

# NOAA'S MERGED LAND–OCEAN SURFACE TEMPERATURE ANALYSIS

BY RUSSELL S. VOSE, DEREK ARNDT, VIVA F. BANZON, DAVID R. EASTERLING, BYRON GLEASON, BOYIN HUANG, ED KEARNS, JAY H. LAWRIKORE, MATTHEW J. MENNE, THOMAS C. PETERSON, RICHARD W. REYNOLDS, THOMAS M. SMITH, CLAUDE N. WILLIAMS JR., AND DAVID B. WUERTZ

A coherent picture of global surface temperature change since the late nineteenth century emerges from a statistical reconstruction of an integrated collection of historical temperature observations over the land and ocean.

**T**he most widely recognized measure of observed climate change is the century-scale trend in globally averaged surface temperature. The global average is a simple theoretical concept, but its computation in practice is far from trivial. The complexity stems mainly from the idiosyncrasies of historical weather observations, most of which were collected for operational purposes, such as aviation and agriculture, rather than climate change detection. In particular, certain practices that are of little operational significance such as relocating a station or changing its instrumentation, may profoundly impact the integrity of the climate record (Aguilar et al.

2003). Furthermore, the availability of observations in space and time reflects the highly uneven historical distribution of human activities across Earth's surface (particularly over the oceans).

At present there are three principal groups that operationally derive global surface temperature from this piecemeal historical record. These include the National Oceanic and Atmospheric Administration (NOAA) National Climatic Data Center (NCDC), the National Aeronautics and Space Administration (NASA) Goddard Institute for Space Studies (GISS; Hansen et al. 2010), and the joint venture by the Met Office Hadley Centre and the Climatic Research Unit (HadCRUT; Brohan et al. 2006). (Note that the Japan Meteorological Agency now has a global dataset, and land-only products are available from other groups, such as the Berkeley Earth Surface Temperature project.) In general, each group uses somewhat different input data and substantially different approaches. For instance, GISS makes extensive use of satellite data, NCDC only uses satellite data in a limited capacity, and HadCRUT uses no satellite data at all. Likewise, GISS and NCDC provide temperature estimates over unsampled areas whereas HadCRUT does not. Despite these underlying differences, however, the groups produce roughly comparable estimates of the long-term trend—that is, about 0.8°C of warming since 1900.

**AFFILIATIONS:** VOSE, ARNDT, BANZON, EASTERLING, GLEASON, HUANG, KEARNS, LAWRIKORE, MENNE, PETERSON, WILLIAMS, AND WUERTZ—NOAA/National Climatic Data Center, Asheville, North Carolina; REYNOLDS—Cooperative Institute for Climate and Satellites—North Carolina, Asheville, North Carolina; SMITH—NOAA/NESDIS/STAR, College Park, Maryland  
**CORRESPONDING AUTHOR:** Russell S. Vose, NOAA/National Climatic Data Center, 151 Patton Avenue, Asheville, NC 28801  
E-mail: russell.vose@noaa.gov

*The abstract for this article can be found in this issue, following the table of contents.*

DOI:10.1175/BAMS-D-11-00241.1

In final form 17 April 2012

This paper describes the current version of the NCDC product, known as the Merged Land–Ocean Surface Temperature analysis (MLOST version 3.5). The primary motivation for the current release is the inclusion of a new land–air temperature dataset that has several major improvements, including a more elaborate approach for addressing changes in station location, instrumentation, and siting conditions. For the sake of completeness, however, the paper provides a holistic overview of MLOST, including the underlying land and ocean datasets, the temperature reconstruction methodology, and trend differences between the current and previous releases. As this paper is intentionally general in nature, it contains pointers to more technical details where appropriate.

**LAND SURFACE AIR TEMPERATURE DATA.** The Global Historical Climatology Network–Monthly dataset (GHCN-M version 3; Lawrimore et al. 2011) is the source of land surface air temperature (LST) data for the new release of MLOST. The previous release used GHCN-M version 2 (Peterson and Vose 1997). GHCN-M is freely available to all users from NCDC. The dataset consists of historical monthly data from surface weather stations that have typically measured air temperature in protected shelters ~1.5 m above the ground. Quality assurance reviews are extensive and include tests for climatological outliers and spatial inconsistencies (Peterson et al. 1998). GHCN-M version 3 contains several additional quality checks relative to its predecessor, such as tests for erroneous isolated values and month-over-month data duplication. The dataset contains historical records for over 7,000 locations worldwide, with hundreds extending back to the early twentieth century. Real-time (i.e., monthly) updates are available via the Global Telecommunication System (GTS) for a network of more than 2,000 stations that collectively replicate the century-scale LST series with negligible differences compared with the full dataset (Lawrimore et al. 2011).

LSTs require adjustments to account for historical changes in station location, temperature instrumentation, and land use. Such changes typically impart either an abrupt jump or a gradual drift in the average temperature at a station relative to its neighbors, obscuring the true climate signal. In GHCN-M version 3, these contaminants are identified through automated, pairwise intercomparisons of station temperatures (Menne and Williams 2009). This “pairwise” approach has two advantages over the technique used in GHCN-M version 2 (and thus the previous release of MLOST): it is able to detect more historical

changes, and it also produces fewer false positives. Site-specific bias adjustments are an outcome of the pairwise comparisons, and approximately half of all stations require at least one adjustment to their record (Lawrimore et al. 2011). Major changes at individual stations can necessitate large adjustments at the local level (e.g., Reno, Nevada, required an adjustment of about 2°C in the 1990s due to a change in station location; Cairo Airport required an adjustment in excess of 3°C in the 1960s for the same reason). More important, national-scale changes in observing practice can result in near-concurrent adjustments to numerous stations, significantly impacting regional temperatures (e.g., the widespread transition to electronic sensors in the United States during the late twentieth century lowered maximum temperatures by about 0.25°C nationwide; Menne et al. 2009). Once averaged over the globe, however, bias adjustments have only a modest impact on the long-term LST series (i.e., about 0.02°C decade<sup>-1</sup>; Lawrimore et al. 2011).

The adjusted GHCN-M temperatures are standardized to account for large temperature gradients arising from differences in station elevation, latitude, and coastal proximity. This is accomplished by computing each station’s mean over a set base period (1961–90) and then subtracting that mean from each temperature value. A separate mean is computed for each month to compensate for changes in temperature across the seasons. The standardized values, known as anomalies, exhibit far less spatial and seasonal variability than the original temperatures. In fact, neighboring stations with very different means usually have comparable anomalies (e.g., Matsumoto and Tokyo differ on average by about 4°C due to a 600-m elevation difference and a physical separation of 150 km, yet their monthly anomalies are usually within 0.5°C of one another; Robeson 1994). If a mean cannot be computed for a station (e.g., because it lacks sufficient data in the base period), then the station is excluded from MLOST. Although this practice reduces the number of stations by about 40%, it only reduces the amount of data by about 20%, and most of that loss is in data-rich areas; in other words, there is minimal impact on global coverage.

Finally, the temperature anomalies are averaged together in a manner that accounts for the uneven distribution of stations in space (e.g., as in Jones et al. 2012). The first step in this process entails subdividing the land surface into grid boxes that are five degrees of latitude by five degrees of longitude in size, and then assigning each station to a grid box based on the station’s coordinates. An average temperature anomaly is then computed for each grid box, year,

and month using all of the stations available in the grid box at that time. Because each grid box has only one value for each year and month, regardless of how many stations were in the grid box, global averages are not unduly influenced by large numbers of stations in some regions. (However, the temporal variance of an individual grid box is dependent upon the number of stations within it, as well as on changes in the number through time; Brohan et al. 2006). The percentage of global land area covered by the resulting grid ranges from about 40% in the early twentieth century to over 60% at present, with a maximum of about 75% during the 1970s. Coverage since 1990 has improved by about 10% because GHCN-M version 3 contains updates to several major data sources, including *World Weather Records and Monthly Climatic Data for the World* (Lawrimore et al. 2011).

**SEA SURFACE TEMPERATURE DATA.** The International Comprehensive Ocean–Atmosphere Data Set (ICOADS version 2.1; Worley et al. 2005) is the source of sea surface temperature (SST) data for MLOST. ICOADS is freely available to all users from the NOAA Earth System Research Laboratory. The dataset consists of marine meteorological observations, primarily from ships and buoys, integrated from numerous historical data sources. Much like GHCN-M, quality assurance reviews are extensive and include tests for data duplication, climatological outliers, and position errors (Slutz et al. 1985; Smith and Reynolds 2003, 2004). The dataset contains at least a million observations per year since 1950, with roughly 100,000 per year in the early twentieth century. Although the dataset is updated each month, NCDC supplements ICOADS with data transmitted over the GTS to increase spatial coverage in near-real time.

As with LSTs, SSTs are averaged into grid boxes to account for the irregular distribution of observations across the ocean surface. In the case of SSTs, a  $2^\circ \times 2^\circ$  latitude/longitude grid is employed, each box value being an average of all daily observations in the box during a month (this grid resolution works well in most instances, but occasionally biases can exist in areas with extremely high gradients, such as the Gulf Stream). If a grid box has both ship and buoy data at the same time, then the buoy observations receive about 6 times as much weight in the final box average to compensate for the much greater statistical noise in the coincident ship reports (e.g., noise arising from mistakes in navigation, instrument calibration, and data transcription). This blending of ship and buoy data is particularly common in the

past several decades, which have witnessed a dramatic increase in the number of drifting and moored buoys on the global scale. Ultimately, all gridbox values are subjected to statistical quality assurance reviews to eliminate residual outliers (Smith and Reynolds 2004). The percentage of ocean area covered by the quality-assured SST grid ranges from about 30% in the early twentieth century to over 70% at present, with marked declines during the two World Wars.

Gridbox averages from ships require adjustments to compensate for several widespread changes in measurement practice. Although a variety of changes have impacted the historical SST record (e.g., Kennedy et al. 2011b), the period prior to the early 1940s is particularly significant. Most notably, before World War II SST observations usually involved bringing canvas or wooden buckets of seawater onto the deck of a ship, whereas thereafter it became more common to measure SST at the engine's cooling system seawater intake. Generally speaking, the older bucket-based data contained a cold bias relative to contemporary ship SST reports, obscuring the true climate signal (Folland and Parker 1995). Because ICOADS contains no bias adjustments, MLOST employs corrections based on nighttime marine air temperatures to address the transition away from bucket-based measurements (Smith and Reynolds 2002). In contrast to LST debiasing, the SST adjustments have a substantial impact when averaged over the entire ocean surface, warming the globe by several tenths of a degree prior to the early 1940s and thus significantly reducing the long-term warming trend (Smith and Reynolds 2002).

Finally, the adjusted SST grids are standardized to account for large gradients arising from differences in latitude, coastal proximity, and other factors. This is accomplished by computing each grid box's mean over a set base period and then subtracting that mean from each temperature value. In contrast to LSTs, a base period of 1971–2000 is used to maximize spatial coverage over the oceans (the two base periods are later reconciled, as described in the next section). As with LSTs, a separate mean is computed for each month to compensate for changes in temperature across the seasons. The resulting anomalies exhibit more spatial coherence than the raw SSTs, facilitating the computation of areal averages.

In summary, in situ ship and buoy data are the source of historical SSTs in the new version of MLOST. As discussed briefly in the next section, satellite-derived SSTs are also employed, but only in a very limited capacity (i.e., to identify recurrent “high frequency” temperature patterns over the ocean). Notably, a prior edition of MLOST (version 3) used

satellite data both to identify the high-frequency patterns and to create the actual year-by-year, month-by-month grids starting in 1985. The latter practice was eliminated in a later release (i.e., MLOST version 3b) because the satellite SSTs were not found to add appreciable value to a monthly analysis on a 2° grid, and they actually introduced a small but abrupt cool bias at the global scale starting in 1985 (Banzon et al. 2010).

**RECONSTRUCTION OF GLOBAL SURFACE TEMPERATURE.** The LST and SST grids are used to construct a coherent picture of historical temperature variations across the globe since the late nineteenth century. From NOAA's perspective, there are two primary requirements in developing this "reconstruction" of global temperature. The first requirement is to capture the major patterns in space and time while smoothing out small-scale, short-term irregularities (i.e., the reconstruction must separate the key trends from background noise). The second requirement is to estimate anomalies in areas without observations (provided there are sufficient proximal data in space and time) so that the resultant product represents as much of the global surface as is reasonably possible. The approach itself is briefly summarized here; for a more detailed description, see Smith and Reynolds (2005) and Smith et al. (2008).

The reconstruction process dictates that LSTs and SSTs should be processed separately and then merged together into a single global reconstruction. There are several reasons for performing separate reconstructions, the first being major differences in spatial coverage between the land and ocean surface (the latter having large, systematic gaps early in the record, such as in the Southern Ocean). Another motive for separate reconstructions is that LST and SST observations are fundamentally different; a land grid box represents at least one month of twice-daily observations from at least one station, whereas an ocean grid box might represent just a few measurements of SST from a single ship. Finally, the time and space scales of temperature variability are shorter over land than ocean because vastly more energy is required to change the latter owing to its higher specific heat and its lower speed of advection.

The reconstruction process also assumes that historical temperature variations can be divided into two distinct components that can be independently extracted from an otherwise noisy observational record. The first component consists of "low frequency" variations that occur over relatively long periods, such as the century-scale increase in global

temperature or the extended period of cooling in the Arctic in midcentury. The second component consists of "high frequency" variations that occur over comparatively short time periods, such as the extremely warm summer of 2011 in the eastern United States or the increase in SST along the western coast of South America during the El Niño event of 2009. To obtain the overall reconstruction, the low- and high-frequency components must be added together.

From a practical perspective, the two components are extracted in a sequential fashion. The first step entails identifying the low-frequency variations through extensive smoothing of the historical record on an annual time scale. This is basically accomplished by averaging the anomaly grids over vast areas, compositing across multiyear periods, assigning an anomaly value of zero to empty grid boxes, and then averaging again with simple space/time filters (Smith and Reynolds 2005). The resulting grids are identical in resolution to their unsmoothed counterparts, but spatial coverage is complete. The next step involves identifying the residual high-frequency variations by applying a pattern recognition technique to the historical record. The technique makes extensive use of empirical orthogonal teleconnections (Van den Dool et al. 2000), which are similar to empirical orthogonal functions, which have been widely employed in atmospheric science. To maximize spatial coverage, different periods are analyzed for land (1982–91) and ocean (1982–2002); furthermore, the grids derived from ship and buoy data are supplanted by similar grids of satellite SST retrievals with enhanced resolution (Smith and Reynolds 2002). Although the high-frequency variations are derived from the modern era, they are generally representative of the entire historical record (Smith and Reynolds 2005). The pattern recognition analysis results in ~60 recurring patterns over land and ~130 over the ocean, totals that are consistent with the relative areas of land and ocean. The resulting high-frequency grid boxes are identical in size to their low-frequency counterparts, and the two are simply added together to form preliminary reconstructions that are spatially complete [note that the SST reconstruction used here is also available as a standalone product known as the Extended Reconstruction of Sea Surface Temperature (ERSST; Smith and Reynolds 2004)].

Given their large-scale focus (generally >2,000 km), the preliminary reconstructions intentionally lack some of the spatial detail of the original data. This is advantageous over the ocean, where background noise is often considerable because many grid boxes may contain only a few observations in any given

month. Over land, however, the approach can lead to an overly smooth representation of reality. To compensate, the preliminary land reconstruction is blended with the original LSTs when they are available. For each grid box, year, and month, the blending process assigns the original LST value a weight proportional to the number of stations in the box at that time; the weight is roughly 50% when two stations are available, increasing to about 95% when nine stations are available (Smith et al. 2008). This blending process increases spatial variability (and thus fidelity) over land, but it has almost no impact on global averages.

The preliminary land and ocean reconstructions are then merged into a single global reconstruction. This is a straightforward process for the most part; grid boxes over land come from the land reconstruction, and grid boxes over the ocean come from the ocean reconstruction. Despite the overt simplicity, however, three subtle processing steps are required to fully merge the separate reconstructions. The first step involves converting the land values from their 1961–90 base period to the 1971–2000 base period used over the ocean (i.e., by computing the average for each land grid box by month for 1971–2000, and then subtracting the appropriate monthly average from each value). The second step entails averaging the  $2^\circ \times 2^\circ$  ocean values into  $5^\circ \times 5^\circ$  boxes that exactly match the existing land grid. The third step requires creating a weighted average of land and ocean values for coastal grid boxes, the weights reflecting the relative proportions of land and ocean area in each box (which is a rational and effective approach for the most part, but on occasion a long record from an island station is downweighted significantly if most of the grid box is ocean).

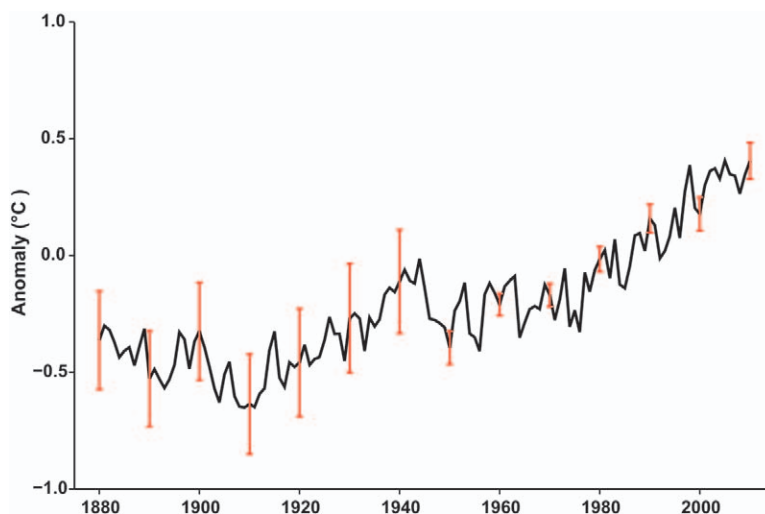
As the final step in the reconstruction process, data-sparse areas are masked out to prevent global averages from depending heavily upon the highly smoothed reconstruction estimates that are common in such locales. For instance, if sea ice [as identified by Rayner et al. (2003)] covers more than half of the grid box in that month, then it is set to missing. If land covers more than half the grid box, and if the box itself contains no land data in that month, and if its latitude is “polar” (i.e., north of  $75^\circ\text{N}$  or south of  $65^\circ\text{S}$ ), then it is set to missing. Finally, if the general area surrounding the grid box contains few observations in that

month (i.e., the sampling rate is less than 20% within a  $25^\circ$  latitude/longitude buffer around the box), then it is set to missing. In general, the masking of data-sparse areas is more common early in the record.

Global averages are derived from the final reconstruction using a relatively simple process. In particular, the global average in each year and month is just the area-weighted average of all grid boxes having a value in that year and month. In turn, annual global averages are simply the arithmetic mean of the 12 monthly averages. Using this approach, the percentage of Earth’s surface represented by the global average generally increases through time, exceeding 80% in each year since 1900. Spatial coverage is poor in most of the nineteenth century, resulting in large errors in the global average (Hansen et al. 2010; Smith et al. 2008). Consequently, the MLOST global time series begins in 1880.

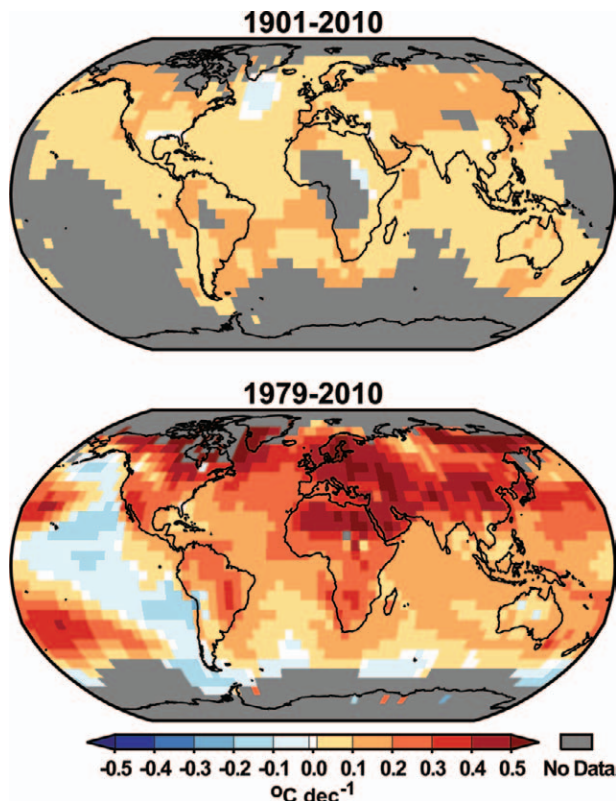
Figure 1 presents global temperature variations from 1880 to 2010 as depicted by the new reconstruction. Not surprisingly, the new global series is broadly consistent with previous analyses (i.e., Solomon et al. 2007), exhibiting a slight decrease until  $\sim 1910$ , an increase until  $\sim 1940$ , a slight decrease until  $\sim 1970$ , and an increase thereafter. The rate of warming (based on least squares regression) is  $0.076^\circ\text{C decade}^{-1}$  since 1901 and  $0.162^\circ\text{C decade}^{-1}$  since 1979. The three warmest years on record are still 2005, 2010, and 1998, and the difference between them remains virtually indistinguishable (a mere  $0.02^\circ\text{C}$ ).

Figure 1 also presents 95% confidence intervals that estimate the “uncertainty” in global mean temperature. In general, the confidence intervals decrease



**FIG. 1. Time series of mean annual temperature anomalies (black) and 95% confidence intervals (red) over the globe for the period 1880–2010.**





**FIG. 2.** Least-squares trends in  $5^\circ \times 5^\circ$  grid boxes for the periods (top) 1901–2010 and (bottom) 1979–2010. Annual anomalies in each grid box were only computed when at least 10 monthly anomalies were available. A minimum of 75% nonmissing annual data were required within the analysis (i.e., trend) period. At least 25% of that minimum number must have existed in each segment of the analysis period when divided into four equal segments. Finally, when the analysis period was divided into 10 equal segments, each of the first two and last two segments were required to have at least some data.

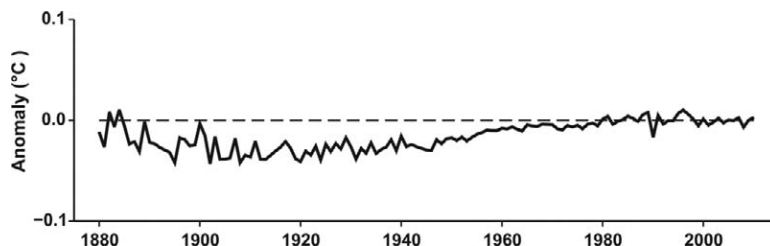
in size from about  $0.5^\circ\text{C}$  in the late nineteenth century to about  $0.1^\circ\text{C}$  in the late twentieth century (e.g., as in Jones et al. 2012; Morice et al. 2012). Very generally, these intervals imply that the total increase in global temperature since 1901 was roughly  $0.8^\circ \pm 0.2^\circ\text{C}$ . The MLOST confidence intervals are primarily intended to quantify two of the largest sources of noise in the global average—namely, gaps in spatial coverage and errors in the SST bias adjustments (Smith et al. 2008). Owing particularly to the latter, the confidence intervals increase significantly in size before the mid-1940s. Because there are a variety of ways to estimate uncertainty, all yielding somewhat different results (e.g., Brohan et al. 2006; Kennedy et al. 2011a,b), it is generally more appropriate to view such

confidence intervals as a broad depiction of historical noise rather than as a precise time series of exact error estimates.

Figure 2 illustrates the patterns of surface temperature change over 1901–2010 and 1979–2010 as depicted by the new reconstruction. As in the global series, the results are broadly consistent with previous analyses (i.e., Solomon et al. 2007), with widespread warming in both time periods and somewhat larger temperature increases over land than ocean. From the century-long perspective, warming is largest over central Eurasia, northwestern North America, central South America, and the South Atlantic Ocean, with evidence of cooling in the North Atlantic Ocean and small areas in the southeastern United States and central Africa. For the recent period, warming is evident in virtually all areas except the eastern Pacific Ocean and parts of the Southern Ocean. The highest rates of warming are over Eurasia, northern Africa, and northern North America. Notably, spatial coverage is more extensive in the recent period due to numerous improvements in ocean climate observing systems.

### DIFFERENCES IN THE NEW RELEASE.

From a large-scale perspective, there are minor but systematic differences between the current version of MLOST (version 3.5) and its predecessor (version 3b). As an illustration, Fig. 3 presents differences between global temperatures in the two reconstructions since the late nineteenth century. The current version of MLOST is very slightly cooler than its predecessor up until about the mid-1960s. Furthermore, the difference series exhibits a small but statistically significant increase of  $0.004^\circ\text{C decade}^{-1}$  since 1901. As a result, the current version of MLOST contains slightly more warming during the twentieth century ( $0.076^\circ$  vs  $0.072^\circ\text{C decade}^{-1}$ ). The difference is largely attributable to the inclusion of GHCN-M version 3, which itself exhibits about  $0.003^\circ\text{C decade}^{-1}$  more warming than the LST dataset used in the past (i.e., GHCN-M version 2; Lawrimore et al. 2011), largely due to improved bias adjustments.

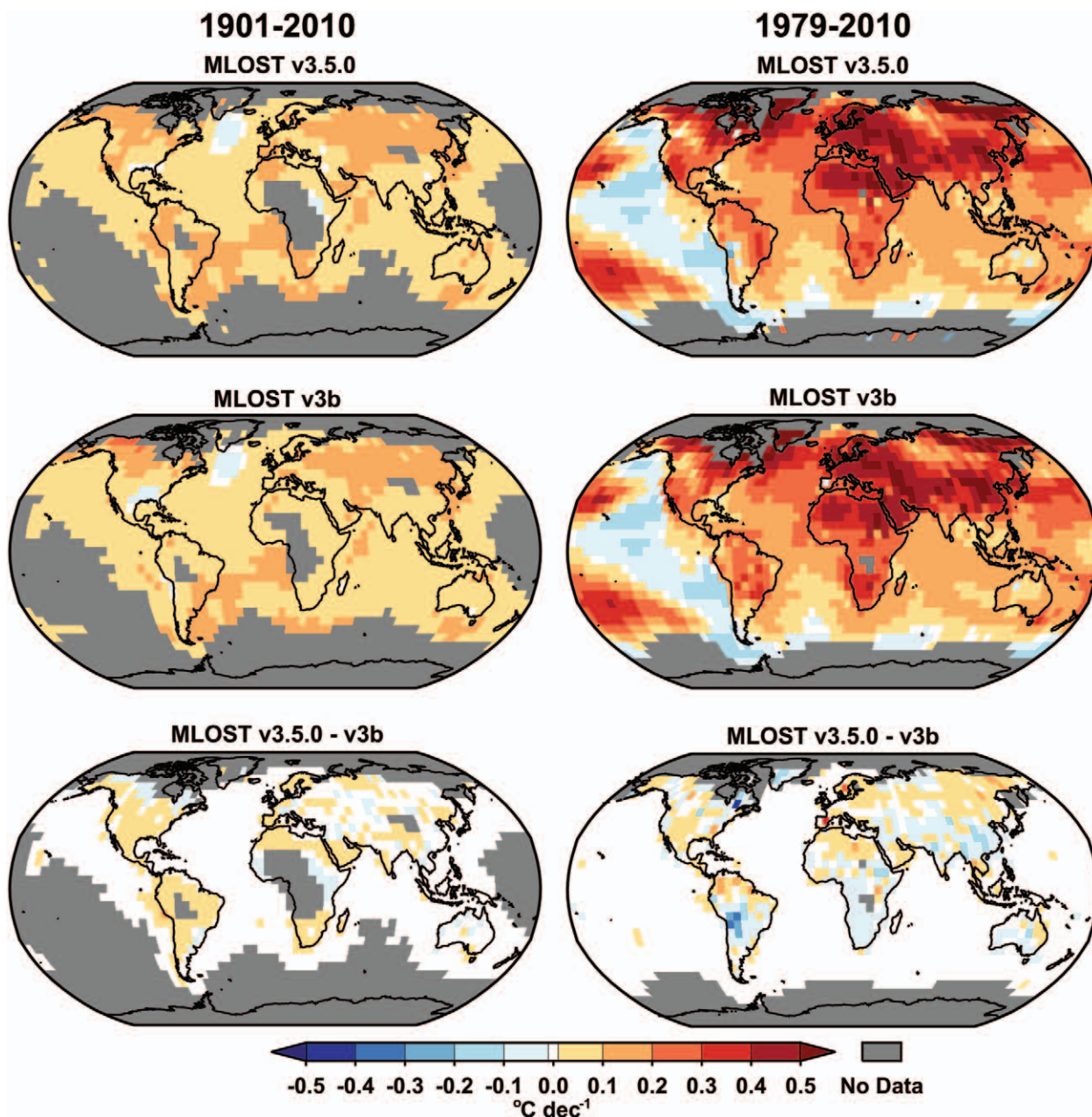


**FIG. 3.** Time series of differences in mean annual temperature anomalies over the globe for the period 1880–2010 (current version of MLOST minus the previous version).

Figure 4 illustrates the patterns of surface temperature change as depicted in both the new reconstruction and its predecessor for the periods 1901–2010 and 1979–2010. Consistent with the use of a new LST dataset, most trend differences appear over the terrestrial surface, with comparable changes in adjacent ocean areas. From the century-long perspective, the new reconstruction contains slightly more warming in most areas, although regional exceptions do exist (e.g., northern North America, northern Australia). Generally speaking, trend differences are less than

$0.1^{\circ}\text{C decade}^{-1}$  at the gridbox level. For the recent period, the new reconstruction exhibits less warming over southern Greenland, eastern Canada, central South America, sub-Saharan Africa, southern Asia (particularly China), and eastern Australia. In contrast, more warming is evident over the southern United States, northern South America, northern Africa, and northern Eurasia.

From a spatial perspective, there are two other subtle changes in the new version of MLOST. First, coverage of polar land areas has improved very



**FIG. 4.** Least-squares trends in  $5^{\circ} \times 5^{\circ}$  grid boxes for (top) the current version of MLOST and (middle) the previous version of MLOST, and (bottom) differences between the two, for (left) 1901–2010 and (right) 1979–2010. Completeness criteria are as in Fig. 2.

slightly in the modern era because grid boxes that contain actual observing stations are now included in the new version (i.e., the actual LST box value is “reinject” into the reconstruction whereas in the past the reconstructed value itself was set to missing because of its overly smoothed character). Second, trend patterns over land exhibit slightly greater spatial coherence (e.g., trends over South America are more uniform for 1979–2010 than in the previous reconstruction).

**SUMMARY.** This paper described the new release of MLOST (version 3.5), the global surface temperature product used by NOAA in monitoring and assessment activities. The primary motivation for the release was the inclusion of a new land dataset (GHCN-M version 3), which contains improved adjustments for changes in station location, instrumentation, and siting conditions. The new version is broadly consistent with previous global analyses, exhibiting a trend of  $0.076^{\circ}\text{C decade}^{-1}$  over the past century,  $0.162^{\circ}\text{C decade}^{-1}$  over the past three decades, and widespread warming in both time periods. In general, the new release exhibits only modest differences with its predecessor, the most obvious being very slightly more warming at the global scale and slightly different trend patterns over the terrestrial surface.

**FUTURE IMPROVEMENTS.** As with all large-scale temperature analyses, MLOST could benefit in the future from improved spatial sampling, particularly early in the record. For the ocean surface, continued integration of historical marine collections into ICOADS is of critical importance. The Atmospheric Circulation Reconstructions over the Earth (ACRE) initiative (Allan et al. 2011) plays an important role in that regard, having facilitated several major data recovery projects over the past several years (e.g., ship logbooks from the English East India Company, Royal Navy logbooks from World War I). For the land surface, MLOST could be augmented with data from the Global Historical Climatology Network daily dataset (Menne et al. 2012), which contains about 3 times as many stations as its monthly counterpart. A longer-term prospect is the recently instigated International Surface Temperature Initiative (Thorne et al. 2011), which seeks to create a single comprehensive databank of the actual land surface observations taken globally.

From a methodological perspective, MLOST could benefit from a more detailed treatment of historical uncertainties. At present MLOST confidence intervals mainly reflect gaps in spatial sampling

and errors in SST bias adjustments, and thus uncertainty estimates increase steadily and significantly prior to World War II. While it is arguable whether a truly comprehensive error model is possible, several avenues for refining statistical uncertainty estimates have recently been proposed, such as the use of large ensembles and benchmarking (e.g., Kennedy et al. 2011a,b; Williams et al. 2012). New approaches for addressing other uncertainties (e.g., the types of buckets used from 1945 to 1960) are also possible (Kennedy et al. 2011a,b). With improved uncertainty estimates will come increased confidence in long-term trends.

## REFERENCES

- Aguilar, E., I. Auer, M. Brunet, T. C. Peterson, and J. Wieringa, 2003: Guidelines on climate metadata and homogenization. WCDMP-53, WMO/TD-1186, 55 pp.
- Allan, R., P. Brohan, G. P. Compo, R. Stone, J. Luterbacher, and S. Bronnimann, 2011: The international Atmospheric Circulation Reconstructions over the Earth (ACRE) initiative. *Bull. Amer. Meteor. Soc.*, **92**, 1421–1425.
- Banzon, V. F., R. W. Reynolds, and T. Smith, 2010: The role of satellite data in extended reconstruction of sea surface temperature. *Proc. Oceans from Space Symp.*, Venice, Italy, Joint Research Center, 27–28.
- Brohan, P., J. J. Kennedy, I. Harris, S. F. B. Tett, and P. D. Jones, 2006: Uncertainty estimates in regional and global observed temperature changes: A new data set from 1850. *J. Geophys. Res.*, **111**, D12106, doi:10.1029/2005JD006548.
- Folland, C. K., and D. E. Parker, 1995: Correction of instrumental biases in historical sea surface temperature data. *Quart. J. Roy. Meteor. Soc.*, **121**, 319–367.
- Hansen, J., R. Ruedy, M. Sato, and K. Lo, 2010: Global surface temperature change. *Rev. Geophys.*, **48**, RG4004, doi:10.1029/2010RG000345.
- Jones, P. D., D. H. Lister, T. J. Osborn, C. Harpham, M. Salmon, and C. P. Morice, 2012: Hemispheric and large-scale land surface air temperature variations: An extensive revision and update to 2010. *J. Geophys. Res.*, **117**, D05127, doi:10.1029/2011JD017139.
- Kennedy, J. J., N. A. Rayner, R. O. Smith, D. E. Parker, and M. Saunby, 2011a: Reassessing biases and other uncertainties in sea surface temperature observations measured in situ since 1850: 1. Measurement and sampling uncertainties. *J. Geophys. Res.*, **116**, D14103, doi:10.1029/2010JD015218.
- , —, —, —, and —, 2011b: Reassessing biases and other uncertainties in sea surface



- temperature observations measured in situ since 1850: 2. Biases and homogenization. *J. Geophys. Res.*, **116**, D14104, doi:10.1029/2010JD015220.
- Lawrimore, J. H., M. J. Menne, B. E. Gleason, C. N. Williams, D. B. Wuertz, R. S. Vose, and J. Rennie, 2011: An overview of the Global Historical Climatology Network temperature dataset, version 3. *J. Geophys. Res.*, **116**, D19121, doi:10.1029/2011JD016187.
- Menne, M. J., and C. N. Williams Jr., 2009: Homogenization of temperature series via pairwise comparisons. *J. Climate*, **22**, 1700–1717.
- , —, and R. S. Vose, 2009: The United States Historical Climatology Network monthly temperature data, version 2. *Bull. Amer. Meteor. Soc.*, **90**, 993–1107.
- , I. Durre, R. S. Vose, B. E. Gleason, and T. G. Houston, 2012: An overview of the Global Historical Climatology Network-Daily database. *J. Atmos. Oceanic Technol.*, **29**, 897–910.
- Morice, C. P., J. J. Kennedy, N. A. Rayner, and P. D. Jones, 2012: Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The HadCRUT4 dataset. *J. Geophys. Res.*, **117**, D08101, doi:10.1029/2011JD017187.
- Peterson, T. C., and R. S. Vose, 1997: An overview of the Global Historical Climatology Network temperature database. *Bull. Amer. Meteor. Soc.*, **78**, 2837–2849.
- , —, R. L. Schmoyer, and V. N. Razuvaev, 1998: Global Historical Climatology Network (GHCN) quality control of monthly temperature data. *Int. J. Climatol.*, **18**, 1169–1179.
- Rayner, N. A., D. E. Parker, E. B. Horton, C. K. Folland, L. V. Alexander, D. P. Rowell, E. C. Kent, and A. Kaplan, 2003: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J. Geophys. Res.*, **108**, 4407, doi:10.1029/2002JD002670.
- Robeson, S. M., 1994: Influence of spatial sampling and interpolation on estimates of air temperature change. *Climate Res.*, **4**, 119–126.
- Slutz, R. J., S. J. Lubker, J. D. Hiscox, S. D. Woodruff, R. L. Jenne, D. H. Joseph, P. M. Steurer, and J. D. Elms, 1985: COADS Comprehensive Ocean–Atmosphere Data Set. Release 1, 262 pp. [Available from Climate Research Program, Environmental Research Laboratories, 325 Broadway, Boulder, CO 80303.]
- Smith, T. M., and R. W. Reynolds, 2002: Bias corrections for historical sea surface temperatures based on marine air temperatures. *J. Climate*, **15**, 73–87.
- , and —, 2003: Extended reconstruction of global sea surface temperatures based on COADS data (1854–1997). *J. Climate*, **16**, 1495–1510.
- , and —, 2004: Improved extended reconstruction of SST (1854–1997). *J. Climate*, **17**, 2466–2477.
- , and —, 2005: A global merged land–air–sea surface temperature reconstruction based on historical observations (1880–1997). *J. Climate*, **18**, 2021–2036.
- , —, T. C. Peterson, and J. Lawrimore, 2008: Improvements to NOAA’s historical Merged Land–Ocean Surface Temperature analysis (1880–2005). *J. Climate*, **21**, 2283–2296.
- Solomon, S., D. Qin, M. Manning, M. Marquis, K. Averyt, M. M. B. Tignor, H. L. Miller Jr., and Z. Chen, Eds., 2007: *Climate Change 2007: The Physical Science Basis*. Cambridge University Press, 996 pp.
- Thorne, P. W., and Coauthors, 2011: Guiding the creation of a comprehensive surface temperature resource for twenty-first century climate science. *Bull. Amer. Meteor. Soc.*, **92**, ES40–ES47.
- Van den Dool, H. M., S. Saha, and A. Johansson, 2000: Empirical orthogonal teleconnections. *J. Climate*, **13**, 1421–1435.
- Williams, C. N., M. J. Menne, and P. W. Thorne, 2012: Benchmarking the performance of pairwise homogenization of surface temperatures in the United States. *J. Geophys. Res.*, **117**, D05116, doi:10.1029/2011JD016761.
- Worley, S. J., S. D. Woodruff, R. W. Reynolds, S. J. Lubker, and N. Lott, 2005: ICOADS release 2.1 data and products. *Int. J. Climatol.*, **25**, 823–842, doi:10.1002/joc.1166.