

Uncertainty in annual rankings from NOAA’s global temperature time series

Anthony Arguez,¹ Thomas R. Karl,¹ Michael F. Squires,¹ and Russell S. Vose¹

Received 12 September 2013; revised 7 November 2013; accepted 8 November 2013; published 27 November 2013.

[1] Annual rankings of global temperature are an important component of climate monitoring. However, there is some degree of uncertainty for every yearly value in the global temperature time series, which leads to uncertainty in annual rankings as well. This study applies a Monte Carlo uncertainty analysis to the National Oceanic and Atmospheric Administration (NOAA) National Climatic Data Center’s global land-ocean surface temperature (NOAATMP) time series. Accounting for persistence between years does not materially affect the results versus presuming statistical independence. The highest probabilities for the warmest year analysis (1880–2012) are associated with the years 2010 (~36%), 2005 (~28%), and 1998 (~11%). The current separation among the warmest observed years is relatively small compared to the standard errors of the NOAATMP time series. However, each year between 1997 and 2012 was warmer than the vast majority of all other years since 1880 at the 95% confidence level. **Citation:** Arguez, A., T. R. Karl, M. F. Squires, and R. S. Vose (2013), Uncertainty in annual rankings from NOAA’s global temperature time series, *Geophys. Res. Lett.*, 40, 5965–5969, doi:10.1002/2013GL057999.

1. Introduction

[2] The National Oceanic and Atmospheric Administration (NOAA) National Climatic Data Center (NCDC) is one of a handful of groups that produces a global temperature product (land and ocean). Other groups include the Met Office Hadley Centre and the Climate Research Unit, which jointly developed the HadCRUT4 data set [Morice *et al.*, 2012], and the Goddard Institute for Space Studies (GISS), which developed the GISS Surface Temperature (GISTEMP) analysis [Hansen *et al.*, 2010]. As part of its continuous monitoring activities, NCDC routinely assesses the rankings of individual months and years for numerous spatial scales (e.g., national, regional, and state), based on the period of record for the time series in question. A higher-profile example of this is the annual rankings of the global temperature series, which is derived from the NOAA NCDC’s global land-ocean surface temperature (NOAATMP) time series [Vose *et al.*, 2012]. This leads to statements like the following: the 15 hottest years on record have occurred in the past 16 years (1997–2012; see Table 1).

[3] Such statements of observational “fact” do not take into account the nontrivial degree of uncertainty for each

annual time series value. A common characterization of this uncertainty is the standard error time series. For annual time series of U.S. and global temperature, the standard errors can be quite large compared to the typical year-to-year fluctuation in the time series values. This can lead to substantial overlap of uncertainty ranges when years are ranked from coldest to warmest. This concept was explored by Guttorp and Kim [2013, hereinafter GK13], who used a Monte Carlo approach to characterize the uncertainty of annual rankings for the U.S. temperature time series. In this note, we apply a variant of the GK13 uncertainty approach to the NOAATMP time series. The data and methodology are presented in section 2, followed by the results in section 3. The discussion and conclusions are offered in section 4.

2. Data and Methodology

[4] In this study, we utilize the global time series and associated total standard errors produced by the NOAATMP analysis. This data set is referred to as Merged Land-Ocean Surface Temperature (MLOST) version 3.5 analysis in Vose *et al.* [2012]. We have chosen not to use the MLOST naming convention, as NOAA NCDC is in the process of revising its naming conventions. NOAATMP consists of monthly gridded temperature anomalies over the globe for the period 1880 to present. Data over land areas originate from weather stations measuring near-surface air temperature, whereas data over ocean areas originate mainly from ships and buoys measuring sea surface temperature (SST). Gridded anomalies are produced using a statistical reconstruction method that sequentially extracts low- and high-frequency components from the historical temperature record [Smith and Reynolds, 2005; Smith *et al.*, 2008]. An outcome of the reconstruction process is a standard error estimate that explicitly quantifies the random, sampling, and bias components of uncertainty [Smith and Reynolds, 2005]. Random errors mainly originate in the input data, but these errors are largely filtered out by the reconstruction process. Sampling error reflects the density and distribution of the original observations. Bias error mainly results from systematic changes in observing practice, such as from historical changes in instrumentation, particularly over the ocean [Smith and Reynolds, 2002].

[5] Table 1 shows the 15 warmest years on record from 1880 to 2012. The NOAATMP time series and its standard errors are plotted in Figure 1. The observed warmest year on record was 2010, followed by 2005. The difference between the 2 years is less than 100th of a degree Celsius, which is substantially smaller than the associated standard errors. Note that the standard errors are significantly higher before the mid-1940s; this is primarily due to errors associated with SST bias adjustments [Vose *et al.*, 2012]. The subtle rise in standard errors from 1880 to the mid-1930s

¹NOAA National Climatic Data Center, Asheville, North Carolina, USA.

Corresponding author: A. Arguez, National Oceanic and Atmospheric Administration National Climatic Data Center, Rm. 506, 151 Patton Ave., Asheville, NC 28801, USA. (Anthony.Arguez@noaa.gov)

Table 1. The 15 Warmest Years on Record (1880–2012) in the NOAAATMP Time Series

Ranking	Year	Anomaly (°C)
1	2010	0.414
2	2005	0.406
3	1998	0.386
4	2003	0.376
5	2002	0.368
6	2006	0.354
7	2009	0.351
8	2007	0.346
9	2004	0.333
10	2012	0.331
11	2001	0.305
12	2011	0.290
13	2008	0.268
14	1997	0.268
15	1999	0.209

is due exclusively to bias uncertainty over the oceans, coinciding with the general switchover from wood buckets to canvas buckets [Kennedy et al., 2011].

[6] A Monte Carlo analysis can be used to assess the uncertainty of annual rankings. The simplest approach is to assume that all years in the annual time series, $X(t)$, are independent. A large number of simulated time series can be created by shifting each estimated annual temperature up or down slightly by a random z score (i.e., Gaussian white noise), $z_i(t)$, scaled by the standard error (Figure 1b), $s(t)$, i.e.,

$$Y_i(t) = X(t) + s(t)z_i(t). \quad (1)$$

[7] Here $Y_i(t)$ represents the i th simulation time series. GK13 called this the “independent model.” We simulate 100,000 realizations of the annual NOAAATMP time series in this manner. From this pool of synthetic time series, we then calculate the probability that a given year was the warmest on record, as well as the probability that a given year was among the top 10 warmest years on record. The synthetic time series are also used to construct two-tailed 95% confidence intervals for each annual ranking.

[8] A more complicated approach for assessing the uncertainty in ranks involves modeling the dependence structure of the annual time series using a Box-Jenkins type of analysis. The NOAAATMP global temperature time series (Figure 1a) is clearly nonstationary. Removal of an ordinary least squares regression line does not yield a stationary time series, as considerable spectral energy resides in the first two harmonics (not shown). The same is true if a generalized least squares regression line is modeled, the approach applied by GK13 to an annual U.S. time series. One way to achieve a stationary time series is to estimate a nonlinear trend using empirical mode decomposition (EMD), a nonparametric scheme for extracting intrinsic mode functions (IMFs) from a time series, leaving behind a “residue” that can be considered a time-dependent mean function [Huang and Shen, 2005]. EMD is increasingly being utilized in climate science applications [Wang et al., 2013]. We utilize the emd function in the R programming language [R Core Team, 2013] to extract the four leading IMFs, resulting in the smooth residue shown in Figure 1a. The residuals of this nonlinear trend fit (not shown) are stationary as determined by Kwiatkowski-

Phillips-Schmidt-Shin test [Kwiatkowski et al., 1992] and Phillips-Perron test [Phillips and Perron, 1988] using the `kps.test` and `pp.test` functions in R, respectively. Following GK13, we estimate the dependence structure of these residuals using the R `auto.arima` function, which determines the optimal autoregressive integrated moving average (ARIMA) parameters of the residual time series. Results are reported using both the Akaike information criterion and the Bayesian information criterion (AIC and BIC, respectively). We calculate 100,000 synthetic time series as in (1), except, instead of using a time series of purely randomized normal deviates, we utilize the ARIMA coefficients along with an innovation series [see Wilks, 2006] to create a time series of z scores, $z_i(t)$, that reflects the dependence structure of the residual time series. This “dependent model” takes the following form:

$$Y'_i(t) = X(t) + s(t)z'_i(t). \quad (2)$$

[9] The dependent simulation time series, $Y'_i(t)$, are used to calculate the uncertainty of the warmest and top 10 warmest years as before. In essence, the dependent approach accounts for the fewer effective degrees of freedom due to the redness of the NOAAATMP time series.

3. Results

[10] As summarized in Table 2, no single year exceeds a 50% probability of having been the warmest year on record globally in the NOAAATMP data set. The year 2010 has the highest probability for all three cases, with values ranging from 36.5% to 38.3%. The year 2005 comes in second

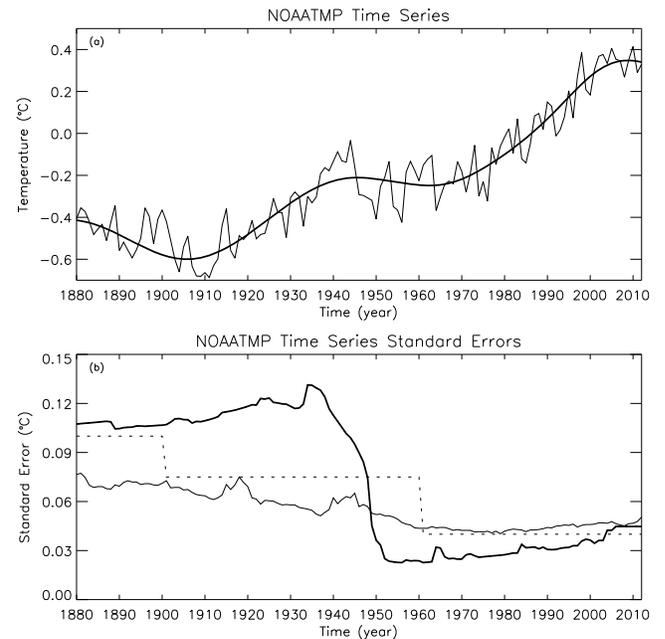


Figure 1. (a) NOAAATMP anomaly time series (1880–2012). The smooth curve indicates the residue of the EMD analysis. (b) The standard errors of the NOAAATMP time series (1880–2012) are presented in bold. The associated HadCRUT4 (solid line) and GISTEMP (dashed line) analyses are also displayed.

Table 2. The Probability (%) That a Year was the Warmest on Record (1880–2012) in the NOAAATMP Time Series, Calculated Using the Independent Years Assumption and the Dependent Years Approach With Both the AIC and the BIC^a

Rank	Independent NOAAATMP		Dependent NOAAATMP (AIC)		Dependent NOAAATMP (BIC)		Independent HadCRUT4		Independent GISTEMP	
	Year	Probability (%)	Year	Probability (%)	Year	Probability (%)	Year	Probability (%)	Year	Probability (%)
1	2010	36.54	2010	38.26	2010	37.75	2010	30.74	2010	39.22
2	2005	28.09	2005	29.16	2005	28.71	2005	24.80	2005	31.90
3	1998	10.63	1998	11.23	1998	11.14	1998	18.24	2007	8.58
4	2003	6.80	2003	7.00	2003	6.91	2003	7.11	2002	6.61
5	2002	4.42	2002	4.18	2002	4.42	2006	5.09	1998	6.19
6	2006	4.12	2007	3.02	2006	3.12	2009	4.93	2003	3.19
7	2009	3.67	2006	2.77	2007	3.02	2002	4.55	2009	1.97
8	2007	2.87	2009	2.22	2009	2.49	2007	2.86	2006	1.82

^aThe corresponding independent approach results for HadCRUT4 and GISTEMP are also shown. Only years with a probability above 1% across all five cases are shown.

(28.1–29.2%), followed by the year 1998 in third (~11%). Five additional years have probabilities between 3% and 7%: 2003, 2002, 2006, 2009, and 2007. Ranks 6–8 across the three cases differ slightly (although they are different orderings of the same 3 years), but the rankings for the top 5 years are the same and are relatively well separated.

[11] With respect to the probability that a given year was among the 10 warmest years on record, the results (Table 3) are remarkably consistent and well separated across all three cases for the top 14 years. Note that the top 14 years coincide with the observed values (Table 1). All 14 years presented have occurred in the last 16 years. All other years had a probability of less than 1% of having been among the warmest 10 years on record.

[12] Figure 2 shows the 95% confidence intervals associated with each annual ranking (using the independent assumption simulations). Not surprisingly, the range of rankings is larger during the earlier half of the record when the NOAAATMP standard errors were larger. The ranges can vary widely. For example, the confidence interval for the year 1934 ranges from 33rd to 116th warmest, whereas for 1995, the range is from thirteenth to nineteenth warmest. The confidence intervals for the 3 years with the warmest anomalies in the NOAAATMP global time series are first to ninth for both 2010 and 2005 and first to tenth for 1998, although they are all clearly skewed toward warmer rankings.

4. Discussion and Conclusions

[13] All of the associated probabilities presented herein for NOAAATMP differ by ~2% or less across the three cases considered, so we contend that the independent and dependent methods yield comparable results for annual rankings, the same conclusion drawn by GK13 for the U.S. time series. Our results suggest that accounting for the dependence of annual values may be a prudent academic exercise, but the practical effect is muted for the global temperature time series. Any effect on the rankings due to the dependence of years is dwarfed by the effect associated with the large ratio of the standard errors to the year-to-year variability and the actual separation of annual values. In terms of practicality, the independent approach has the advantage of being automatable, whereas the dependent approach requires some degree of manual inspection to ensure the appropriateness of the EMD and ARIMA time series modeling involved. Therefore, we recommend the use of the Monte Carlo approach using the independent assumption to diagnose the uncertainty of annual rankings in climate monitoring applications involving global temperature time series.

[14] For example, the NCDC’s most recent annual state of the climate report (<http://www.ncdc.noaa.gov/sotc/global/2012/13>) includes the following analysis: “The year 2012 was the 10th warmest year since records began in 1880.... Including 2012, all 12 years to date in the 21st century

Table 3. The Probability (%) That a Year was Among the 10 Warmest on Record (1880–2012) in the NOAAATMP Time Series, Calculated Using the Independent Years Assumption and the Dependent Years Approach With Both the AIC and the BIC^a

Rank	Independent NOAAATMP		Dependent NOAAATMP (AIC)		Dependent NOAAATMP (BIC)		Independent HadCRUT4		Independent GISTEMP	
	Year	Probability (%)	Year	Probability (%)	Year	Probability (%)	Year	Probability (%)	Year	Probability (%)
1	2010	99.13	2010	99.42	2010	99.37	2010	99.25	2010	99.90
2	2005	98.94	2005	99.32	2005	99.26	2005	98.72	2005	99.84
3	1998	98.35	1998	98.24	1998	98.24	1998	98.48	2007	98.42
4	2003	97.02	2003	97.53	2003	97.33	2003	93.86	2002	97.83
5	2002	95.14	2002	95.72	2002	95.68	2006	91.90	1998	97.58
6	2006	86.13	2006	86.37	2006	86.43	2009	91.68	2003	95.12
7	2009	84.35	2009	84.93	2009	84.77	2002	90.58	2009	92.31
8	2007	81.77	2007	82.99	2007	82.62	2007	87.85	2006	92.19
9	2004	74.40	2004	73.52	2004	73.86	2012	63.98	2012	77.79
10	2012	71.34	2012	71.86	2012	72.00	2004	62.35	2011	59.85
11	2001	47.93	2001	46.58	2001	46.43	2001	55.85	2001	44.18
12	2011	34.78	2011	34.15	2011	34.35	2011	29.16	2004	29.65
13	2008	18.27	2008	16.75	2008	17.16	1997	18.19	2008	12.10
14	1997	11.65	1997	11.79	1997	11.69	2008	16.42	1997	2.46

^aThe corresponding independent approach results for HadCRUT4 and GISTEMP are also shown. Only years with a probability above 1% are shown.

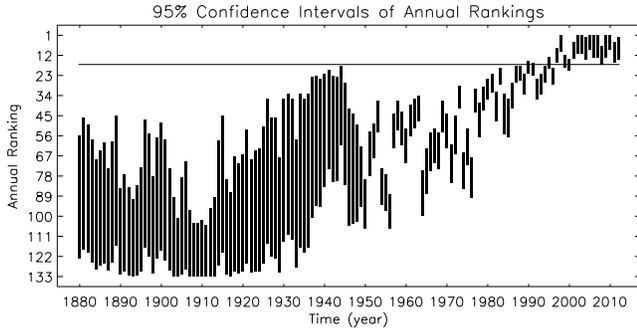


Figure 2. The 95% confidence intervals for each annual ranking of the NOAAATMP anomaly time series (1880–2012) using the independent approach. The horizontal black line indicates a ranking of seventeenth warmest, the lowest confidence interval value between 2001 and 2012.

(2001–2012) rank among the 14 warmest in the 133-year period of record.” Using the independent approach, we calculate a 12% probability that 2012 was the tenth warmest year on record. The two-tailed 95% confidence interval suggests that 2012 is likely ranked anywhere between the second and fourteenth warmest years (Figure 2). We calculate a 68% likelihood that all years between 2001 and 2012 were among the 14 warmest years (there is a 31% chance that only 11 of the years rank among the top 14). Of the 12 confidence intervals over 2001–2012 shown in Figure 2, the rankings all range between the first and seventeenth warmest.

[15] Although the three global temperature time series are by no means derived from independent data values [Hansen *et al.*, 2010], it is instructive to compare the probabilities generated using the independent approach for HadCRUT4 and GISTEMP as well (as with NOAAATMP, using the dependent approach on the HadCRUT4 and GISTEMP global temperature series does not materially alter the probabilities). The GISTEMP standard errors (Figure 1b) consist of three period averages (R. Reto, personal communication, 2013). HadCRUT4 characterizes uncertainty by reporting the 95% confidence intervals from its 100 ensemble members, and we estimate its standard error time series (Figure 1b) by presuming normality. HadCRUT4 and GISTEMP both show 2010 and 2005 as the warmest and second warmest years on record, respectively. As is the case for NOAAATMP, the highest probabilities (associated with the year 2010) are in the 30–40% range, shy of a majority of the simulations. While rankings 3–8 consist of the same set of years, their ordering varies across the data sets, most notably for the year 2007, which has the third highest probability in GISTEMP and eight highest in both NOAAATMP and HadCRUT4. For all but one of the top 8 years (2002), the NOAAATMP probability lies between the associated HadCRUT4 and GISTEMP probabilities. Similarly, the top 10 analysis (Table 3) shows that the years with probabilities above 1% consist of the same 14 years for all three data sets (all 14 have occurred since 1997), with the ordering of the probabilities being reasonably comparable, except once again for the year 2007. The overall story for all three data sets is the same: while we cannot irrefutably pinpoint which year was the warmest year on record through 2012, there is strong statistical evidence (e.g., Figure 2) that each individual year in the 1997–2012 period has been consistently and significantly warmer than the vast majority of all other years since 1880.

[16] We also repeated the independent approach uncertainty analysis for NOAAATMP after multiplying its standard error time series by two. Characterizing uncertainty is an inexact science by nature, and this simple exercise tests the sensitivity of the Monte Carlo analysis with respect to the magnitudes of the standard errors. As expected, hypothetically doubling the standard error time series would reduce the probability that 2010 was the warmest year on record (from 36% to 22%), and the years 2005 (from 28% to 18%) and 1998 (from 11% to 10%) would also show varying degrees of reduced probabilities. However, the vast majority of the hypothetical shift away from the top 3 years would go toward increasing the probabilities of other years in the 1997–2012 period. In fact, even after doubling the standard errors, there is a 97% (down from 100%) probability that the warmest year occurred between 1997 and 2012, and no year prior to 1997 reaches a 10% probability of having been among the 10 warmest years on record.

[17] As a final check on the sensitivity of our methodology, we repeated the independent approach uncertainty analysis for NOAAATMP using different end years between 1997 and 2012 (Table 4). Prior to 1997, the warmest year to date was 1995. However, 15 of the 17 years since 1995 exceeded the 1995 global temperature average in NOAAATMP (only 1996 and 2000 were less warm), and four of these years went on to become the warmest years on record, at least briefly. After 1997 came to a close, it became the warmest year on record with a ~92% probability and was eclipsed the very next year by 1998 which garnered a probability above 99%. Subsequently, 2005 and 2010 became the warmest years on record, and due to smaller separations between them, the probabilities have come down. This demonstrates that the probabilities can change a great deal with the addition of just a single year. Another dramatic example of this was the U.S. temperature time series after the addition of 2012, which was significantly warmer than any previous year on record according to GK13 and NCDC (<http://www.ncdc.noaa.gov/sotc/national/2012/13>).

Table 4. The Probability (%) That a Year was the Warmest on Record (Since 1880) in the NOAAATMP Time Series, Calculated Using the Independent Years Assumption and Variable End Years Between 1997 and 2012^a

End Year	Ranking #1		Ranking #2		Ranking #3	
	Year	Probability (%)	Year	Probability (%)	Year	Probability (%)
1997	1997	91.52	1995	7.82	-	-
1998	1998	99.25	-	-	-	-
1999	1998	99.23	-	-	-	-
2000	1998	99.23	-	-	-	-
2001	1998	94.40	2001	5.08	-	-
2002	1998	63.07	2002	34.77	2001	2.03
2003	1998	44.13	2003	31.90	2002	23.03
2004	1998	41.44	2003	29.69	2002	21.33
2005	2005	47.89	1998	22.52	2003	15.47
2006	2005	44.94	1998	20.51	2003	13.96
2007	2005	43.20	1998	19.30	2003	13.03
2008	2005	43.18	1998	19.28	2003	13.02
2009	2005	41.14	1998	17.96	2003	12.05
2010	2010	36.94	2005	28.44	1998	10.85
2011	2010	36.91	2005	28.42	1998	10.83
2012	2010	36.54	2005	28.42	1998	10.63

^aOnly years with a probability above 1% are shown. Years shown in bold indicate years that at one point in time were the warmest years to date.

[18] Given the potential for annual ranking uncertainty estimates to vary from year to year, which can be particularly volatile when annual values persist near record levels as we have seen over the last two decades, articulating an uncertainty range alongside an annual ranking makes the climate scientist's already formidable communications challenge even more difficult. However, stakeholders are better served by, and often clamor for, a more thorough accounting of climatic conditions. This entails climate monitoring centers to provide not only a historical perspective of the most recent annual or monthly observation but also the context of the uncertainty inherent in that historical perspective.

[19] **Acknowledgments.** We kindly thank P. Guttorp for providing his R code and engaging us on this topic. We also acknowledge R. Reto and P. Jones for their fruitful discussions. Lastly, we wish to thank P. Thorne and B. Huang for their valuable insights while reviewing earlier versions of this manuscript.

[20] The Editor thanks two anonymous reviewers for their assistance in evaluating this paper.

References

- Guttorp, P., and T. Y. Kim (2013), Uncertainty in ranking the hottest years of US surface temperatures, *J. Clim.*, *26*, 6323–6328.
- Hansen, J., R. Ruedy, M. Sato, and K. Lo (2010), Global surface temperature change, *Rev. Geophys.*, *48*, RG4004, doi:10.1029/2010RG000345.
- Huang, N. E., and S. S. P. Shen (Eds.) (2005), *Hilbert-Huang Transform and its Applications*, World Sci., Singapore.
- Kennedy, J. J., N. A. Rayner, R. O. Smith, M. Saunby, and D. E. Parker (2011), Reassessing biases and other uncertainties in sea-surface temperature observations since 1850: 2. Biases and homogenization, *J. Geophys. Res.*, *116*, D14104, doi:10.1029/2010JD015220.
- Kwiatkowski, D., P. C. B. Phillips, P. Schmidt, and Y. Shin (1992), Testing the null hypothesis of stationarity against the alternative of a unit root, *J. Econometrics*, *54*, 159–178.
- 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.
- Phillips, P. C. B., and P. Perron (1988), Testing for a unit root in time series regression, *Biometrika*, *75*, 335–346.
- R Core Team (2013), *R: A Language and Environment for Statistical Computing*, R Found. for Stat. Comput., Vienna, Austria.
- Smith, T. M., and R. W. Reynolds (2005), A global Merged Land-Air-Sea Surface Temperature reconstruction based on historical observations (1880–1997), *J. Clim.*, *18*, 2021–2036.
- Smith, T. M., R. W. Reynolds, T. C. Peterson, and J. Lawrimore (2008), Improvements to NOAA's historical merged land-ocean surface temperature analysis (1880–2005), *J. Clim.*, *21*, 2283–2296.
- Smith, T. S., and R. W. Reynolds (2002), Bias corrections for historical sea surface temperatures based on marine air temperatures, *J. Clim.*, *15*, 73–87.
- Vose, R. S., et al. (2012), NOAA's merged land-ocean surface temperature analysis, *Bull. Am. Meteorol. Soc.*, *93*, 1677–1685.
- Wang, J., Z. Yan, P. D. Jones, and J. Xia (2013), On “observation minus reanalysis” method: A view from multidecadal variability, *J. Geophys. Res. Atmos.*, *118*, 7450–7458, doi:10.1002/jgrd.50574.
- Wilks, D. S. (2006), *Statistical Methods in the Atmospheric Sciences*, Int. Geophys. Ser., vol. 59, 2nd ed., Academic, Burlington, Mass.