (2000). Empirical Orthogonal Teleconnections. *Journal of Climate*, *13*(8), 1421–1435. https://doi.org/10.1175/1520-0442(2000)013%3C1421:EOT%3E2.0.CO;2

Folland, C. K., Rayner, N. A., Brown, S. J., Smith, T. M., Shen, S. S. P., Parker, D. E., et al. (2001). Global temperature change and its uncertainties since 1861.

974 *Geophysical Research Letters*, *28*(13), 2621–2624.
975 https://doi.org/10.1029/2001GL012877

976 Freeman, E., Woodruff, S. D., Worley, S. J., Lubker, S. J., Kent, E. C., Angel, W. E., et
977 al. (2017). ICOADS Release 3.0: a major update to the historical marine climate
978 record. *International Journal of Climatology*, *37*(5), 2211–2232.
979 https://doi.org/10.1002/joc.4775

980 Hawkins, E., Ortega, P., Suckling, E., Schurer, A., Hegerl, G., Jones, P., et al. (2017).
981 Estimating Changes in Global Temperature since the Preindustrial Period.
982 *Bulletin of the American Meteorological Society*, *98*(9), 1841–1856.
983 https://doi.org/10.1175/BAMS-D-16-0007.1

984 Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., et
985 al. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal*
986 *Meteorological Society*, *146*(730), 1999–2049. https://doi.org/10.1002/qj.3803

987 Huang, B., Thorne, P. W., Smith, T. M., Liu, W., Lawrimore, J., Banzon, V. F., et al.
988 (2016). Further Exploring and Quantifying Uncertainties for Extended
989 Reconstructed Sea Surface Temperature (ERSST) Version 4 (v4). *Journal of*
990 *Climate*, *29*(9), 3119–3142. https://doi.org/10.1175/JCLI-D-15-0430.1

991 Huang, B., Thorne, P. W., Banzon, V. F., Boyer, T., Chepurin, G., Lawrimore, J. H., et
992 al. (2017). Extended Reconstructed Sea Surface Temperature, Version 5
993 (ERSSTv5): Upgrades, Validations, and Intercomparisons. *Journal of Climate*,

30(20), 8179–8205. https://doi.org/10.1175/JCLI-D-16-0836.1

Huang, B., Menne, M. J., Boyer, T., Freeman, E., Gleason, B. E., Lawrimore, J. H., et al. (2020). Uncertainty Estimates for Sea Surface Temperature and Land Surface Air Temperature in NOAAGlobalTemp Version 5. *Journal of Climate*, *33*, 1351–1379. https://doi.org/10.1175/JCLI-D-19-0395.1

Huang, B., Yin, X., Boyer, T., Liu, C., Menne, M., Rao, Y. D., et al. (2025). Extended Reconstructed Sea Surface Temperature, Version 6 (ERSSTv6). Part I: An Artificial Neural Network Approach. *Journal of Climate*, *38*(4), 1105–1121. https://doi.org/10.1175/JCLI-D-23-0707.1

IPCC. (2021). *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press.

Jones, P. D. (1994). Hemispheric Surface Air Temperature Variations: A Reanalysis and an Update to 1993. *Journal of Climate*, *7*(11), 1794–1802. Retrieved from https://journals.ametsoc.org/view/journals/clim/7/11/1520-0442_1994_007_1794_hsatva_2_0_co_2.xml

Jones, P. D., & Briffa, K. R. (1996). What can the instrumental record tell us about longer timescale paleoclimatic reconstructions? In P. D. Jones, R. S. Bradley, & J. Jouzel (Eds.), *Climatic Variations and Forcing Mechanisms of the Last 2000*

Years (pp. 625–644). Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-642-61113-1_30

Jones, P. D., Groisman, P. Y., Coughlan, M., Plummer, N., Wang, W.-C., & Karl, T. R. (1990). Assessment of urbanization effects in time series of surface air temperature over land. *Nature*, *347*(6289), 169–172. https://doi.org/10.1038/347169a0

Jones P. D., Osborn T. J., & Briffa K. R. (1997). Estimating Sampling Errors in Large-Scale Temperature Averages. *Journal of Climate*, *10*(10), 2548–2568. Retrieved from https://journals.ametsoc.org/view/journals/clim/10/10/1520-0442_1997_010_2548_eseils_2.0.co_2.xml

Ku, H. H. (1966). Notes on the use of propagation of error formulas. *Journal of Research of the National Bureau of Standards, Section C: Engineering and Instrumentation*, *70C*(4), 263–273.

Lenssen, N., Schmidt, G., Hansen, J., Menne, M., Persin, A., Ruedy, R., & Zyss, D. (2019). Improvements in the GISTEMP Uncertainty Model. *Journal of Geophysical Research: Atmospheres*, *124*(12), 6307–6326. https://doi.org/10.1029/2018JD029522

Lenssen, N., Schmidt, G., Hendrickson, M., Jacobs, P., Menne, M., & Ruedy, R. (2024). A NASA GISTEMPv4 Observational Uncertainty Ensemble. *Journal of Geophysical Research: Atmospheres*, *129*(17), e2023JD040179.

https://doi.org/10.1029/2023JD040179

Li, Q., Dong, W., Li, W., Gao, X., Jones, P., Kennedy, J., & Parker, D. (2010). Assessment of the uncertainties in temperature change in China during the last century. *Chinese Science Bulletin*, *55*(19), 1974–1982. https://doi.org/10.1007/s11434-010-3209-1

Li, Q., Sun, W., Yun, X., Huang, B., Dong, W., Wang, X. L., et al. (2021). An updated evaluation of the global mean land surface air temperature and surface temperature trends based on CLSAT and CMST. *Climate Dynamics*, *56*(1), 635–650. https://doi.org/10.1007/s00382-020-05502-0

Li, Z., & Li, Q. (2024). Method, Apparatus, Device, and Storage Medium for Reconstructing Observational Data, Chinese Patent CN118132942B, China National Intellectual Property Administration.

Li, Z., Sun, W., Liang, C., Xing, X., & Li, Q. (2023). Arctic warming trends and their uncertainties based on surface temperature reconstruction under different sea ice extent scenarios. *Advances in Climate Change Research*, *14*(3), 335–346. https://doi.org/10.1016/j.accre.2023.06.003

Liu, C., Freeman, E., Kent, E. C., Berry, D. I., Worley, S. J., Smith, S. R., et al. (2022). Blending TAC and BUFR Marine In Situ Data for ICOADS Near-Real-Time Release 3.0.2. *Journal of Atmospheric and Oceanic Technology*, *39*(12), 1943–1959. https://doi.org/10.1175/JTECH-D-21-0182.1

1054 Menne, M. J., & Williams, C. N. (2009). Homogenization of Temperature Series via
1055 Pairwise Comparisons. *Journal of Climate*, *22*(7), 1700–1717.
1056 https://doi.org/10.1175/2008JCLI2263.1

1057 Menne, M. J., Williams, C. N., & Vose, R. S. (2009). The U.S. Historical Climatology
1058 Network Monthly Temperature Data, Version 2. *Bulletin of the American*
1059 *Meteorological Society*, *90*(7), 993–1008.
1060 https://doi.org/10.1175/2008BAMS2613.1

1061 Menne, M. J., Williams, C. N., Gleason, B. E., Rennie, J. J., & Lawrimore, J. H. (2018).
1062 The Global Historical Climatology Network Monthly Temperature Dataset,
1063 Version 4. *Journal of Climate*, *31*(24), 9835–9854.
1064 https://doi.org/10.1175/JCLI-D-18-0094.1

1065 Morice, C. P., Kennedy, J. J., Rayner, N. A., & Jones, P. D. (2012). Quantifying
1066 uncertainties in global and regional temperature change using an ensemble of
1067 observational estimates: The HadCRUT4 data set. *Journal of Geophysical*
1068 *Research: Atmospheres*, *117*(D8). https://doi.org/10.1029/2011JD017187

1069 Morice, C. P., Kennedy, J. J., Rayner, N. A., Winn, J. P., Hogan, E., Killick, R. E., et al.
1070 (2021). An Updated Assessment of Near-Surface Temperature Change From
1071 1850: The HadCRUT5 Data Set. *Journal of Geophysical Research:*
1072 *Atmospheres*, *126*(3). https://doi.org/10.1029/2019JD032361

1073 Morice, C. P., Berry, D. I., Cornes, R. C., Cowtan, K., Cropper, T., Hawkins, E., et al.

(2024). An observational record of global gridded near surface air temperature change over land and ocean from 1781. *Earth System Science Data Discussions*, 1–32. https://doi.org/10.5194/essd-2024-500

Parker, D. E. (1994). Effects of changing exposure of thermometers at land stations. *International Journal of Climatology*, *14*(1), 1–31. https://doi.org/10.1002/joc.3370140102

Rohde, R., & Hausfather, Z. (2020). The Berkeley Earth Land/Ocean Temperature Record. *Earth System Science Data*, *12*(4), 3469–3479. https://doi.org/10.5194/essd-12-3469-2020

Rohde, R., Muller, R., Jacobsen, R., Perlmutter, S., Rosenfeld, A., Wurtele, J., et al. (2013). Berkeley Earth Temperature Averaging Process. *Geoinformatics & Geostatistics: An Overview*, *1*(2). https://doi.org/10.4172/2327-4581.1000103

Sun, W., Li, Q., Huang, B., Cheng, J., Song, Z., Li, H., et al. (2021). The Assessment of Global Surface Temperature Change from 1850s: The C-LSAT2.0 Ensemble and the CMST-Interim Datasets. *Advances in Atmospheric Sciences*, *38*(5), 875–888. https://doi.org/10.1007/s00376-021-1012-3

Sun, W., Yang, Y., Chao, L., Dong, W., Huang, B., Jones, P., & Li, Q. (2022). Description of the China global Merged Surface Temperature version 2.0. *Earth System Science Data*, *14*(4), 1677–1693. https://doi.org/10.5194/essd-14-1677-2022

Wallis, E. J., Osborn, T. J., Taylor, M., Jones, P. D., Joshi, M., & Hawkins, E. (2024). Quantifying exposure biases in early instrumental land surface air temperature observations. *International Journal of Climatology*, *44*(5), 1611–1635. https://doi.org/10.1002/joc.8401

Wang, X. (2008a). Accounting for Autocorrelation in Detecting Mean Shifts in Climate Data Series Using the Penalized Maximal t or F Test. *Journal of Applied Meteorology and Climatology*, *47*(9), 2423–2444. https://doi.org/10.1175/2008JAMC1741.1

Wang, X. (2008b). Penalized Maximal F Test for Detecting Undocumented Mean Shift without Trend Change. *Journal of Atmospheric and Oceanic Technology*, *25*(3), 368–384. https://doi.org/10.1175/2007JTECHA982.1

Wei, S., Li, Q., Xu, Q., Li, Z., Zhang, H., & Lin, J. (2025). Updates to C-LSAT 2.1 and the development of high-resolution land surface air temperature and diurnal temperature range datasets. *Earth System Science Data*, *17*(9), 4985–5005. https://doi.org/10.5194/essd-17-4985-2025

World Meteorological Organization. (2007). *The Role of Climatological Normals in a Changing Climate*. Geneva. Retrieved from https://library.wmo.int/records/item/52499-the-role-of-climatological-normals-in-a-changing-climate

World Meteorological Organization. (2017). *WMO Guidelines on the Calculation of*

Climate Normals. Geneva. Retrieved from https://library.wmo.int/records/item/55797-wmo-guidelines-on-the-calculation-of-climate-normals

Xu, Q., Wei, S., Li, Z., & Li, Q. (2025). A New Evaluation of Observed Changes in Diurnal Temperature Range. *Geophysical Research Letters*, *52*(2), e2024GL113406. https://doi.org/10.1029/2024GL113406

Xu, W., Li, Q., Jones, P., Wang, X. L., Trewin, B., Yang, S., et al. (2018). A new integrated and homogenized global monthly land surface air temperature dataset for the period since 1900. *Climate Dynamics*, *50*(7), 2513–2536. https://doi.org/10.1007/s00382-017-3755-1

Yin, X., Huang, B., Menne, M., Vose, R., Zhang, H.-M., Adeyeye, A., et al. (2024). NOAAGlobalTemp Version 6: An AI-Based Global Surface Temperature Dataset. *Bulletin of the American Meteorological Society*, *105*(11), E2184–E2193. https://doi.org/10.1175/BAMS-D-24-0012.1

Yun, X., Huang, B., Cheng, J., Xu, W., Qiao, S., & Li, Q. (2019). A new merge of global surface temperature datasets since the start of the 20th century. *Earth System Science Data*, *11*(4), 1629–1643. https://doi.org/10.5194/essd-11-1629-2019

References From the Supporting Information

Chao, L., Huang, B., Yuanjian, Y., Jones, P., Cheng, J., Yang, Y., & Li, Q. (2020). A New Evaluation of the Role of Urbanization to Warming at Various Spatial

Scales: Evidence From the Guangdong-Hong Kong-Macau Region, China. Geophysical Research Letters, 47(20), e2020GL089152. https://doi.org/10.1029/2020GL089152

Elvidge, C. D., Zhizhin, M., Ghosh, T., Hsu, F.-C., & Taneja, J. (2021). Annual Time Series of Global VIIRS Nighttime Lights Derived from Monthly Averages: 2012 to 2019. Remote Sensing, 13(5), 922. https://doi.org/10.3390/rs13050922

Li, Q., Zhang, H., Liu, X., & Huang, J. (2004). Urban heat island effect on annual mean temperature during the last 50 years in China. Theoretical and Applied Climatology, 79(3), 165–174. https://doi.org/10.1007/s00704-004-0065-4

Li, Q., Dong, W., Li, W., Gao, X., Jones, P., Kennedy, J., & Parker, D. (2010). Assessment of the uncertainties in temperature change in China during the last century. Chinese Science Bulletin, 55(19), 1974–1982. https://doi.org/10.1007/s11434-010-3209-1

Li, W. (2023). Analysis of the Impacts of Urbanization on Land Surface Air Temperature and Thermal Comfort Changes (MA thesis). Sun Yat-sen University.

Pesaresi, M., Florczyk, A., Schiavina, M., Melchiorri, M., & Maffenini, L. (2019). GHS settlement grid, updated and refined REGIO model 2014 in application to GHS-BUILT R2018A and GHS-POP R2019A. https://doi.org/10.2905/42E8BE89-54FF-464E-BE7B-BF9E64DA5218

Peterson, T. C., & Easterling, D. R. (1994). Creation of homogeneous composite climatological reference series. International Journal of Climatology, 14(6), 671–679. https://doi.org/10.1002/joc.3370140606

Wang, F., Ge, Q., Wang, S., Li, Q., & Jones, P. D. (2015). A New Estimation of Urbanization's Contribution to the Warming Trend in China. Journal of Climate, 28(22), 8923–8938. https://doi.org/10.1175/JCLI-D-14-00427.1

Wang, X. (2008a). Accounting for Autocorrelation in Detecting Mean Shifts in Climate Data Series Using the Penalized Maximal t or F Test. Journal of Applied Meteorology and Climatology, 47(9), 2423–2444. https://doi.org/10.1175/2008JAMC1741.1

Wang, X. (2008b). Penalized Maximal F Test for Detecting Undocumented Mean Shift without Trend Change. Journal of Atmospheric and Oceanic Technology, 25(3), 368–384. https://doi.org/10.1175/2007JTECHA982.1

Wang, X., & Feng, Y. (2013, July). RHtestsv4 user manual. Climate Research Division, Atmospheric Science and Technology Directorate, Science and Technology Branch, Environment Canada. Retrieved from https://github.com/ECCC-CDAS(open in a new window)

Xu, Q., Wei, S., Li, Z., & Li, Q. (2025). A New Evaluation of Observed Changes in Diurnal Temperature Range. Geophysical Research Letters, 52(2), e2024GL113406. https://doi.org/10.1029/2024GL113406

1174 Xu, W., Li, Q., Wang, X. L., Yang, S., Cao, L., & Feng, Y. (2013). Homogenization of
1175 Chinese daily surface air temperatures and analysis of trends in the extreme
1176 temperature indices. Journal of Geophysical Research: Atmospheres, 118(17),
1177 9708–9720. https://doi.org/10.1002/jgrd.50791

1178 Xu, W., Li, Q., Jones, P., Wang, X. L., Trewin, B., Yang, S., et al. (2018). A new
1179 integrated and homogenized global monthly land surface air temperature dataset
1180 for the period since 1900. Climate Dynamics, 50(7), 2513–2536.
1181 https://doi.org/10.1007/s00382-017-3755-1

1182 Yang, X., Hou, Y., & Chen, B. (2011). Observed surface warming induced by
1183 urbanization in east China. Journal of Geophysical Research: Atmospheres,
1184 116(D14). https://doi.org/10.1029/2010JD015452

1185 Zhang, P., Ren, G., Qin, Y., Zhai, Y., Zhai, T., Tysa, S. K., et al. (2021). Urbanization
1186 Effects on Estimates of Global Trends in Mean and Extreme Air Temperature.
1187 Journal of Climate, 34, 1923–1945. https://doi.org/10.1175/JCLI-D-20-0389.1

1188 Zhou, L., Dickinson, R. E., Tian, Y., Fang, J., Li, Q., Kaufmann, R. K., et al. (2004).
1189 Evidence for a significant urbanization effect on climate in China. Proceedings
1190 of the National Academy of Sciences, 101(26), 9540–9544.
1191 https://doi.org/10.1073/pnas.0400357101
1192