A NASA GISTEMPv4 Observational Uncertainty Ensemble

Nathan Lenssen[1,2], Gavin A. Schmidt[3], Michael Hendrickson[3,6], Peter Jacobs[4], Matthew J. Menne[5], Reto Ruedy[3,6]

[1]Dept. of Earth and Environmental Sciences, Columbia University, New York, New York, USA
[2]Dept. of Applied Mathematics and Statistics, Colorado School of Mines, Golden, Colorado, USA
[3]NASA Goddard Institute for Space Studies, New York, New York, USA
[4]NASA Goddard Space Flight Center, Greenbelt, Maryland, USA
[5]NOAA National Centers for Environmental Information, Asheville, North Carolina, USA
[6]Autonomic Integra LLC, New York, New York, USA

Key Points:

- A 200-member ensemble quantifies uncertainty in historical surface temperature anomalies as estimated by GISTEMP
- This ensemble enables accurate statistical analyses of key global change metrics such as trends and rankings of temperature records.
- The median ensemble estimate agrees with the operational GISTEMP analyses and other global products

Corresponding author: Nathan Lenssen, lenssen@mines.edu

Abstract

The historical global temperature record is an essential data product for quantifying the variability and change of the Earth system. In recent years, better characterization of observational uncertainty in global and hemispheric trends has become available, but the methodologies are not necessarily applicable to analyses at smaller regional areas, or monthly or seasonal means, where station sparsity and other systematic issues contribute to greater uncertainty. This study presents a gridded uncertainty ensemble of historical surface temperature anomalies from the Goddard Institute for Space Studies (GISS) Surface Temperature (GISTEMP) product. This ensemble characterizes the complex spatial and temporal correlation structure of uncertainty, enabling better uncertainty propagation for climate and applied science at regional and sub-annual scales. This work details the methodology for generating the uncertainty ensemble, presents key statistics of the uncertainty evolution over space and time, and provides best practices for using the uncertainty ensemble in future studies. Summary statistics from the uncertainty ensemble agree with the previous GISTEMP global uncertainty assessment, providing confidence in both.

1 Introduction

In recent years, better characterization of uncertainty in observed global and hemispheric temperature trends has become available (Lenssen et al., 2019; Morice et al., 2020; Huang et al., 2020; Rohde & Hausfather, 2020), but the methodologies are not necessarily applicable to smaller spatiotemporal scales such as regional or monthly averages, where station sparsity and other systematic issues contribute to greater uncertainty. This study describes an ensemble of temperature reconstructions for the Goddard Institute for Space Studies (GISS) Surface Temperature product (GISTEMP) product (Hansen et al., 2010; Lenssen et al., 2019) which spans the possible regional and monthly uncertainty while properly accounting for the underlying spatiotemporal correlation structure.

Previous work quantified uncertainty in the GISTEMP estimate of large-scale annual mean series and developed critical components necessary for quantifying the uncertainty in the GISTEMP historical surface temperature record (Lenssen et al., 2019). Critically, Lenssen et al. (2019) formalized the various sources of uncertainty in the GISTEMP procedure and divided total uncertainty into independent, quantifiable components that represented the major sources of uncertainty in the land and ocean analyses. In the Land Surface Air Temperature (LSAT) record, the primary sources of uncertainty are sampling uncertainty and station homogenization uncertainty. Sampling uncertainty is an umbrella term for uncertainties introduced into global and regional means due to incomplete spatial and temporal coverage. Station homogenization uncertainty accounts for possible errors arising from the adjustment of single station records to correct artificial break points due to changes in observing methods or station locations. Using this framework, operational GISTEMP now provides an estimate of global mean uncertainty.

Extending the results of Lenssen et al. (2019) to regional and monthly mean temperature is a significant undertaking. There are two primary difficulties: (1) moving from global and large-scale spatial means to small-scale spatial means and (2) quantifying the temporal dependence of the uncertainty to provide accurate estimates of the uncertainty in changes in the mean. The temporal structure of the uncertainty is the most important problem, and is particularly important to capture correctly for accurate uncertainty quantification in global and regional trends. The simple 95% confidence intervals for the global mean discussed in Lenssen et al. (2019) do not include information about the temporal structure of uncertainty. It is well known that significant temporal autocorrelation in uncertainty exists, primarily driven over the land surface by the homogenization of the station record (Menne et al., 2018). The temporal structure of this homogenization uncertainty is highly persistent and not well represented by common statistical models for time series such as auto-regressive or more complex ARIMA models.

Creating ensembles of equally likely realizations of the global temperature record is the current best practice for quantifying and presenting uncertainty in gridded monthly historical temperature analyses (Morice et al., 2012, 2020; Huang et al., 2020). The Hadley Centre with HadCRUT4 (Morice et al., 2012) and HadCRUT5 (Morice et al., 2020) as well as NOAA's GlobalTemp Version 5 (Huang et al., 2020) have shifted their global temperature uncertainty products from simple confidence intervals to such uncertainty ensembles. In addition, the newer deep neural network (DNN)-based infilling of HadCRUT5 by the German Climate Computation Center (DKRZ), or the DKRZ-DNN global product, uses the HadCRUT5 ensemble to quantify uncertainty in their infilling method (Kadow et al., 2020). Each of these ensembles represent the complex and persistent temporal structure of the uncertainties inherent in the global temperature record, enable more accurate estimates of uncertainty in global and regional temperature change, and make it straightforward to include observational uncertainty in subsequent analyses.

This study presents a monthly, gridded GISTEMPv4 uncertainty ensemble from 1880-2020. Following the operational GISTEMP analysis, Land Surface Air Temperature (LSAT) is calculated from station records from NOAA NCEI's Global Historical Climatology Network (GHCN) monthly version 4 (GHCNm v4; (Menne et al., 2018)). Sea Surface Temperature (SST) data from NOAA's Extended Reconstructed Sea Surface Temperature version 5 (ERSSTv5; (Huang et al., 2017)) is merged with the LSAT analysis to form the GISTEMP global land-ocean analysis (Hansen et al., 2010; Lenssen et al., 2019).

One of the primary motivations behind the GISTEMP uncertainty ensemble is to increase the awareness of observational uncertainty in studies relying on historical temperature data. The global historical temperature record, and GISTEMP in particular, is widely accessed, cited, and used in subsequent studies: From the 10-most cited papers that cite Lenssen et al. (2019), direct applications of GISTEMP include: the validation of historical runs of global general circulation models (Swart et al., 2019; Held et al., 2019; Danabasoglu et al., 2020; Notz & SIMIP Community, 2020), retrospectively evaluating past climate model projections (Hausfather et al., 2020), verifying estimates of climate sensitivity (Tokarska et al., 2020), quantifying changes in mean climate and extremes over the historical period (Myhre et al., 2019), and estimating the cost of carbon emission in the global economy (Carleton et al., 2020). Despite the scientific and societal importance of the problems addressed in these studies, and their reliance on the historical global temperature record, none of them include observational uncertainty as part of their methodologies.

A potential reason for the near ubiquitous omission of observational uncertainty in analyses involving historical climate data is the lack of accessible, interpretable, and easily-implemented uncertainty products. Observational ensembles are a large step forward, as posterior distributions of a key result in an analysis that relies on historical temperature can be constructed nearly trivially by running the analysis of interest on each uncertainty ensemble member. However, these ensembles are relatively new, only appearing in the last decade, with very few studies utilizing them as yet. Thus, this study also includes a description of easy-to-implement best practices for including observational uncertainty in studies that use historical surface temperature products.

The remainder of this paper is organized as follows. Section 2 outlines the source data used for the analyses. Section 3 provides a brief background on the LSAT uncertainty model discussed in detail in (Lenssen et al., 2019). Section 4 presents the methods used to generate the GISTEMP uncertainty ensemble. Section 5 summarizes the statistical properties of the GISTEMP uncertainty ensemble. Section 6 discusses the implications of the results and provides best practices for implementing the GISTEMP uncertainty ensemble in future studies.

2 Input Data

2.1 LSAT Data: GHCNm Version 4

The GHCNm version 4 dataset is a quality-controlled collection of station-based land temperature records at the monthly temporal resolution (Menne et al., 2018). All station records included in the dataset are processed to correct for irregularities arising due to change of station location, measurement method, and surrounding land cover. In addition to a single authoritative station record, GHCNm v4 also contains a 100+ member uncertainty ensemble that spans the parametric uncertainty arising from choices in the homogenization procedure as detailed in Menne et al. (2018). This study uses the GHCNm v4 ensemble to capture the station and bias uncertainties as is discussed further in Section 3.

2.2 SST Data: ERSSTv5

The latest version of NOAA's gridded sea surface temperature analysis, ERSSTv5, is used to quantify the historical monthly SST anomalies globally (Huang et al., 2017). The product is distributed on a $2° \times 2°$ grid that is interpolated to the 8000 GISTEMP equal area boxes to be compatible with the operational GISTEMP python analysis. The uncertainty quantification in ERSSTv4/v5 breaks down ocean uncertainty into parametric uncertainty, or uncertainty arising from choices the ERSST method, and reconstruction uncertainty, or uncertainty arising from estimating global SST from limited SST records. The ERSSTv5 uncertainty model contains small updates to the parameters from the ERSSTv4 uncertainty method, but is otherwise identical (Liu et al., 2015; Huang et al., 2016, 2017). ERSSTv5 provides a 1,000 member uncertainty ensemble of gridded SST fields as well as a 500 member operational uncertainty ensemble, enabling other operational products to take advantage of their uncertainty assessment.

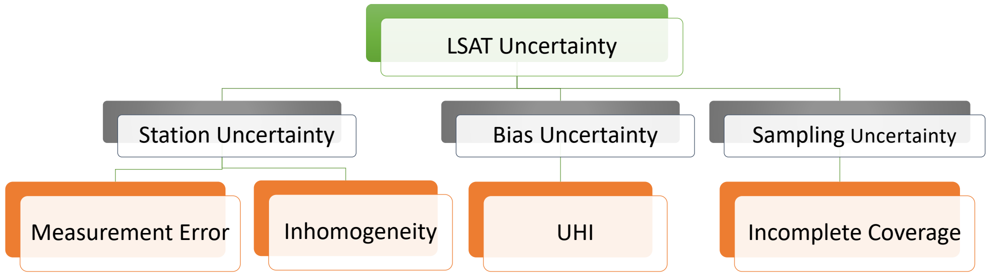


Figure 1. Decomposition the total LSAT uncertainty into the three major categories and the most common sources. The connections on the chart denote dependence, implying statistical independence between cells that are not connected.

2.3 ERA5 Reanalysis

This study uses the monthly ECMWF Reanalysis version 5 (ERA5) from 1951-2020 as an approximate, full-coverage historical LSAT record (Hersbach et al., 2020). The 2 m temperature field is averaged to the final $2° \times 2°$ GISTEMP uncertainty ensemble grid to facilitate direct comparison between ERA5 and GISTEMP. ERA5 is chosen as the reanalysis as a it best replicates the observed global mean over its period (Lenssen et al., 2019; Hersbach et al., 2020). As shown in Lenssen et al. (2019), global and large-scale GISTEMP uncertainty estimates derived from ERA5 agree with the JRA-55 and MERRA2 reanalyses.

3 LSAT Uncertainty

There are three major, statistically independent, categories of uncertainty that arise in the LSAT record (Figure 1). A brief introduction to these uncertainties is provided though see Morice et al. (2012), and Lenssen et al. (2019) for more details. The uncertainty ensemble model accounts for station and bias uncertainties through the GHCNm v4 ensemble as detailed in Section 4.1 and sampling uncertainties following the methodology outlined in Section 4.2.

Station uncertainty arises from errors in the temperature record of a single station. The first sources of station uncertainty are instrumental errors from limited thermometer precision. These are relatively small and uncorrelated in space and time, making them essentially a non-issue for monthly records (Morice et al., 2012). The other and more significant sources of station uncertainty are inhomogeneities, or non-climatic shifts in mean in station records. These can arise from local microclimate shifts or changes in the station measurement method. As mentioned in Section 2.1, the GHCNm dataset attempts to detect and correct for these inhomogeneities, but this is a difficult problem, and uncertainty due to statistical estimation of these corrections adds uncertainty to estimates of regional and global temperature while reducing any bias (e.g. (Hausfather et al., 2013)).

Bias uncertainty refers to anthropogenic changes in local climate that are not representative of changes in the regional or global climate system. Generally, this category of uncertainty refers to the enhanced warming observed in cities commonly referred to as the urban heat island (UHI) effect. Again, the GHCNm dataset accounts for these, but corrections add uncertainty to the surface temperature record. This issue has also been tackled in GISTEMP through the use of nightlights to characterize the more urban environments (Hansen et al., 2010).

Sampling uncertainty arises from estimating regional and global temperature due to limited spatial and temporal coverage. The distribution of the global observation network does not fully cover the land surface and has changed over time. GISTEMP uses the spatial correlation of temperature anomalies to increase the coverage (Hansen et al., 2010; Cowtan et al., 2018). By interpolating the station-level anomalies, GISTEMP is able to make a more accurate estimate of the global temperature, but at a cost of introducing uncertainty into fine-scale regional means.

Due to the different sources and relative contributions of these uncertainties, representing uncertainty at each location and month as independent or correlated Gaussian random variables is an incomplete method. In particular, uncertainty arising due to errors in the homogenization process have long-term persistence that are not well-suited to ARIMA time series models. Construction of an LSAT uncertainty ensemble using an iterative process where each step accounts for one of the two major categories of uncertainty, is better able to better represent the spatiotemporal structure of the uncertainties, and can be understood in isolation.

4 Methods

4.1 GHCN-ERSST-GISTEMP Ensemble

The core of the GISTEMP uncertainty is the GHCN-ERSST-GISTEMP ensemble which is generated by running 100 potential station records from the GHCNm v4 uncertainty

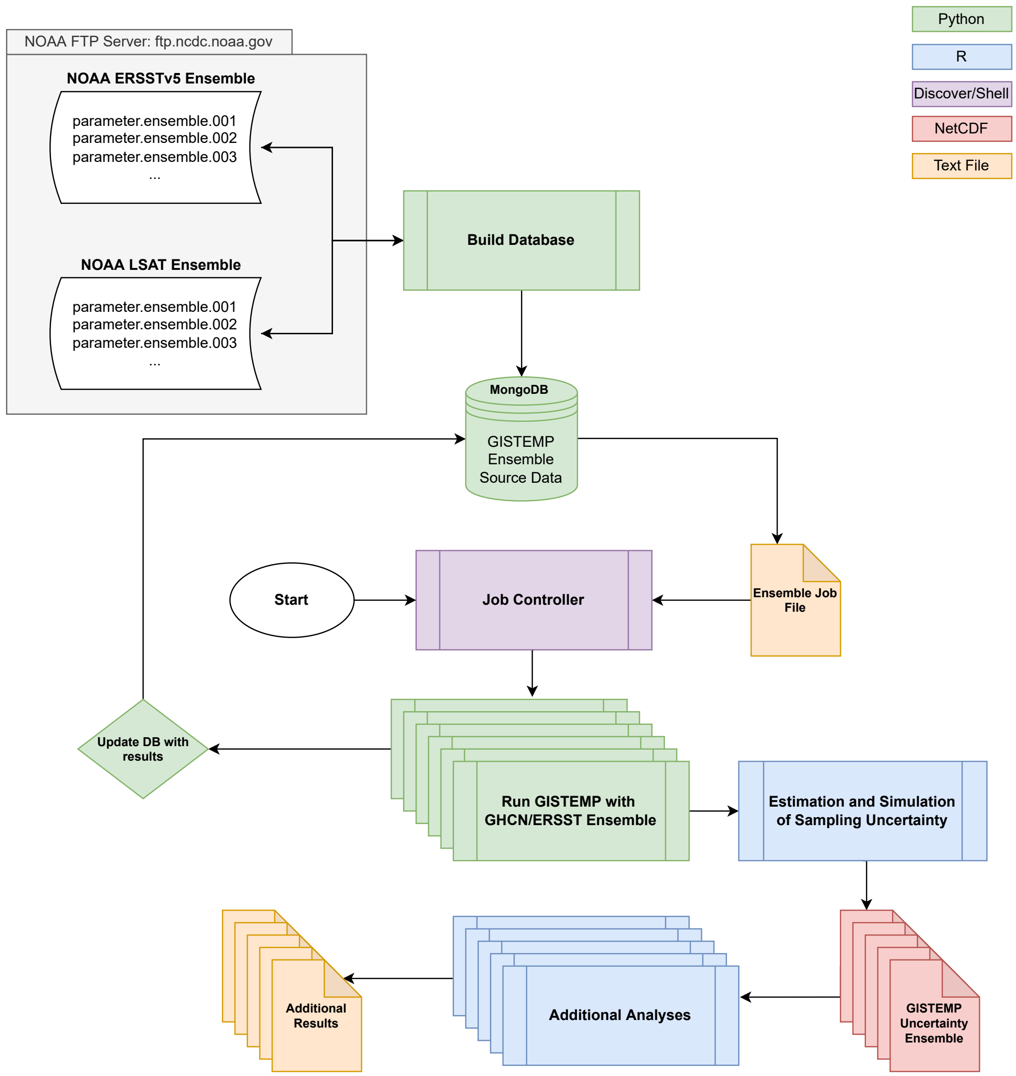


Figure 2. Organization of the analysis from the raw NOAA data in the upper-left corner to the final country-level mean estimates in the bottom-left corner. The legend in the upper-right denotes the primary language or datatype of each node.

paired with 100 of the ERSSTv5 uncertainty ensemble members. These station record-ocean record pairs are then run through the operational Python GISTEMP analysis code (Barnes & Jones, 2011). This process is outlined visually in the code flowchart (Figure 2) through steps at the top of the chart leading up to the block labeled “Run GISTEMP with NOAA Ensemble Data.” The GHCN ensemble is 100 possible station records in the same format as the version of GHCN used in production GISTEMP. Temperature fields and mean time series are calculated as described in Chapter 1 (Hansen et al., 2010; Lenssen et al., 2019). The 100 member GHCN-ERSST-GISTEMP ensemble accounts for all quantified SST uncertainty as well as homogenization and bias LSAT uncertainty. Thus, all that remains is to quantify the LSAT sampling uncertainty arising from limited station coverage as detailed in Section 4.2.

Managing the output of the ensemble members to ensure computations are working as intended and output is documented appropriately, is a critical part of the workflow. The steps in the flowchart prior to the "Run GISTEMP with GHCN/ERSST Ensemble" block describe the data and code management processes needed to organize the analysis on the NASA Center for Climate Simulations (NCCS) high performance cloud computing environment. By porting the analysis, the GHCN-ERSST-GISTEMP ensemble is able to be generated in under an hour as opposed to the days it would take to run on a typical laptop.

4.2 Sampling Uncertainty Ensemble

The 100 member GHCN-ERSST-GISTEMP ensemble detailed above in Section 4.1 accounts for the station and bias uncertainties. To incorporate the sampling uncertainty, 2 possible realizations of the sampling uncertainty are simulated and added to each of the 100 members of the GHCN/GISTEMP, resulting in a final uncertainty ensemble of 200 members. This step is performed in R (R Core Team, 2020) and is denoted by the blue "Estimation and Simulation of Sampling Uncertainty" on the analysis flowchart (Figure 2).

The sampling uncertainty is quantified using an improved version of GISTEMP sampling uncertainty analysis detailed in Lenssen et al. (2019). The ERA5 reanalysis from 1950-2020 is used as an approximate historical climate with full global coverage. For each decade from 1880-2020, a proxy station record is created by masking full-field ERA5 record to match the station coverage for that decade. That is, a station mask is created for $1880-1889, 1890-1899, \ldots, 2000-2009, 2010-2020$ where the last proxy station record includes 2020. Following Lenssen et al. (2019), a grid-cell is considered covered in a decade if it has a station with coverage for at least 5 of the 10 years. Annual coverage requires a station has coverage for at least 3 seasons which requires at least two months in the season.

The GISTEMP interpolation step with 1,200km smoothing is applied to the masked ERA5 data resulting in estimates of regional temperature on a $2^\circ \times 2^\circ$ grid. The true temperature anomaly fields from ERA5 are differenced to calculate reconstruction error fields for each time-step in the ERA5 record. These reconstruction error fields are an estimate of the uncertainty in the LSAT field due to limited station coverage as well as uncertainty introduced by the GISTEMP interpolation method.

Due to the interpolation in the GISTEMP method, the reconstruction error fields have spatial structure that must be accounted for. Here, the empirical reconstruction error fields are used as draws from the sampling uncertainty distribution as they inherently contain the correct spatial correlation structure. As the sampling uncertainty is independent to the homogenization uncertainty, the monthly empirical reconstruction error fields are added to each of the gridded GHCN-ERSST-GISTEMP ensemble members discussed in Section 4.1.

The sampling uncertainty also has temporal persistence due to autocorrelation in the monthly temperature anomaly fields. To conservatively account for this temporal persistence, a random block of length 1–12 months is selected from the empirical reconstruction error fields. This method is an extension of the 12 month persistence of uncertainty used in the HadCRUT5 method (Morice et al., 2020) which reduces artifacts in time series calculated using the uncertainty ensemble. Note that the month of the empirical reconstruction error field is selected to align with the month of the GISTEMP ensemble as the underlying temperature variability and therefore uncertainty varies seasonally.

4.3 Calculation of Global and Large-Scale Series Ensembles

One would naively expect to be able to use the ensemble to calculate the magnitude of uncertainty in a time series such as the global annual mean by simply calculating the time series of interest in each of the 200 GISTEMP uncertainty ensemble members. However, this method does not account for the uncertainty in such series due to grid cells that do not have estimates. To include this uncertainty, an additional 200 member GISTEMP uncertainty ensemble is created that includes this uncertainty by modifying the sampling uncertainty step discussed above. As before, the empirical reconstruction error fields are added to the GHCN-ERSST-GISTEMP ensemble at each monthly time step for grid-cells where GISTEMP is making estimates. However, for grid-cells where GISTEMP does not make estimates due to limited station or ship coverage, the missing value is replaced with the "true" value from ERA5 from the same ERA5 time-step that was used for the empirical reconstruction error field. This method both includes the uncertainty added to large-scale series by not making temperature anomaly estimates at all locations in the interpolation step, as well as includes the correct spatial structure by ensuring this uncertainty comes from the same underlying "true" temperature field that was used to estimate the sampling uncertainty.

This full-coverage uncertainty ensemble fully captures the uncertainty of global and other large scale means. The GISTEMP averaging procedure is applied to each of the 200 members, resulting in a 200 member ensemble each of global, hemispheric, band, and land-only/ocean-only monthly and annual temperature series.

4.4 Decomposition of the LSAT Uncertainty

The full-coverage GISTEMP uncertainty ensemble described in Section 4.3 also allows the decomposition of global annual LSAT uncertainty into sampling and homogenization for comparison with Figure 4 of Lenssen et al. (2019). The global annual LSAT homogenization uncertainty is calculated as the spread of the 200 members land-only global annual mean temperature anomaly from the GHCN-ERSST-GISTEMP, excluding any addition of sampling uncertainty. The global annual LSAT sampling uncertainty is calculated as the spread of 200 members simulated following the full-coverage GISTEMP sampling ensemble detailed above where empirical reconstruction error values are used when GISTEMP provides an estimate and ERA5 values of corresponding ERA5 time-steps are used when GISTEMP does not provide an estimate.

5 Results and Discussion

The global and hemispheric annual mean series are calculated with the GISTEMP uncertainty ensemble by applying the GISTEMP averaging scheme to each of the 200 gridded ensemble members. The ensemble median matches very well with operational GISTEMP for each of these series (Figure 3). The 95% confidence intervals of mean series are constructed as the empirical 95% confidence interval from the 200 annual mean series. The 95% confidence interval of the ensemble mean and hemispheric series covers the operational series at every time point, which along with the near-perfect agreement between the ensemble median and the operational series, validates the GISTEMP uncertainty ensemble's ability to accurately replicate the global mean calculation.

The ensemble estimate of uncertainty in the global annual mean is uniformly lower than the analysis of Lenssen et al. (2019) (Figure 4). This could be due to either lower SST or LSAT uncertainty or a positive correlation between the SST and LSAT, assumed zero in

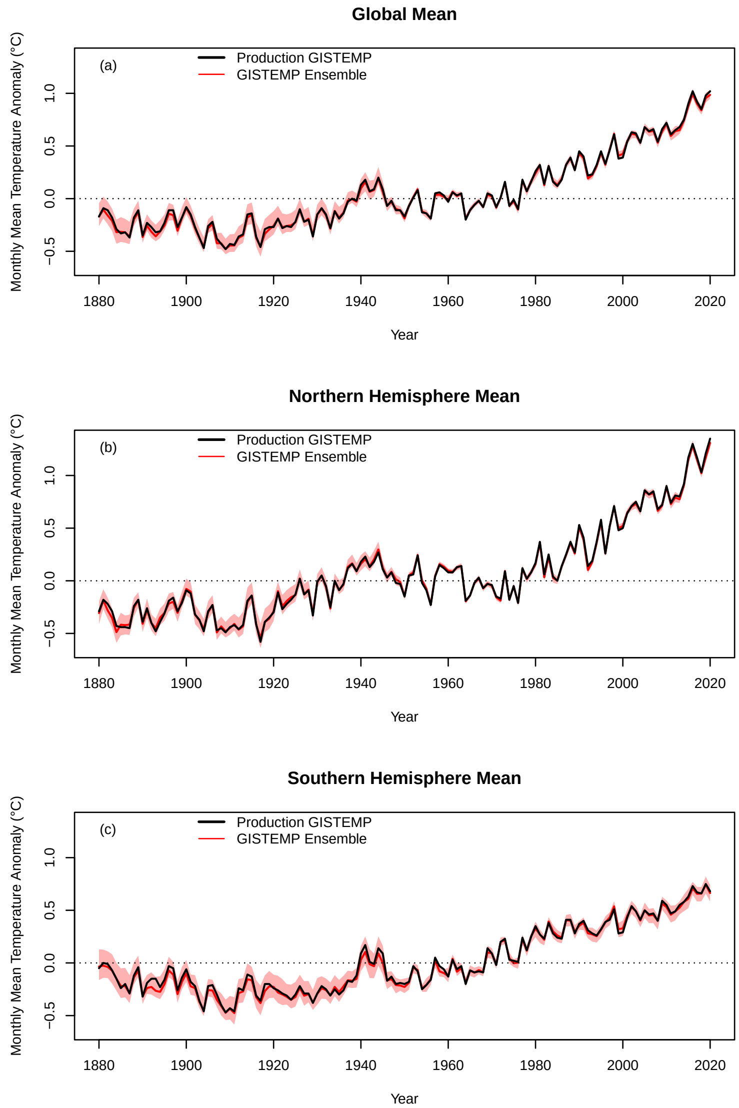


Figure 3. A comparison of the global and hemispheric annual mean series as calculated from operational GISTEMP and the GISTEMP ensemble. The solid red line is the median of the GISTEMP ensemble.

the Lenssen et al. (2019) analysis. The SST uncertainty is quantified by the same ERSSTv5 ensemble used in the first analysis and can't be the source of the discrepancy. Decomposing the ensemble global annual land surface uncertainty into its two components suggests that much smaller homogenization uncertainty is the primary reason for the smaller global uncertainty estimate presented here (Figure 5). This is expected as the homogenization uncertainty in the global mean in Lenssen et al. (2019) was quantified on a $5° \times 5°$ grid, but did not include any additional smoothing or corrections. However, the GISTEMP averaging

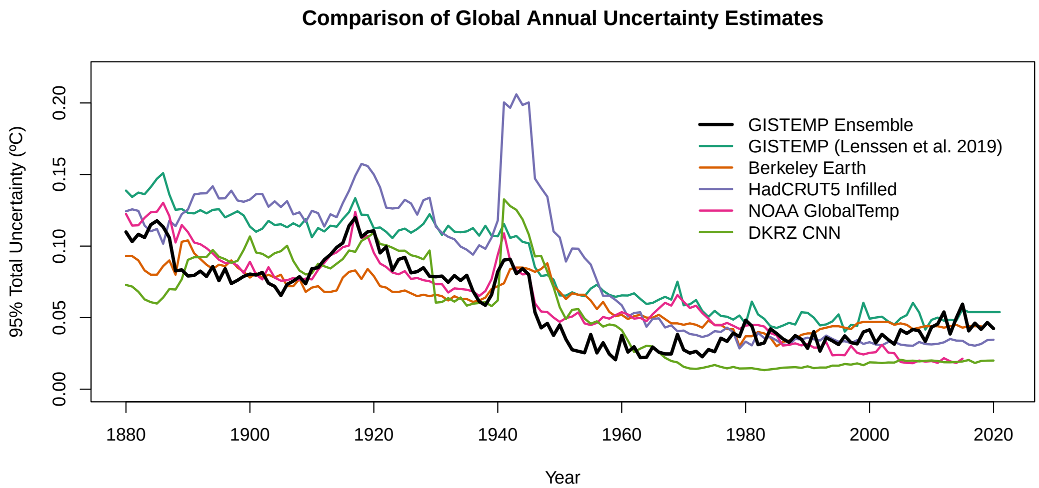


Figure 4. The global annual mean 95% confidence intervals for the new GISTEMP ensemble, the same calculation as performed in Lenssen et al. (2019), and the two products that publish operational confidence intervals.

method includes all station information within 1,200 km when estimating the temperature anomaly of a gridbox as well as corrects all station records to account for mean-shift biases by comparing nearby station records over complete time periods (Hansen & Lebedeff, 1987). This additional homogenization step dramatically reduces the homogenization uncertainty, particularly in the early part of the record.

The estimate of the uncertainty in the global annual mean temperature anomaly from the GISTEMP ensemble is close to the other major global analyses that publish uncertainty estimates (Figure 4): HadCRUT5 (Morice et al., 2020), NOAA GlobalTemp (Huang et al., 2020), Berkeley Earth (Rohde & Hausfather, 2020), and DKRZ-DNN (Kadow et al., 2020). The GISTEMP Ensemble estimate closely mirrors that of NOAA GlobalTemp as expected due to the shared source data uncertainty quantification between the two products. The largest deviation in products is the very large uncertainty in HadCRUT5 during the WWII period in the early 1940s. This large uncertainty is a conservative estimate of the large biases found in SST data during this period that have not yet been incorporated in operational ship record databases (Chan et al., 2019).

The GISTEMP ensemble provides uncertainty at monthly temporal resolution, allowing analyses of uncertainty in monthly mean temperature change. Monthly uncertainties

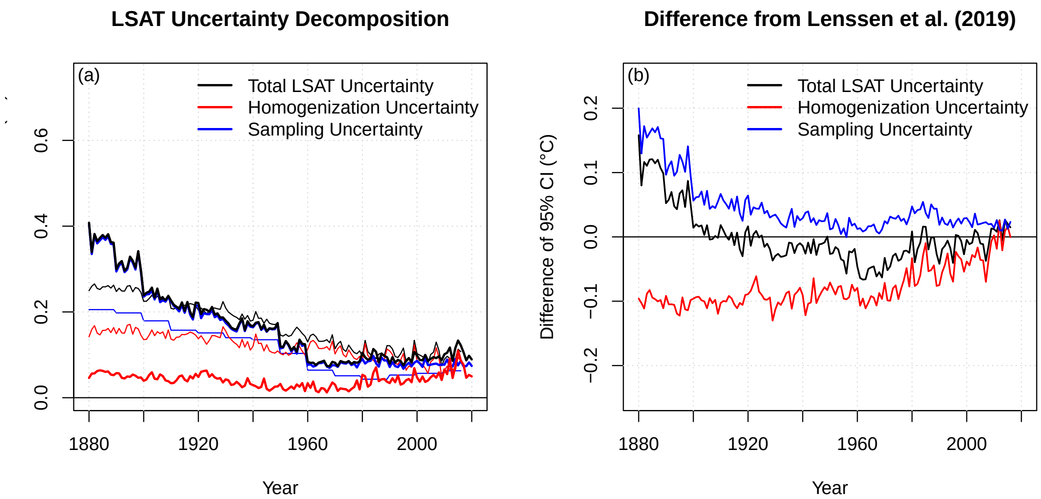


Figure 5. (a) The annual LSAT uncertainty (2σ) decomposed into the sampling and homogenization uncertainty components. The dark lines show the calculation from the ensemble analysis and the thin lines show the Lenssen et al. (2019) calculation. (b) The difference of the LSAT uncertainties between the ensemble and the global analysis of Lenssen et al. (2019).

are slightly smaller in the NH winter (0.06ºC) than the summer months (0.08ºC), because of the greater uncertainty in SH winters – particularly in Antarctica which has the worst coverage and largest homogenization uncertainty. We see that the uncertainty in the January and July global mean temperature series is again much smaller than the warming signal (Figure 6). Using the July 2020 uncertainty estimate as the approximate uncertainty of the GISTEMP July 2023 anomaly of 1.18 $°C$, we conclude that July 2023 is the warmest global month on record with nearly 100 % certainty.

Decomposing the annual LSAT mean into the homogenization and sampling components reveals that the LSAT uncertainty, at least globally, it dominated by sampling uncertainty for the majority of the record (Figure 5). In the Lenssen et al. (2019), homogenization and sampling uncertainty in the global LSAT mean were calculated separately and then combined assuming independence. However, it is shown here that the additional homogenization done in GISTEMP when averaging multiple station records reduces homogenization uncertainty to levels substantially smaller than raw GHCN, at least at the global scale.

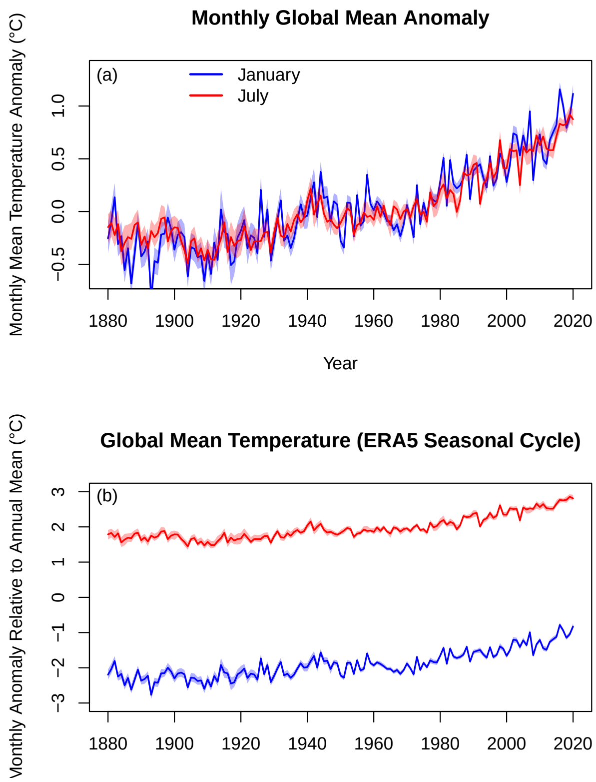


Figure 6. The global mean January and July time series and 95% confidence intervals for (a) the global mean temperature anomalies with respect to a 1950-1979 climatology and (b) the global mean anomaly scaled by the seasonal cycle using ERA5 to estimate the seasonal cycle of global temperature.

Sampling uncertainty is the dominant source of uncertainty for nearly the entire land surface (Figure 7), in agreement with the global LSAT decomposition. The sampling uncertainties are particularly dominant for regions with dense station coverage where homogenization uncertainties are nearly zero. The only area where homogenization uncertainty dominates is the central Amazon and parts of Antarctica, both regions with known major inhomogeneities and very few nearby station observations to use in correction methods.

Looking at the GISTEMP latitudinal band mean estimates from the uncertainty ensemble, large uncertainty in the polar regions appears to be driving the global uncertainty (Figure 8). This uncertainty is driven by land and sea ice regions near the poles (Figure 9). Again, there is very good agreement between the operational and ensemble GISTEMP with the ensemble confidence interval always covering the operational series. In general, the

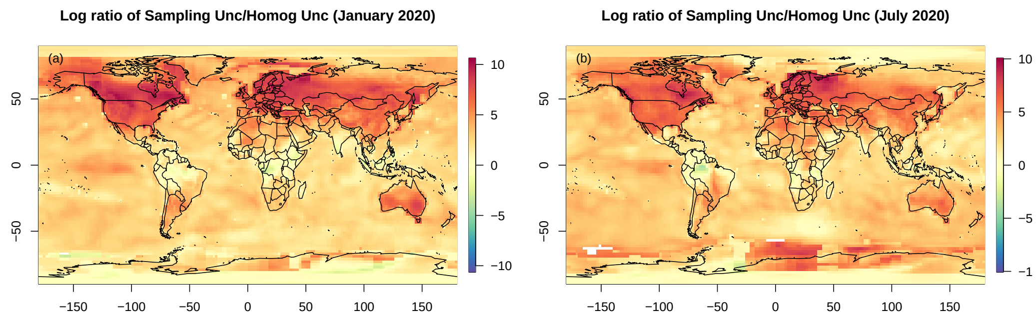


Figure 7. The log ratio of sampling and GHCN uncertainty for (a) January 2020 and (b) July 2020. Red regions show where sampling uncertainty dominates and blue regions show where homogenization uncertainty dominates.

land surface has greater uncertainty than the ocean at the monthly scale, except during the observationally sparse 1940s (Figure 9).

The GISTEMP uncertainty ensemble allows easy investigation of the spatial pattern of uncertainty at any month in the record. the uncertainty for January 1910, 1940, and 1970 look broadly similar with generally greater uncertainty over land than the ocean, in regions of high temperature variability such as the (particularly winter) mid and high latitudes, and in regions with sparse station coverage (Figure 9).

The availability of a few uncertainty ensembles allows comparison of the GISTEMP ensemble with estimates from HadCRUT5, NOAA GlobalTemp, and DKRZ-DNN (Figure 10). Note that each ensemble member from all products have been regridded to the HadCRUT5 $5° \times 5°$ grid before calculating the standard deviation to facilitate comparison. Comparing GISTEMP and HadCRUT5, it is immediately evident that GISTEMP generally estimates larger uncertainty over land, particularly in the polar regions. This result is expected as HadCRUT5 estimates the sampling and interpolation uncertainty through posterior draws from the stationary Gaussian Process used to interpolate data to grid boxes that don't have reporting stations. However, this stationary Gaussian Process uses one set of parameters fit to the entire entire globe and, as such, overestimates the signal-to-noise ratio over the polar regions where there is both less information of the true spatial field, greater underlying variance in the true temperature field, and a shorter spatial autocorrelation (Morice et al., 2020). By using the empirical estimates of the sampling uncertainty in the GISTEMP ensemble, this non-stationarity is avoided and provides a better estimate of the point-wise uncertainty. However, these apparently large differences do not imply large differences in global annual mean uncertainty (Figure 4) due to the relatively small area of the polar regions.

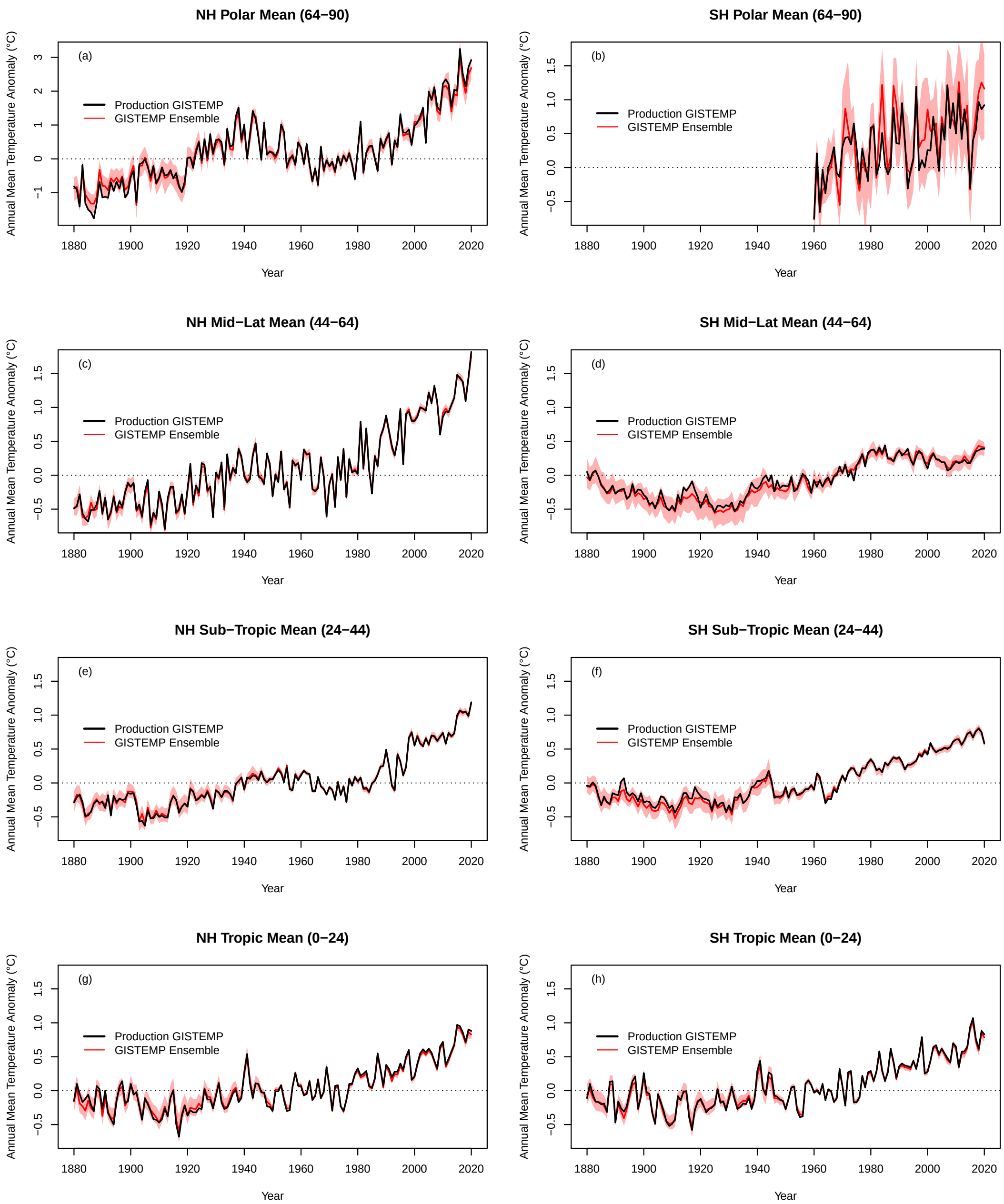


Figure 8. A comparison of the annual mean series from the 8 GISTEMP latitudinal bands as calculated from operational GISTEMP and the GISTEMP uncertainty ensemble. The solid red line is the median of the GISTEMP ensemble. Note the different y-scale on the top-left NH Polar plot.

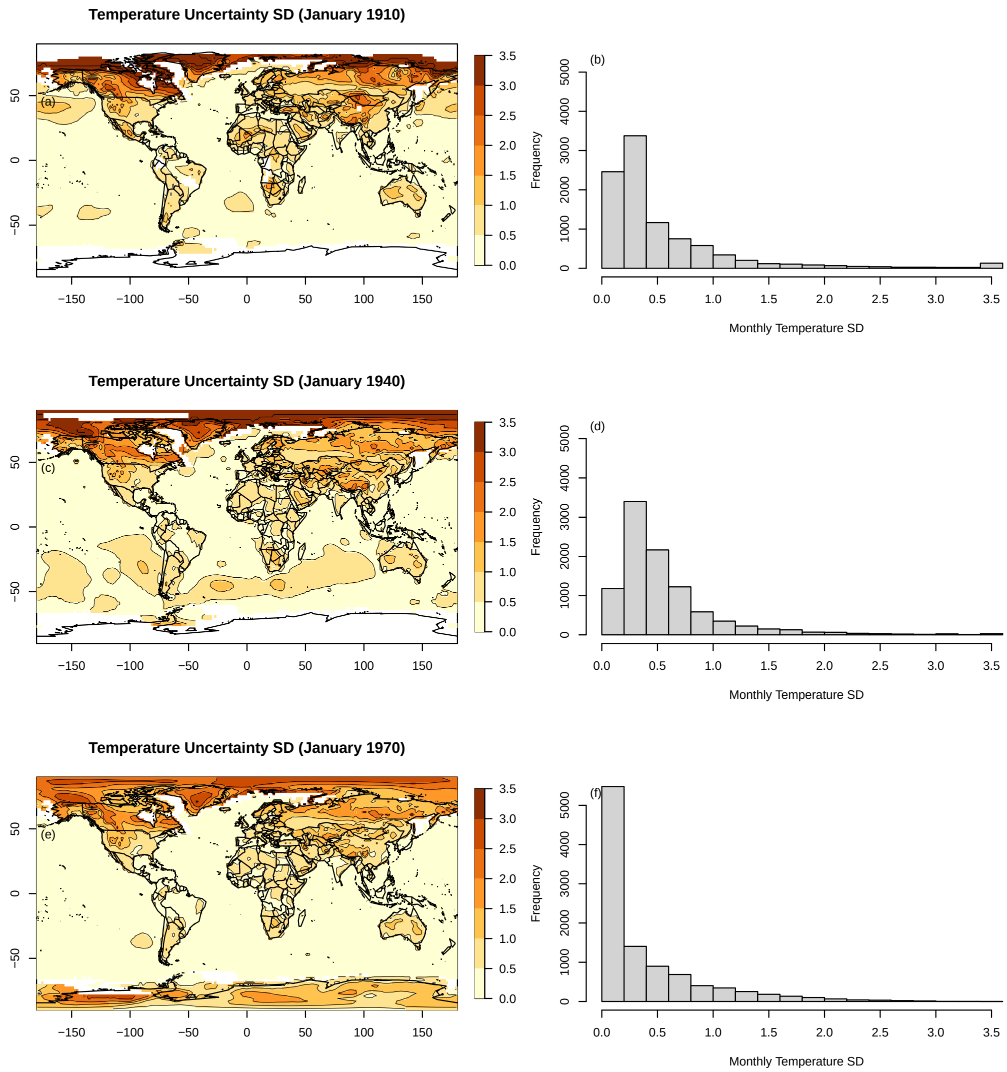


Figure 9. The standard deviation of the GISTEMP uncertainty ensemble for three monthly fields. The corresponding histogram to each field is shown to the right. The visualization has been capped at a standard deviation of 3.5 to avoid the very large Antarctic uncertainty dominating the maps.

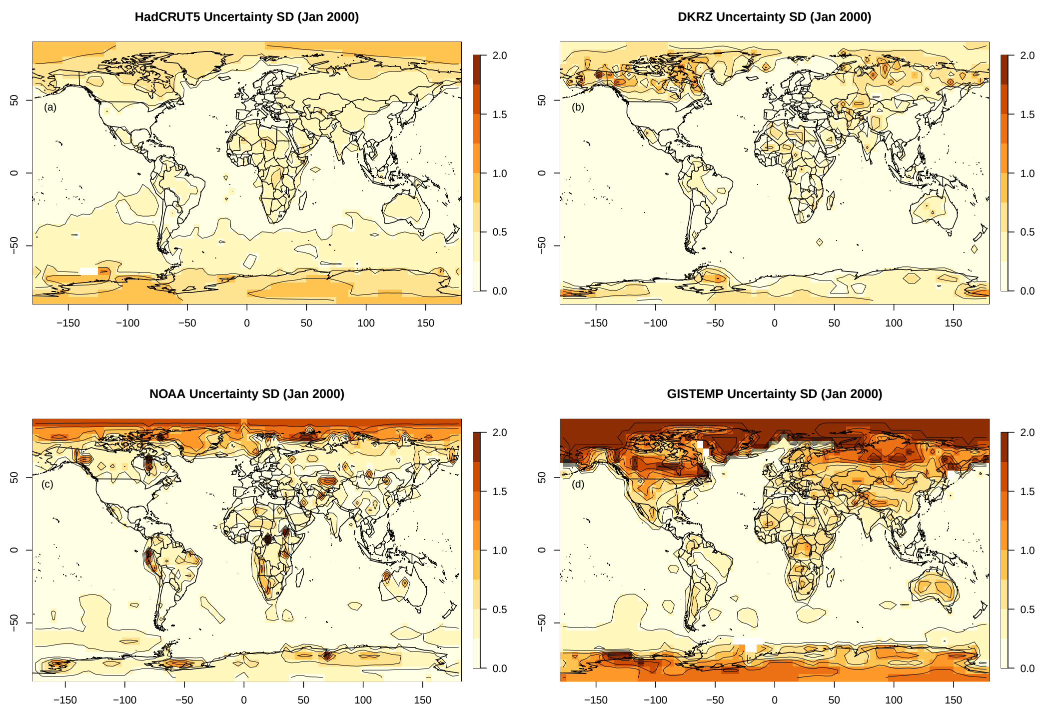


Figure 10. Comparison of all global gridded ensemble uncertainty products total uncertainty for January 2000

6 Discussion and Conclusions

Here, an uncertainty ensemble for the GISTEMP temperature product has been presented and analyzed. Accounting for all sources of uncertainty at the monthly level increases enables inclusion of historical temperature uncertainty in future studies. The median estimates from the GISTEMP uncertainty ensemble agree very well with operational GISTEMP and the resulting global mean uncertainty agrees with the calculation of Lenssen et al. (2019). This work is a major step forward in the GISTEMP uncertainty model, enabling the inclusion of observational uncertainty in studies on historical global change.

Uncertainty ensembles of gridded products and key time series make including observational uncertainty in subsequent analyses simple. Given an analysis developed using the operational version of GISTEMP, the only additional step is to rerun the analysis on each member of the uncertainty ensemble. Then, results can be summarized using the mean or median estimate of the result of interest as well as empirical confidence intervals. Including uncertainty in surface temperature data is particularly important in regions with high uncertainty such as the polar regions as well as areas with lower forced signals, such as investigations of warming over the eastern tropical pacific and the southern ocean.

This release of an additional uncertainty ensemble form a major global temperature product highlights the need for post-analysis of the available surface temperature uncertainty products. Critical future work includes a comprehensive assessment of the similarities and differences between the 4 temperature ensembles presented here. Such an analysis should investigate both the statistics of the ensembles as well as the results from using these 4 ensembles to answer critical questions about our climate system such as the rate of arctic warming (Rantanen et al., 2022) or the time of emergence of forced signals.

It is the authors' hope that the release of the GISTEMP uncertainty ensemble, alongside the already existing HadCRUT5, NOAA GlobalTemp, and DKRZ DNN uncertainty ensembles, will prompt the community to incorporate observational uncertainty in future studies involving historical surface temperature data whenever possible.

7 Open Research

The final gridded and series ensembles are available on the GISTEMP website `https://data.giss.nasa.gov/gistemp/` as well as AWS cloud storage. The operational GISTEMP code used to create the GHCN-ERSST-GISTEMP is also available on the GISTEMP website. All of the R analysis code is available on github at `https://github.com/nlenssen/gistempAWS`. The full code base, source data and intermediate analyses are available on an AWS instance upon request.

Acknowledgments

The GISTEMP analysis is funded by the NASA Modeling, Analysis and Prediction program. N.L. was also funded from the National Science Foundation Graduate Research Fellowship (NSF DGE 16-44869) and the Food and Agriculture Organization (FAO) of the United Nations. Thanks to Duo Chan, Dorit Hammerling, Boyin Huang, Chris Kadow, Colin Morice, Doug Nychka, and Robert Rhode for helpful discussions. Thanks to Hoot Thompson, Garrison Vaughan, Dan'l Pierce, and Dan Duffy at the NASA Center for Climate Simulation (NCCS) for supercomputing and cloud computing support.

References

Barnes, N., & Jones, D. (2011). Clear Climate Code: Rewriting legacy science software for clarity. *IEEE Software*, *28*(6), 36–42. doi: 10.1109/ms.2011.113

Carleton, T. A., Jina, A., Delgado, M. T., Greenstone, M., Houser, T., Hsiang, S. M., ... others (2020). *Valuing the global mortality consequences of climate change accounting for adaptation costs and benefits* (Tech. Rep.). Cambridge, MA: National Bureau of Economic Research.

Chan, D., Kent, E. C., Berry, D. I., & Huybers, P. (2019). Correcting datasets leads to more homogeneous early-twentieth-century sea surface warming. *Nature*, *571*(7765), 393–397.

Cowtan, K., Jacobs, P., Thorne, P., & Wilkinson, R. (2018, 07). Statistical analysis of coverage error in simple global temperature estimators. *Dynamics and Statistics of the Climate System*, *3*(1), dzy003. Retrieved from `https://doi.org/10.1093/climsys/dzy003` doi: 10.1093/climsys/dzy003

Danabasoglu, G., Lamarque, J.-F., Bacmeister, J., Bailey, D., DuVivier, A., Edwards, J., ... others (2020). The community earth system model version 2 (cesm2). *Journal of Advances in Modeling Earth Systems*, *12*(2), e2019MS001916.

Hansen, J., & Lebedeff, S. (1987). Global trends of measured surface air temperature. *Journal of Geophysical Research: Atmospheres*, *92*(D11), 13345–13372. Retrieved from `http://dx.doi.org/10.1029/JD092iD11p13345` doi: 10.1029/JD092iD11p13345

Hansen, J., Ruedy, R., Sato, M., & Lo, K. (2010). Global surface temperature change. *Reviews of Geophysics*, *48*(4). Retrieved from `http://dx.doi.org/10.1029/2010RG000345` (RG4004) doi: 10.1029/2010RG000345

Hausfather, Z., Drake, H. F., Abbott, T., & Schmidt, G. A. (2020). Evaluating the performance of past climate model projections. *Geophysical Research Letters*, *47*(1), e2019GL085378.

Hausfather, Z., Menne, M. J., Williams, C. N., Masters, T., Broberg, R., & Jones, D. (2013). Quantifying the effect of urbanization on u.s. historical climatology network temperature records. *Journal of Geophysical Research: Atmospheres*, *118*(2), 481–494. doi: 10.1029/2012jd018509

Held, I., Guo, H., Adcroft, A., Dunne, J., Horowitz, L., Krasting, J., ... others (2019). Structure and performance of gfdl's cm4. 0 climate model. *Journal of Advances in Modeling Earth Systems*, *11*(11), 3691–3727.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., ... Thépaut, J.-N. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999-2049. Retrieved from `https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.3803` doi: https://doi.org/10.1002/qj.3803

Huang, B., Menne, M. J., Boyer, T., Freeman, E., Gleason, B. E., Lawrimore, J. H., ... Zhang, H.-M. (2020). Uncertainty estimates for sea surface temperature and land surface air temperature in noaaglobaltemp version 5. *Journal of Climate*, *33*(4), 1351 - 1379. Retrieved from `https://journals.ametsoc.org/view/journals/clim/33/4/jcli-d-19-0395.1.xml` doi: 10.1175/JCLI-D-19-0395.1

Huang, B., Thorne, P. W., Banzon, V. F., Boyer, T., Chepurin, G., Lawrimore, J. H., ... Zhang, H.-M. (2017). Extended reconstructed sea surface temperature, version 5 (ersstv5): Upgrades, validations, and intercomparisons. *Journal of Climate*, *30*(20), 8179 - 8205. Retrieved from `https://journals.ametsoc.org/view/journals/clim/30/20/jcli-d-16-0836.1.xml` doi: 10.1175/JCLI-D-16-0836.1

Huang, B., Thorne, P. W., Smith, T. M., Liu, W., Lawrimore, J., Banzon, V. F., ... Menne, M. (2016). Further exploring and quantifying uncertainties for extended reconstructed sea surface temperature (ersst) version 4 (v4). *Journal of Climate*, *29*(9), 3119–3142.

Kadow, C., Hall, D. M., & Ulbrich, U. (2020). Artificial intelligence reconstructs missing climate information. *Nature Geoscience*, *13*(6), 408–413.

Lenssen, N. J. L., Schmidt, G. A., Hansen, J. E., Menne, M. J., Persin, A., Ruedy, R., & Zyss, D. (2019). Improvements in the GISTEMP uncertainty model. *Journal of Geophysical Research: Atmospheres*, *124*(12), 6307-6326. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018JD029522 doi: 10.1029/2018JD029522

Liu, W., Huang, B., Thorne, P. W., Banzon, V. F., Zhang, H.-M., Freeman, E., … Woodruff, S. D. (2015). Extended reconstructed sea surface temperature version 4 (ersst.v4): Part ii. parametric and structural uncertainty estimations. *Journal of Climate*, *28*(3), 931 - 951. Retrieved from https://journals.ametsoc.org/view/journals/clim/28/3/jcli-d-14-00007.1.xml doi: 10.1175/JCLI-D-14-00007.1

Menne, M. J., Williams, C. N., Gleason, B. E., Rennie, J. J., & Lawrimore, J. H. (2018, sep). The Global Historical Climatology Network monthly temperature dataset, version 4. *Journal of Climate*. Retrieved from https://doi.org/10.1175/jcli-d-18-0094.1 doi: 10.1175/jcli-d-18-0094.1

Morice, C. P., Kennedy, J. J., Rayner, N. A., & Jones, P. D. (2012, 4). Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The HadCRUT4 data set. *Journal of Geophysical Research: Atmospheres*, *117*(D8). Retrieved from http://doi.wiley.com/10.1029/2011JD017187 doi: 10.1029/2011JD017187

Morice, C. P., Kennedy, J. J., Rayner, N. A., Winn, J. P., Hogan, E., Killick, R. E., … Simpson, I. R. (2020). An updated assessment of near-surface temperature change from 1850: the HadCRUT5 dataset. *Journal of Geophysical Research*.

Myhre, G., Alterskjær, K., Stjern, C. W., Hodnebrog, Ø., Marelle, L., Samset, B. H., … others (2019). Frequency of extreme precipitation increases extensively with event rareness under global warming. *Scientific reports*, *9*(1), 1–10.

Notz, D., & SIMIP Community. (2020). Arctic sea ice in cmip6. *Geophysical Research Letters*, *47*(10), e2019GL086749.

R Core Team. (2020). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from https://www.R-project.org/

Rantanen, M., Karpechko, A. Y., Lipponen, A., Nordling, K., Hyvärinen, O., Ruosteenoja, K., … Laaksonen, A. (2022). The arctic has warmed nearly four times faster than the globe since 1979. *Communications Earth & Environment*, *3*(1), 168.

Rohde, R. A., & Hausfather, Z. (2020). The berkeley earth land/ocean temperature record. *Earth System Science Data*, *12*(4), 3469–3479. Retrieved from `https://essd.copernicus.org/articles/12/3469/2020/` doi: 10.5194/essd-12-3469-2020

Swart, N. C., Cole, J. N., Kharin, V. V., Lazare, M., Scinocca, J. F., Gillett, N. P., ... others (2019). The canadian earth system model version 5 (canesm5. 0.3). *Geoscientific Model Development*, *12*(11), 4823–4873.

Tokarska, K. B., Stolpe, M. B., Sippel, S., Fischer, E. M., Smith, C. J., Lehner, F., & Knutti, R. (2020). Past warming trend constrains future warming in cmip6 models. *Science advances*, *6*(12), eaaz9549.

Figure 1.

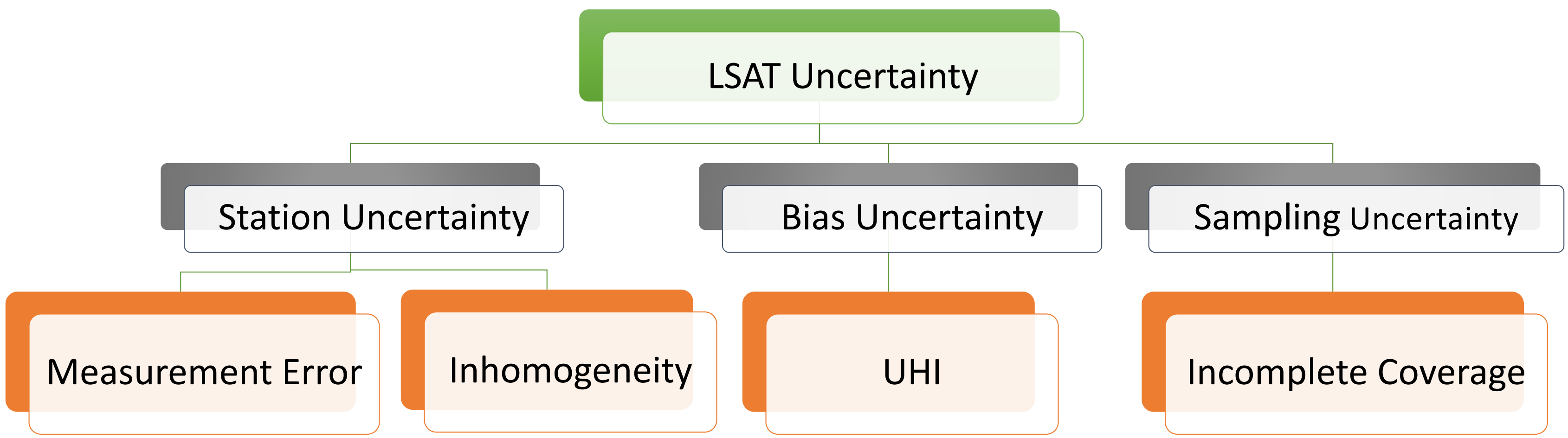

LSAT Uncertainty
Station Uncertainty
Bias Uncertainty
Sampling Uncertainty
Measurement Error
Inhomogeneity
UHI
Incomplete Coverage

Figure 2.

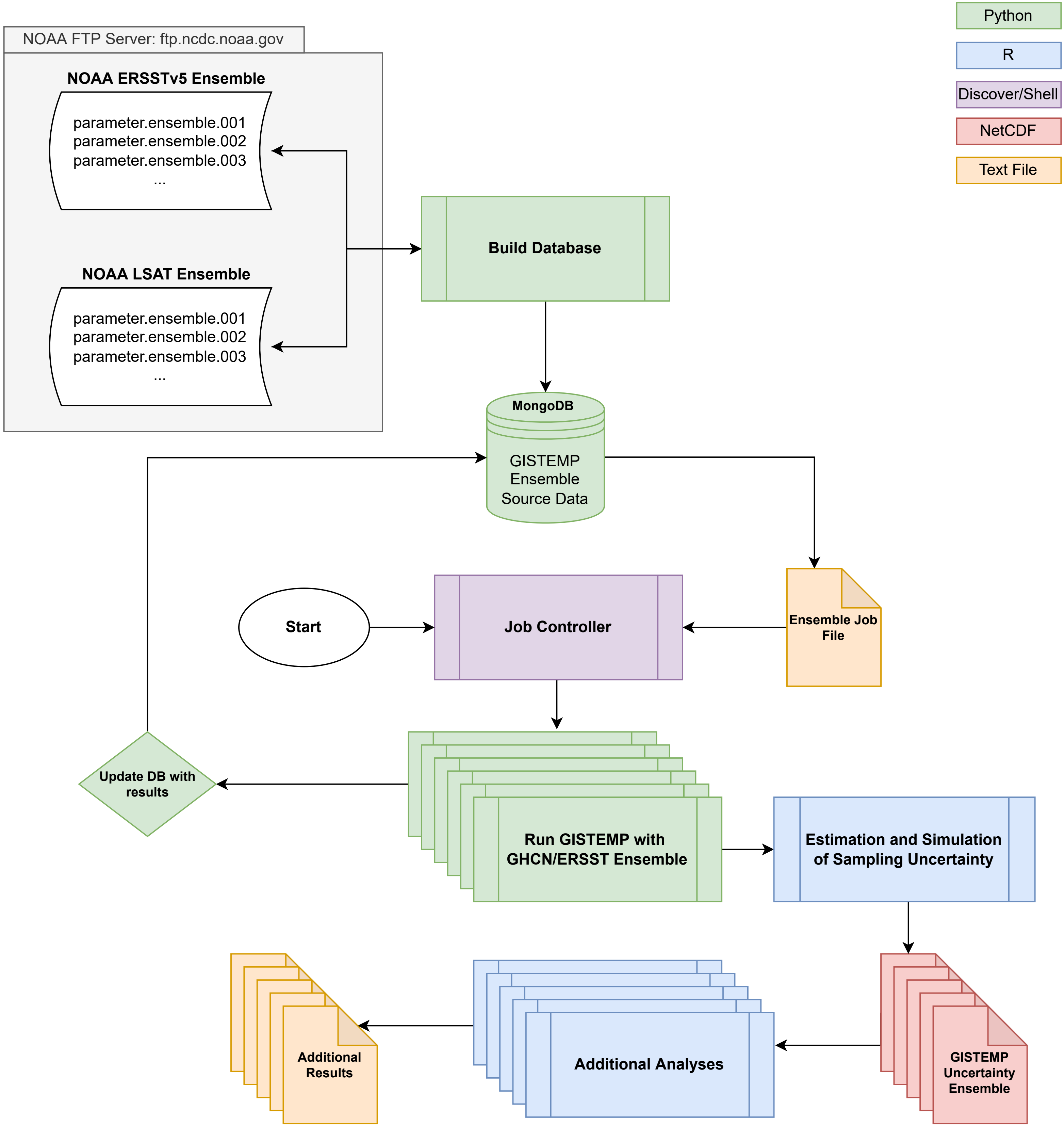

Python
R
Discover/Shell
NetCDF
Text File
NOAA FTP Server: ftp.ncdc.noaa.gov
NOAA ERSSTv5 Ensemble
parameter.ensemble.001
parameter.ensemble.002
parameter.ensemble.003
...
NOAA LSAT Ensemble
parameter.ensemble.001
parameter.ensemble.002
parameter.ensemble.003
...
Build Database
MongoDB
GISTEMP Ensemble Source Data
Start
Job Controller
Ensemble Job File
Update DB with results
Run GISTEMP with GHCN/ERSST Ensemble
Estimation and Simulation of Sampling Uncertainty
GISTEMP Uncertainty Ensemble
Additional Analyses
Additional Results

Figure 3.

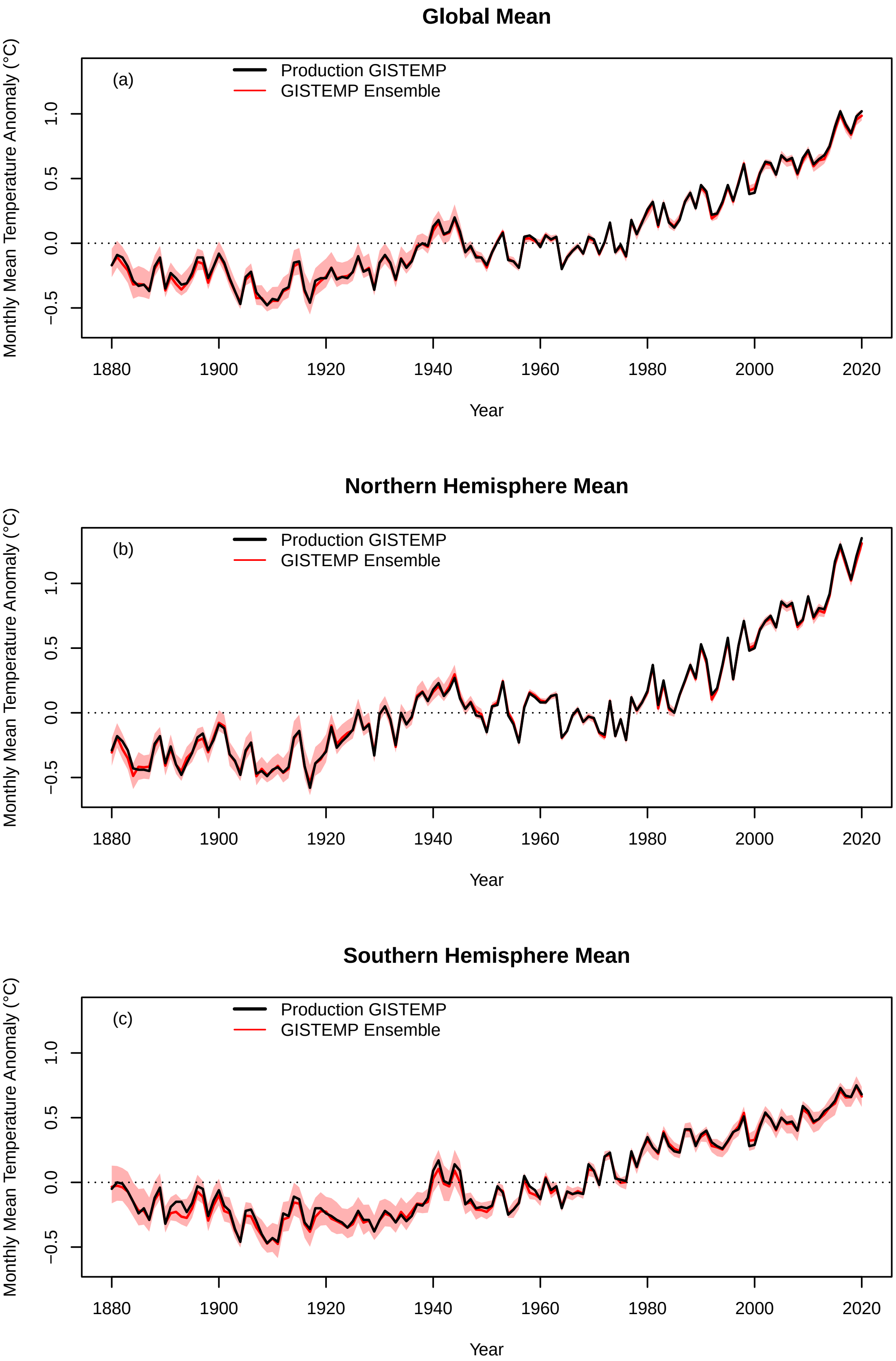
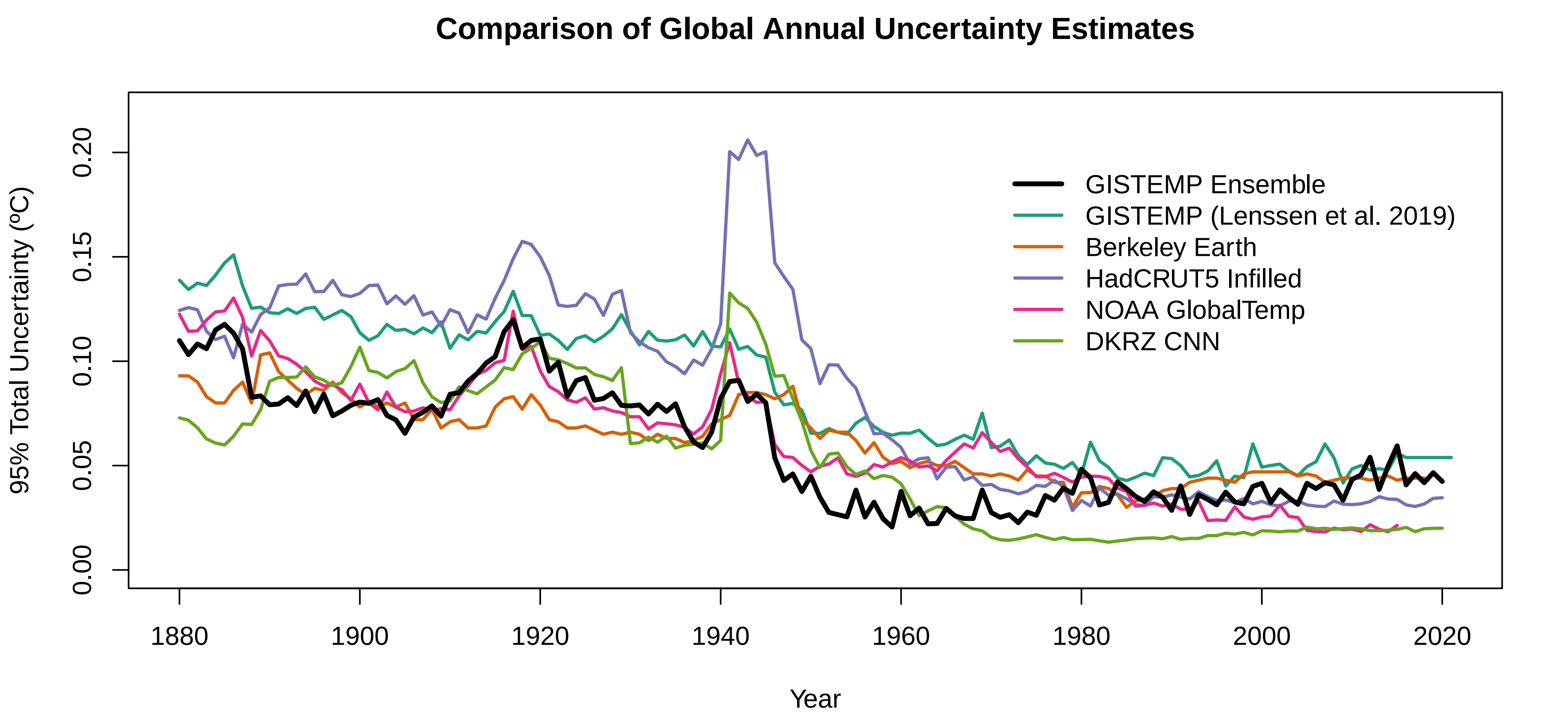
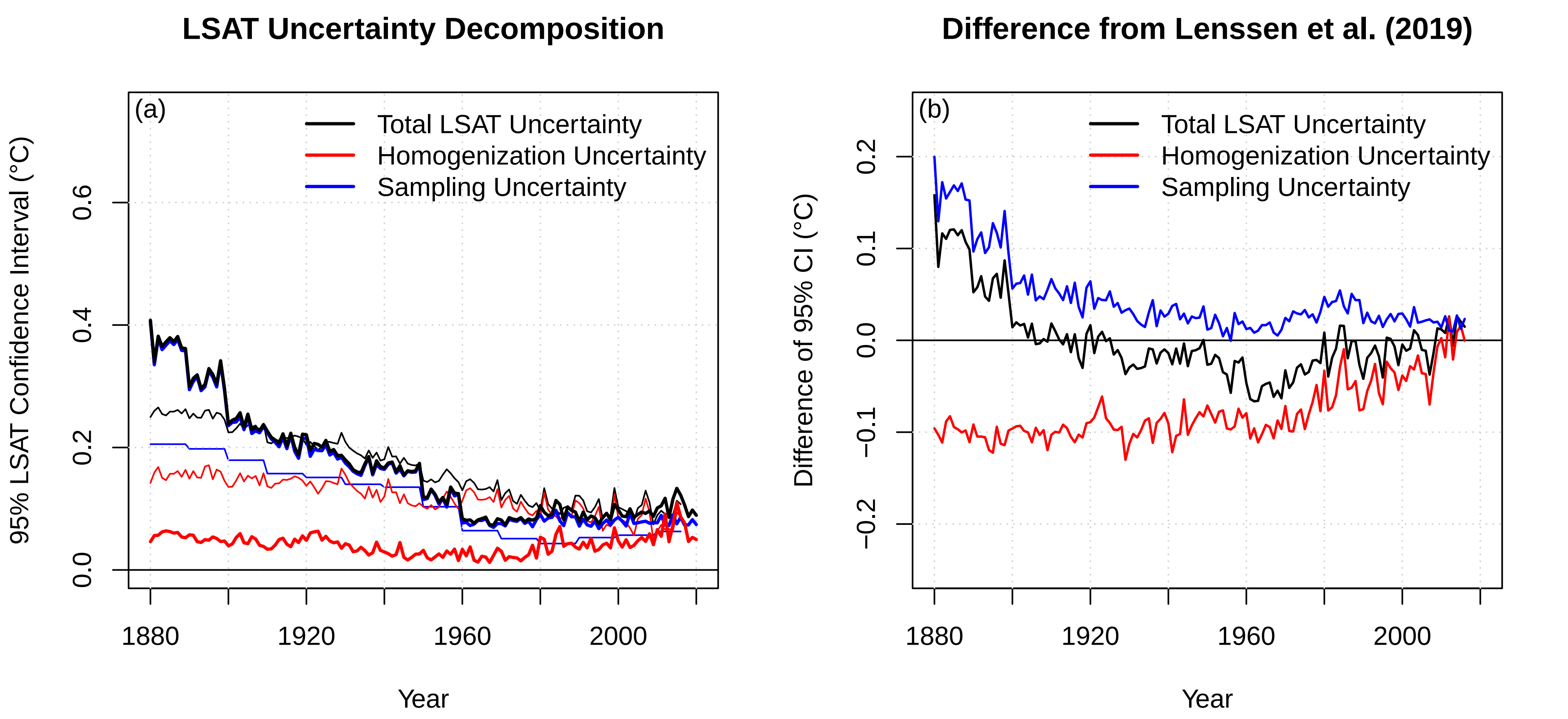

Global Mean
(a)
Production GISTEMP
GISTEMP Ensemble
Monthly Mean Temperature Anomaly (°C)
−0.5
0.0
0.5
1.0
1880
1900
1920
1940
1960
1980
2000
2020
Year
Northern Hemisphere Mean
(b)
Production GISTEMP
GISTEMP Ensemble
Monthly Mean Temperature Anomaly (°C)
−0.5
0.0
0.5
1.0
1880
1900
1920
1940
1960
1980
2000
2020
Year
Southern Hemisphere Mean
(c)
Production GISTEMP
GISTEMP Ensemble
Monthly Mean Temperature Anomaly (°C)
−0.5
0.0
0.5
1.0
1880
1900
1920
1940
1960
1980
2000
2020
Year

Figure 4.



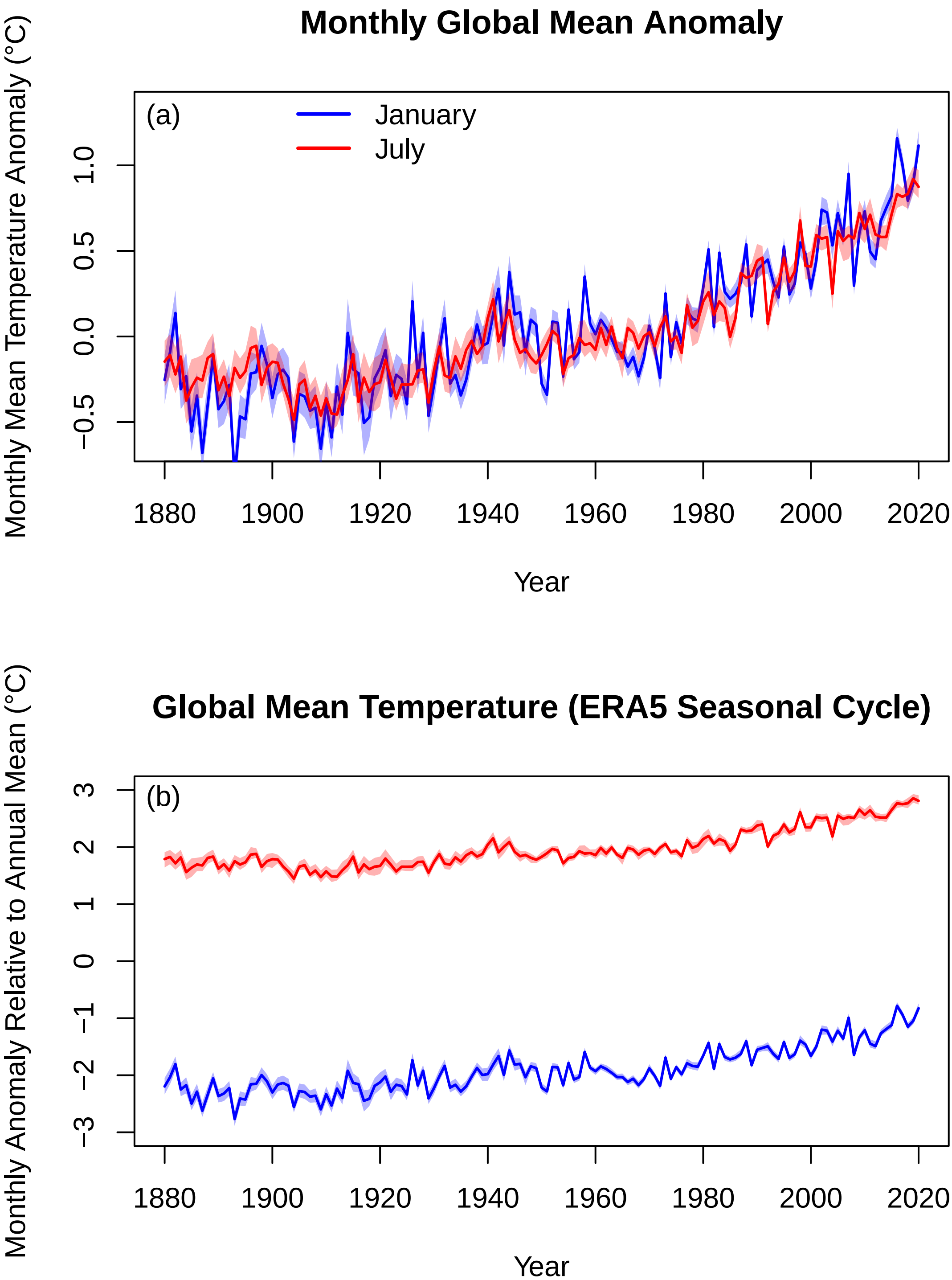
Comparison of Global Annual Uncertainty Estimates
95% Total Uncertainty (ºC)
0.00
0.05
0.10
0.15
0.20
1880
1900
1920
1940
1960
1980
2000
2020
Year
GISTEMP Ensemble
GISTEMP (Lenssen et al. 2019)
Berkeley Earth
HadCRUT5 Infilled
NOAA GlobalTemp
DKRZ CNN

Figure 5.



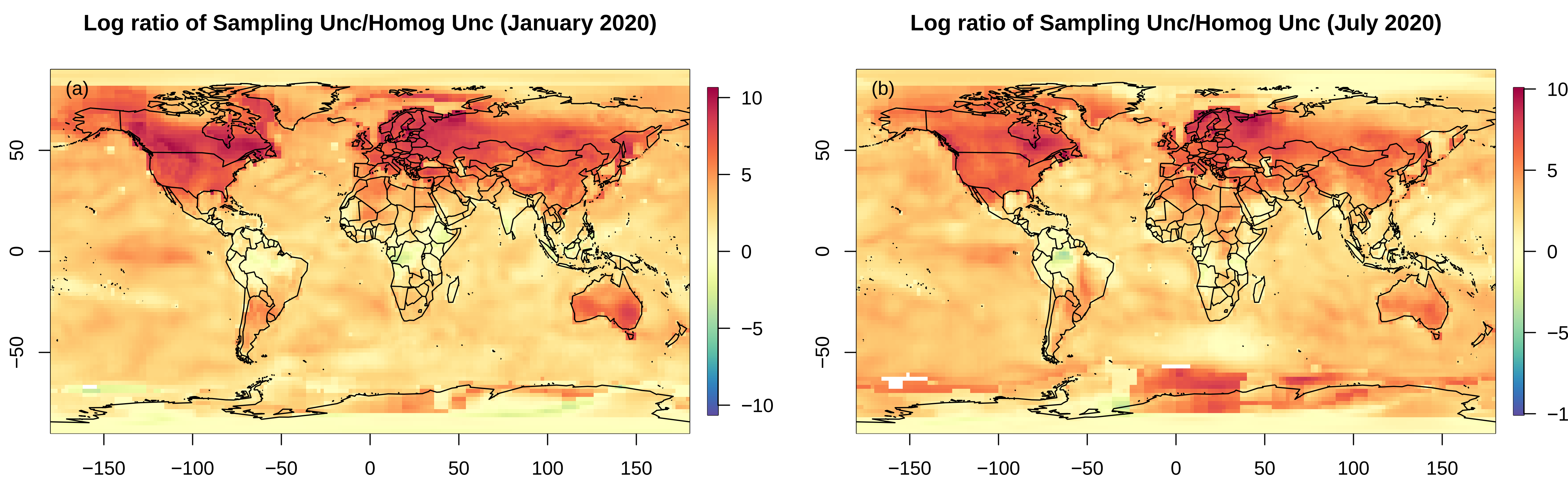
LSAT Uncertainty Decomposition
(a)
Total LSAT Uncertainty
Homogenization Uncertainty
Sampling Uncertainty
95% LSAT Confidence Interval (°C)
0.0
0.2
0.4
0.6
1880
1920
1960
2000
Year
Difference from Lenssen et al. (2019)
(b)
Total LSAT Uncertainty
Homogenization Uncertainty
Sampling Uncertainty
Difference of 95% CI (°C)
−0.2
−0.1
0.0
0.1
0.2
1880
1920
1960
2000
Year

Figure 6.



Monthly Global Mean Anomaly
(a)
January
July
Monthly Mean Temperature Anomaly (°C)
−0.5
0.0
0.5
1.0
1880
1900
1920
1940
1960
1980
2000
2020
Year
Global Mean Temperature (ERA5 Seasonal Cycle)
(b)
Monthly Anomaly Relative to Annual Mean (°C)
−3
−2
−1
0
1
2
3
1880
1900
1920
1940
1960
1980
2000
2020
Year

Figure 7.



Log ratio of Sampling Unc/Homog Unc (January 2020)
(a)
Log ratio of Sampling Unc/Homog Unc (July 2020)
(b)

Figure 8.

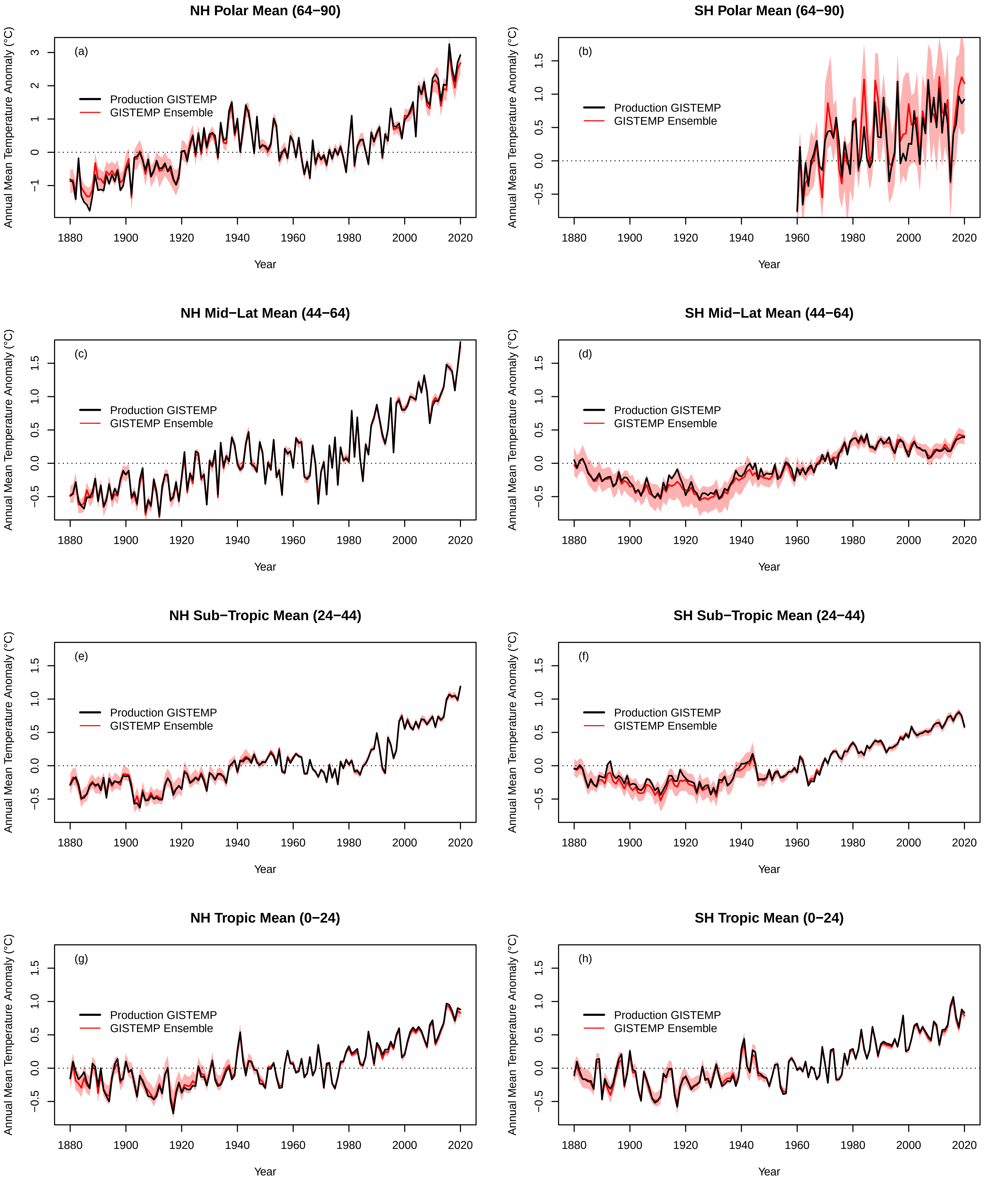

NH Polar Mean (64−90)
(a)
SH Polar Mean (64−90)
(b)
NH Mid−Lat Mean (44−64)
(c)
SH Mid−Lat Mean (44−64)
(d)
NH Sub−Tropic Mean (24−44)
(e)
SH Sub−Tropic Mean (24−44)
(f)
NH Tropic Mean (0−24)
(g)
SH Tropic Mean (0−24)
(h)
Production GISTEMP
GISTEMP Ensemble
Annual Mean Temperature Anomaly (°C)
Year
1880
1900
1920
1940
1960
1980
2000
2020

Figure 9.

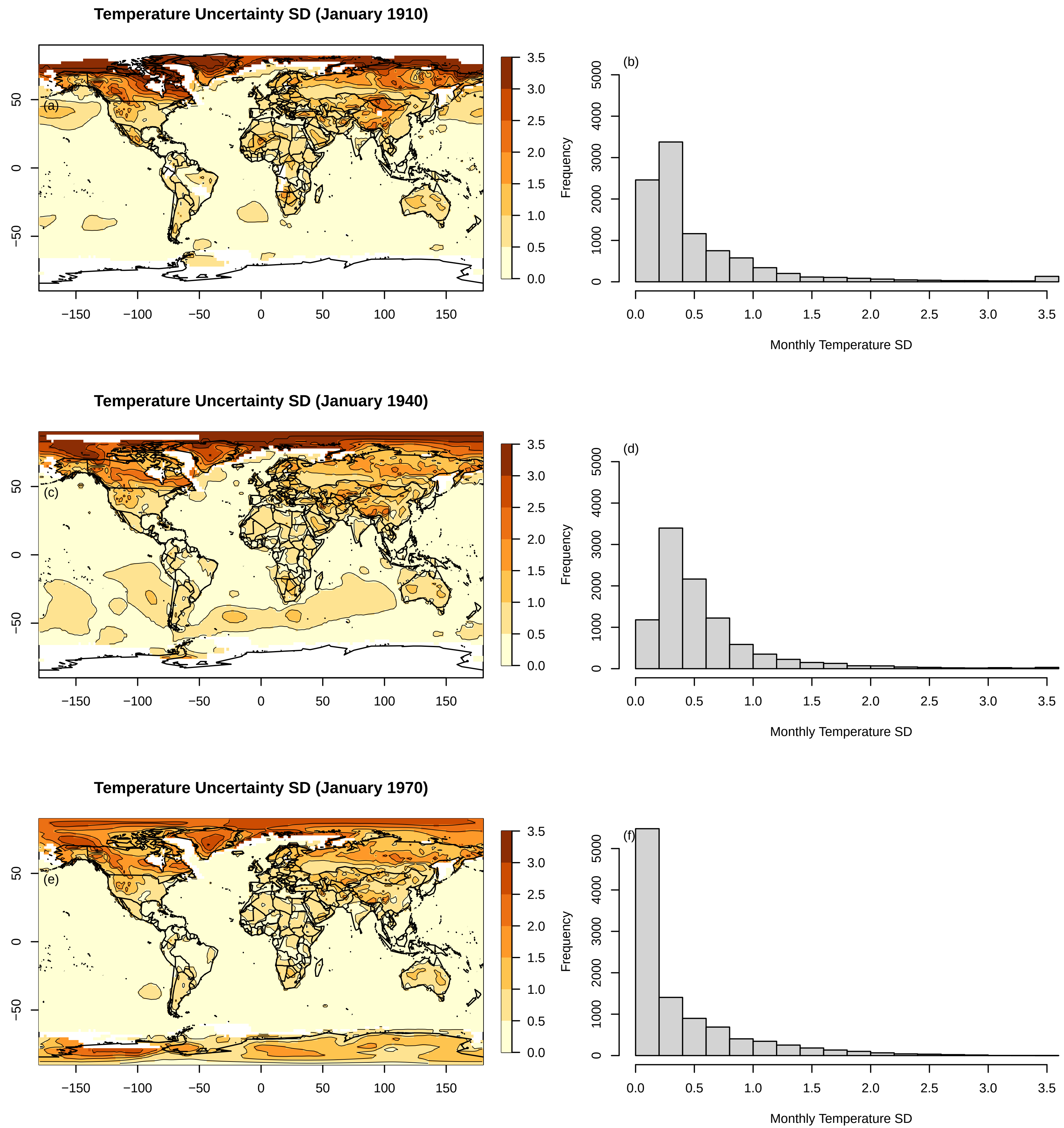
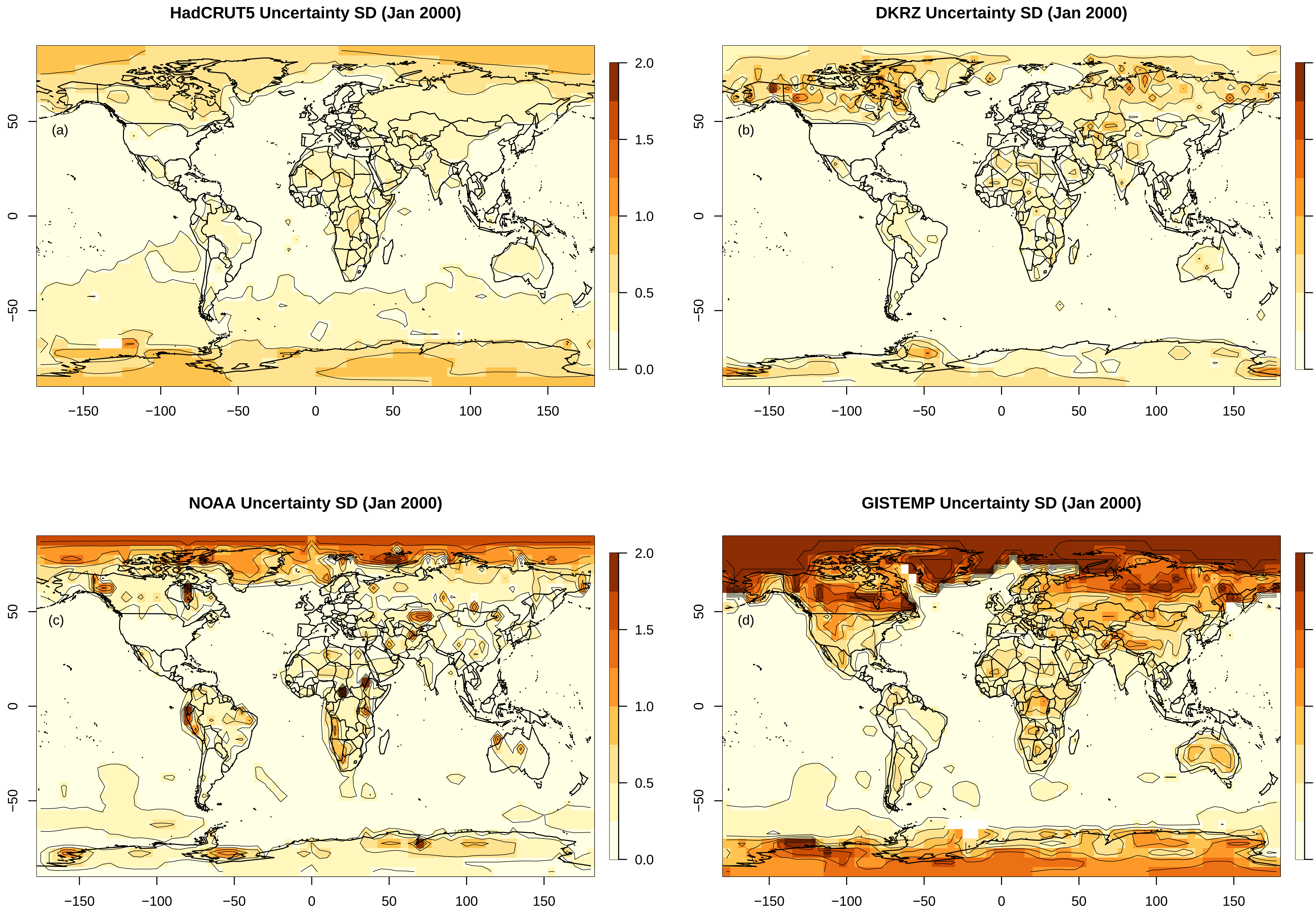

Temperature Uncertainty SD (January 1910)
(a)
(b)
Frequency
Monthly Temperature SD
Temperature Uncertainty SD (January 1940)
(c)
(d)
Frequency
Monthly Temperature SD
Temperature Uncertainty SD (January 1970)
(e)
(f)
Frequency
Monthly Temperature SD
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
−150
−100
−50
0
50
100
150
−50
0
50
1000
2000
3000
4000
5000

Figure 10.



HadCRUT5 Uncertainty SD (Jan 2000)
DKRZ Uncertainty SD (Jan 2000)
NOAA Uncertainty SD (Jan 2000)
GISTEMP Uncertainty SD (Jan 2000)
(a)
(b)
(c)
(d)
−50
0
50
−150
−100
−50
0
50
100
150
0.0
0.5
1.0
1.5
2.0