BAMS
In Box

Should We Expect Each Year in the Next Decade (2019–28) to Be Ranked among the Top 10 Warmest Years Globally?

Anthony Arguez, Shannan Hurley, Anand Inamdar, Laurel Mahoney, Ahira Sanchez-Lugo, and Lilian Yang

Annual rankings of global temperature are widely cited by media and the general public, not only to place the most recent year in a historical perspective, but also as a first-order metric of recent climate change that is easily digestible by the general public. According to observations from NOAA's Merged Land Ocean Global Surface Temperature Analysis Dataset (NOAAGlobalTemp) 5.0, the year 2018 was the fourth warmest year on record globally since 1880 (see Table 1). The most recent 5 years in the record (2014–18) comprise the 5 warmest years on record, with 2016 currently ranking as the warmest year. Given this streak of record or near-record global warmth in recent years, should we expect each year in the next decade (2019–28) to be ranked among the top 10 warmest years globally? In other words, given historical observations (including the most recent ones), can we assume that near-record annual rankings are already "baked into the cake" for the next several years?.

To answer these questions, we analyze the monthly version of NOAAGlobalTemp to select a methodology for projecting the end-of-year (i.e., "monitoring year") annual global ranking as well as the rankings for the subsequent 9 years (i.e., years 2–10 or "outlook years"). We utilize autoregressive modeling with and without extension of the long-term trend, and also consider an adjustment based on real-time ENSO conditions. In addition, to address the propensity for virtually every recent year to be ranked (at least initially) as a top 10 warmest year, we introduce a "temperature score" product that will allow NOAA to better communicate the coolness or warmth of a recent year versus the long-term trend. To the best of our knowledge, no such projected ranking or temperature score products are currently produced operationally by any other major climate monitoring center around the world, nor are we aware of any research documenting the global annual ranking errors associated with projected rankings of any type. It is our expectation that these new tools will improve the communication of climate change impacts to the general public.

https://doi.org/10.1175/BAMS-D-19-0215.1
Corresponding author: Anthony Arguez, anthony.arguez@noaa.gov
In final form 25 January 2020
©2020 American Meteorological Society
For information regarding reuse of this content and general copyright information, consult the AMS Copyright Policy.

Unauthenticated | Downloaded 12/18/20 10:59 AM UTC

AFFILIATIONS: Arguez and Sanchez-Lugo—NOAA/National Centers for Environmental Information, Asheville, North Carolina; **Hurley**—Old Dominion University, Norfolk, Virginia; **Mahoney**—Fairfax, Virginia; **Inamdar**—North Carolina Institute for Climate Studies, Asheville, North Carolina; **Yang**—California State University, Long Beach, Long Beach, California

Methodology

Data. We utilize the monthly operational version of global surface temperature time series produced by the NOAAGlobalTemp analysis. NOAAGlobalTemp includes land-based near-surface air temperatures from the Global Historical Climatology Network Monthly dataset (GHCN-M; version 4.0.1) and sea surface temperatures (SSTs) from the Extended Reconstructed Sea Surface Temperature dataset (ERSST; version 5). The operational monthly time series are accessible online.[1] Prior to mid-2019, the operational version of NOAAGlobalTemp was based on GHCN-M version 3.3.0 and ERSST version 4. We repeated our analyses using this previous operational version of NOAAGlobalTemp, and the results were virtually indistinguishable from those presented here. We also utilize the oceanic Niño index (ONI), a 3-month running average of SST anomalies in the Niño 3.4 region. The current operational version of ONI is produced by NOAA's Climate Prediction Center using ERSST, version 5, and is also accessible online.[2]

[1] www.ncei.noaa.gov/data/noaa-global-surface-temperature/access/operational/timeseries

[2] www.cpc.ncep.noaa.gov/data/indices/oni.ascii.txt

Projected rankings. The purpose of our experiment is to select a projected ranking algorithm for operational use. We begin by utilizing the monthly global land-only and ocean-only NOAAGlobalTemp time series from January 1975 through December 2018. We chose the 1975–2018 period of record because it provides a 40+-yr baseline for estimating month-to-month fluctuations that are likely to be representative of real-time fluctuations, given the largely unabated and stable upward trend since the mid-1970s. Moreover, all annual NOAAGlobalTemp anomalies from 1880 (the earliest reading available) through the mid-1970s are well below anomalies of the top 10 warmest years in Table 1, even when considering the uncertainty of the NOAAGlobalTemp time series values. Our results exhibit improved performance when modeling the land-only and ocean-only series separately and subsequently merging the simulations (using effective land–ocean proportions of 27.4%/72.6% as derived from the 1975–2018 time series) versus modeling the merged land+ocean time series; only this bifurcated modeling approach is considered henceforth.

The period 1999–2018 is used as a 20-yr reforecasting period over which we calculate various error statistics compared to observed annual global temperature rankings and anomalies; these error statistics will provide the basis for selecting which algorithm to use

Table 1. The 10 warmest years on record (1880–2018) in the NOAAGlobalTemp dataset. Anomalies (°C) are relative to the 1971–2000 base period. The running ranking is the ranking based on the period of record from 1880 through the year in question. The temperature score (a measure of relative annual warmth or coolness) is based on the 1975–2018 period.

Ranking	Year	Anomaly	Running ranking	Temperature score
1	2016	0.675	1	10
2	2015	0.615	1	10
3	2017	0.590	3	9
4	2018	0.507	4	5
5	2014	0.423	1	4
6	2010	0.409	1	7
7	2013	0.359	2	2
8	2005	0.355	1	8
9	1998	0.340	1	10
10	2009	0.330	3	4

Unauthenticated | Downloaded 12/18/20 10:59 AM UTC

from a handful of options. For each monthly prediction step in the 240-month (i.e., 20 year) reforecasting period, residuals are calculated by removing the ordinary least squares (OLS) trend from the monthly (land only or ocean only) NOAAGlobalTemp anomalies from January 1975 through the most recent month of observed data. This is a necessary step because of the nonstationarity of the global monthly temperature record. Following a Monte Carlo approach with 10,000 simulations per scenario (i.e., prediction month), the projected residuals of 1 to 120 months are then computed via autoregressive (AR) modeling using the Bayesian information criterion. For example, simulations for January 1999 through December 2008 are based on NOAAGlobalTemp residuals from January 1975 through December 1998; in this example, December 1998 represents the most recent month of observed data, which would have become available operationally in January 1999. This overall Monte Carlo simulation approach is repeated for each monthly time step from January 1999 through December 2018.

Three different variations of this methodology are examined. In the first case, denoted "AR without trend extension," we add back in the OLS trend for observed years without extending the trend through future values, effectively imposing a mean trend of zero over the 10-yr forecast period. In the second case, denoted "AR with trend extension," the trend line is extended through future values, effectively assuming that the mean observed trend continues at the same rate during the 10-yr forecast period. Comparing the results from these two cases allows us to characterize the skill associated with the trend itself. Following the example above with data through December 1998, the AR without (with) trend extension approach determines residuals from the OLS trend from January 1975 through December 1998, simulates month-to-month evolutions for January 1999 through December 2008 via autoregressive analysis of these residuals, and adds back the OLS trend without (with) extending the trend through the January 1999 to December 2008 period. Last, we also test the "AR+ENSO" case, which is equivalent to the AR with trend extension approach, but offset by an adjustment to the simulated residuals based on the most recent ONI value(s). This adjustment is determined by correlation analysis, which shows maximal correlation between the most recent ONI value and the running mean of NOAAGlobalTemp residuals at lags of 1–4 months.

The land-only and ocean-only simulations are then merged into land+ocean simulations. Annual averages and rankings are then computed from the monthly simulations (plus observed months for the monitoring year) and are compared to the observed averages/rankings for 1999–2018. From these differences, we compute the mean absolute rank errors and the mean absolute simulation (temperature) errors for each of the three cases and for various lead time ranges. The 95% prediction intervals are also reported, from which we calculate the prediction interval accuracy and the average prediction interval widths.

Previous studies (see Additional resources) have shown that, in addition to ENSO, the Arctic Oscillation (AO) and the Atlantic multidecadal oscillation (AMO) can contribute to suitable reconstruction of global surface temperatures. However, the effects are small over the 1975–2018 period, especially as compared to ENSO's impact. Their effects, along with the episodic effects associated with major volcanic eruptions, are only indirectly accounted for in our methodology in that the autoregressive relationships are influenced by these factors. Moreover, other recent studies have shown that simple autoregressive-based modeling of detrended residuals of annual U.S. and global time series produce robust estimates for quantifying ranking uncertainties. Therefore, we only consider the NOAAGlobalTemp and ONI series in the present investigation, which is adequate for modeling global annual rankings.

In this study, we are focused on the "running ranking," which we define as the ranking an individual year attains when it is first ranked versus all prior years (see Table 1). For example, the year 1998 was at one point in time the warmest year on record, and therefore registered a

Unauthenticated | Downloaded 12/18/20 10:59 AM UTC

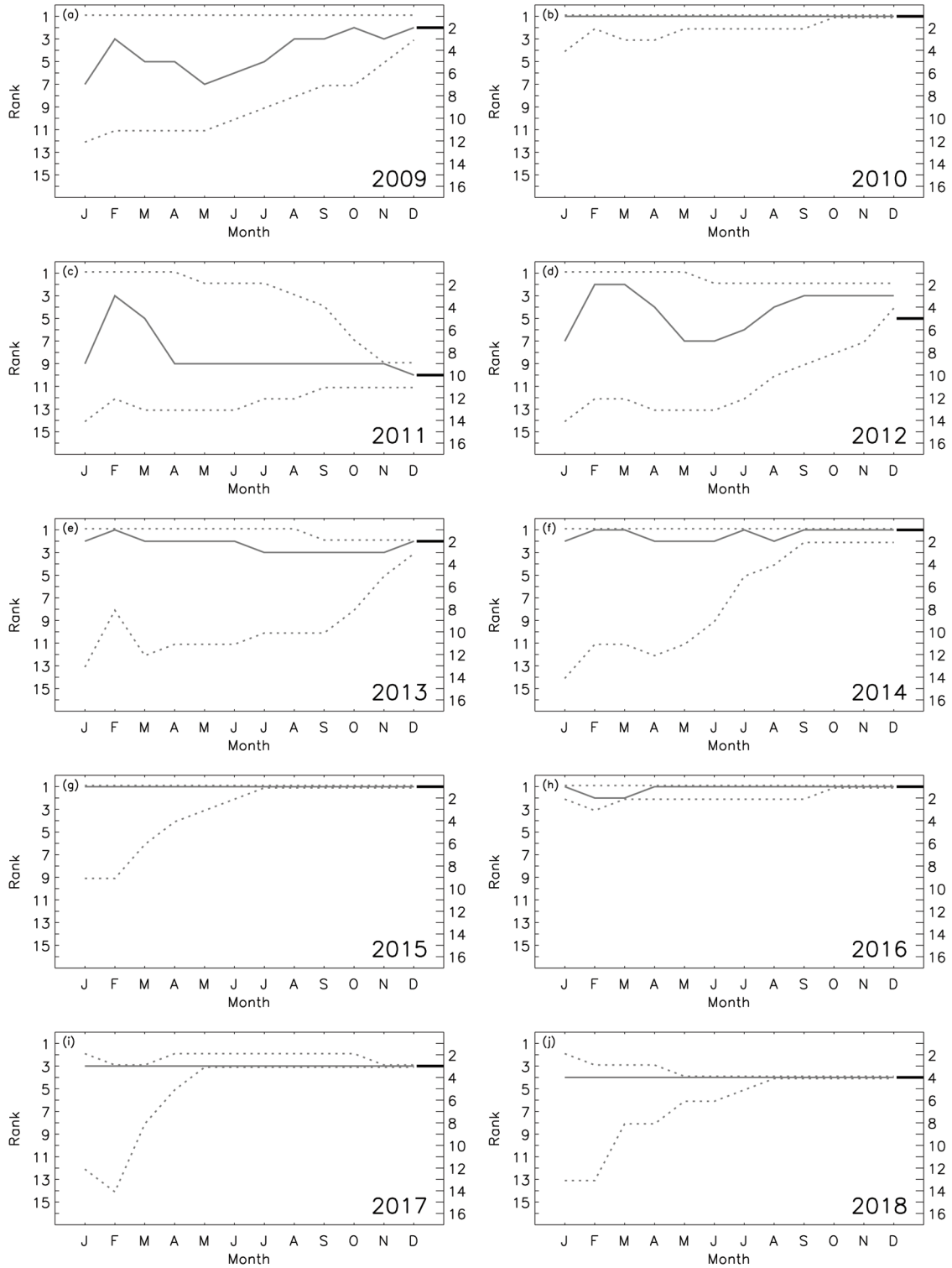

Fig. 1. (a)–(j) Monthly evolution of the median projected ranking (solid gray curve) and the 95% prediction interval (dashed gray curve) during the monitoring year for individual years between 2009 and 2018. The month indicates the month during which the projection was performed using monthly data through the previous month. The black tick mark indicates the observed end-of-year ranking.

running ranking of 1 using data through 1998, although its record has been eclipsed several times and exhibits a "retrospective ranking" of 9 using the current operational dataset through 2018. Similarly, our results are a snapshot given the existing methodology used to effectuate NOAAGlobalTemp. While we expect the algorithm's performance to be largely independent of any changes made to the way that NOAAGlobalTemp (or any other annual global temperature

Unauthenticated | Downloaded 12/18/20 10:59 AM UTC

Table 2. Skill statistics associated with the AR without trend extension, AR with trend extension, and AR+ENSO approaches based on a 1999–2018 reforecasting period.

Lead time →		Year 1 (months)				Year 2	Year 5	Year 10
Method Statistics ↓		1–3	4–6	7–9	10–12			
AR without trend extension	Mean absolute rank error	0.4	0.6	1.1	1.2	1.8	2.6	5.9
	Mean absolute simulation error (°C)	0.011	0.019	0.033	0.048	0.067	0.103	0.156
	Prediction interval accuracy (%)	98.3	100.00	100.0	100.0	98.2	93.8	90.9
	Average prediction interval width	1.8	4.1	6.4	9.0	10.9	14.3	18.4
AR with trend extension	Mean absolute rank error	0.4	0.6	1.1	1.2	1.8	2.1	2.2
	Mean absolute simulation error (°C)	0.011	0.019	0.033	0.046	0.065	0.078	0.084
	Prediction interval accuracy (%)	98.3	100.0	100.0	100.0	100.0	100.0	100.0
	Average prediction interval width	1.8	4.1	6.3	8.6	10.0	11.1	10.8
AR+ENSO	Mean absolute rank error	0.4	0.7	0.9	1.1	1.8	2.1	2.2
	Mean absolute simulation error (°C)	0.011	0.021	0.036	0.055	0.065	0.079	0.084
	Prediction interval accuracy (%)	98.3	100.0	100.0	100.0	100.0	100.0	100.0
	Average prediction interval width	1.9	4.3	6.7	9.0	10.1	11.1	10.9

time series) is calculated, we do envision monitoring the algorithm's performance and proposing future fine tuning of the algorithm if warranted.

Annual temperature scores. Since virtually all newly ended years since 1988 would have had a top 10 running ranking, it would be useful to distinguish between warmer and colder years relative to the sustained long-term trend for communicating climate monitoring impacts. For example, the years 2008 and 2011 were considerably cooler than surrounding years and below the overall trend line, whereas 1998 and 2016 were not only considered the warmest years on record when first reported, but their values were also warmer than surrounding years. We propose a simple new annual "temperature score" algorithm that provides such context of natural variability relative to long-term trends, and complements the associated ranking.

First, the OLR trend was removed from the annual land+ocean NOAAGlobalTemp time series from 1975 to 2018 to identify residuals. These residuals are then divided by the group standard deviation to arrive at values analogous to standard scores (or z scores). Finally, these standard scores are transformed to a scale from 1 to 10, such that each score had a 10% probability of occurrence based on the cutoff values of the Gaussian distribution. These temperature scores provide a real-time perspective relative to the long-term trend, with a value of 1 representing a very cold year and a value of 10 representing a very warm year.

Unauthenticated | Downloaded 12/18/20 10:59 AM UTC

Results

Projected rankings. During the monitoring year, all three methods perform similarly (see Table 2). While the AR+ENSO approach appears to slightly outperform in terms of rank errors during lead times of 7–12 months, that is, January–June of the monitoring year, the differences are not statistically significant. Similarly, the AR with trend extension approach (and the AR without trend extension approach to a lesser extent) appears to slightly outperform the AR+ENSO approach in terms of simulation error and prediction interval width, but again the differences are not statistically significant.

During outlook years, there are no appreciable differences between the AR+ENSO and the AR with trend extension approaches. The mean absolute rank errors remain near 2.0 from lead times of 2–10 years, whereas the mean absolute simulation errors vary from about 0.065°C at year 2 to about 0.084°C at year 10. However, both of these methods far outperform the AR without trend extension approach for years 5 through 10, suggesting a high degree of predictability associated with extrapolation of the instantaneously determined trend. By year 10, the mean absolute rank error of the AR without trend extension approach is 5.9 and the mean absolute simulation error is about 86% higher than in the other approaches. Moreover, even as the prediction intervals widen considerably with lead time, the accuracy of the prediction intervals degrades with an error rate of ~6% at year 5 and ~9% at year 10, whereas the approaches with extrapolation retain an error rate at or near 0%.

Figure 1 shows the evolution of projected (median) rankings and the prediction interval during the monitoring year for 2009–18 for the AR with trend extension approach. In half of the cases, the expected (median) end-of-year ranking was correctly forecast at lead times of 9–12 months. The observed end-of-year ranking fell within the prediction interval, and matched the projected ranking determined in December (using data through November), in all cases but one: December 2012. This case was associated with a rather large month-over-month change of about −0.3°C from November to December 2012.

Annual temperature scores. Figure 2 shows the annual temperature scores for the globe from 1975 to 2018. Notably, the very strong El Niño events of 1982/83, 1997/98, and 2015/16 are associated with temperature scores of 10 (in 1983, 1998, and 2016, respectively), indicating exceptionally warm years relative to the trend. In contrast, the strong La Niña events of 2007/08 and 2010/11 are both associated with temperature scores of 1 (in 2008 and 2011, respectively). Although 2008 and 2011 were cold years relative to the trend, they were over 0.2°C warmer than 1983 (which exhibits a running ranking of first warmest and temperature score of 10) and are warmer than all years prior to 1998. More recently, the years

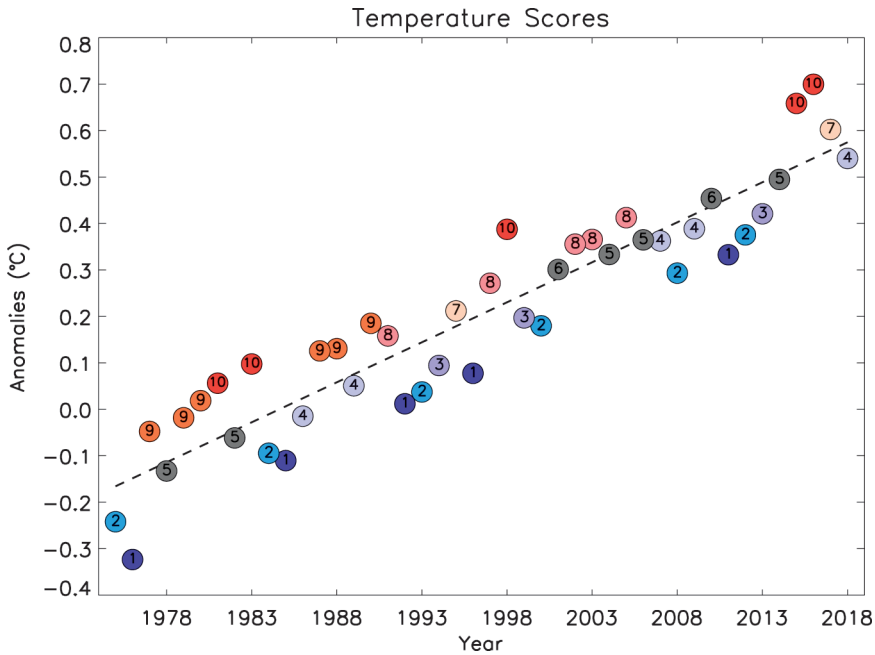

Fig. 2. Global annual temperature scores based on the annual NOAAGlobalTemp time series from 1975 to 2018. The ordinate indicates the NOAAGlobalTemp anomaly relative to the 1971–2000 base period. The inscribed value indicates the temperature score. Blue shading indicates "cool" scores from 1 to 4, gray shading indicates "near neutral" scores from 5 to 6, and red shading indicates "warm" scores from 7 to 10. The dashed line indicates the OLS regression line from 1975 to 2018.

Unauthenticated | Downloaded 12/18/20 10:59 AM UTC

2014 and 2013 were initially ranked as the first and second warmest years on record, yet their corresponding temperature scores are 4 and 2, respectively, both lying on the colder side of the trend line. Thus, Fig. 2 graphically demonstrates how what we consider to be a "warm year" or "cold year" has changed over time.

We performed sensitivity analysis of the temperature score metric to determine whether the metric is stable. This was verified by repeating the temperature score calculation incrementally over the last 20 years (1999–2018). Thus, the temperature score for 1999 was computed 20 times—first using data from 1975 to 1999, then 1975 to 2000, and so forth through the 1975–2018 computation—whereas only one temperature score was computed for 2018, with a linearly decreasing sample size in between years 1999 and 2018. Overall, the median absolute year-over-year score difference is ~0.15. Moreover, no year-over-year temperature score change exceeds ±1, and no individual year's score range over the 1999–2018 test period is greater than 2, suggesting that the temperature score metric is indeed rather stable.

Discussion and conclusions

Based on our results, we propose using the AR with trend extension approach operationally for characterizing the annual ranking probabilities during the course of the monitoring year as well as for the outlook years. Projections of the next ten years using NOAAGlobalTemp data through December 2018 suggest a greater than 99% (a 75.3%) probability that most (all) of the years between 2019 and 2028 will also be top 10 warmest years (see Table 3) under the "running ranking" perspective. Notably, even when utilizing the AR without trend extension method, the results indicate a strong likelihood that most (>99% probability) of these years will be among the top 10 warmest years, and there is an 82.0% probability that all 10 years will rank in the top 15 warmest years. Thus, accounting for historical month-to-month variability in global surface temperatures, it would likely take an abrupt climate shift for even a few years within the next decade to register outside the top 10 warmest years. This is a testament to the exceptional warmth experienced over the last few decades, punctuated by the last 4 years (2015–18), which have separated themselves from "the pack."

Although global temperatures are generally trending upward over the past several decades (exhibiting a high signal-to-noise ratio), there are still meaningful fluctuations of global surface temperature associated with natural variability and variations in anthropogenic forcing, especially over decadal or intradecadal periods (lower signal-to-noise ratio). Given the strong likelihood for future years to remain near record levels, we recommend that global monitoring analyses incorporate the temperature score to better communicate the differentiation of warmer and colder years relative to the long-term trend. Taken in tandem, the new approaches for temperature scores and projected rankings provide the general public with additional context for characterizing recent and expected global temperature conditions.

Table 3. Projected (median) rankings for 2019–28 based on the NOAAGlobalTemp dataset from 1975 to 2018 for the AR without trend extension and AR with trend extension methods. The 95% prediction intervals are indicated by the 2.5% and 97.5% levels of the Monte Carlo simulations.

	AR without trend extension			AR with trend extension		
Year	2.5% level	Median	97.5% level	2.5% level	Median	97.5% level
2019	1	4	7	1	4	7
2020	1	5	8	1	4	8
2021	1	5	10	1	4	9
2022	1	6	12	1	4	9
2023	1	6	13	1	4	9
2024	1	7	14	1	4	10
2025	1	7	15	1	4	11
2026	1	8	16	1	3	12
2027	1	8	17	1	3	12
2028	1	9	18	1	3	13

Unauthenticated | Downloaded 12/18/20 10:59 AM UTC

Additional resources

For examples of global annual temperature rankings utilized in climate monitoring reporting and in general media, see NOAA (2019) and Schwartz and Popovich (2019), respectively. The description of NOAAGlobalTemp and its constituent components can be found in Vose et al. (2012), Menne et al. (2018), Huang et al. (2017), Lawrimore et al. (2011), Huang et al. (2015), Liu et al. (2015), and Huang et al. (2016). More information about the Niño 3.4 region (used to construct the ONI) is given in Bamston et al. (1997). For more information on the stability of the upward trend in the global annual temperature time series since the mid-1970s, please see Karl et al. (2015) and Lewandowsky et al. (2015). Arguez et al. (2013) analyze the effects of statistical uncertainty on global annual temperature rankings, including comparisons with other major datasets. For more information regarding standard autoregressive modeling with the Bayesian information criterion, please consult Wilks (2006). Folland et al. (2013, 2018) describe how ENSO, AO, AMO, volcanic eruptions, and other factors contribute to suitable reconstruction of global surface temperatures time series. Guttorp and Kim (2013) and Arguez et al. (2013) show that autoregressive-based modeling provides accurate and robust estimates of ranking uncertainties for U.S. and global temperature time series, respectively. Santer et al. (2011) quantify the impact of time scale on signal-to-noise ratios of air temperature.

Acknowledgments. The authors kindly acknowledge Karin Gleason, Derek Arndt, Ellen Mecray, Alec Courtright, and Art DeGaetano for fruitful discussions and several anonymous reviewers for their thoughtful comments. Inamdar was supported by NOAA through the Cooperative Institute for Climate and Satellites–North Carolina under Cooperative Agreement NA14NES432003. This project was partially funded by NOAA by virtue of a Presidential Early Career Award for Scientists and Engineers Grant and via partnership with the NASA DEVELOP program.

Unauthenticated | Downloaded 12/18/20 10:59 AM UTC

References

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, https://doi.org/10.1002/2013GL057999.

Bamston, A. G., M. Chelliah, and S. B. Goldenberg, 1997: Documentation of a highly ENSO-related SST region in the equatorial Pacific. *Atmos.–Ocean*, **35**, 367–383, https://doi.org/10.1080/07055900.1997.9649597.

Folland, C. K., A. W. Colman, D. M. Smith, O. Boucher, D. E. Parker, and J.-P. Vernier, 2013: High predictive skill of global surface temperature a year ahead. *Geophys. Res. Lett.*, **40**, 761–767, https://doi.org/10.1002/grl.50169.

——, O. Boucher, A. W. Colman, and D. E. Parker, 2018: Causes of irregularities in trends of global mean surface temperature since the late 19th century. *Sci. Adv.*, **4**, eaao5297, https://doi.org/10.1126/sciadv.aao5297.

Guttorp, P., and T. Y. Kim, 2013: Uncertainty in ranking the hottest years of U.S. surface temperatures. *J. Climate*, **26**, 6323–6328, https://doi.org/10.1175/JCLI-D-12-00760.1.

Huang, B., and Coauthors, 2015: Extended Reconstructed Sea Surface Temperature version 4 (ERSST.v4): Part I. Upgrades and intercomparisons. *J. Climate*, **28**, 911–930, https://doi.org/10.1175/JCLI-D-14-00006.1.

——, and Coauthors, 2016: Further exploring and quantifying uncertainties for Extended Reconstructed Sea Surface Temperature (ERSST) version 4 (v4). *J. Climate*, **29**, 3119–3142, https://doi.org/10.1175/JCLI-D-15-0430.1.

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

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

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 monthly mean temperature data set, version 3. *J. Geophys. Res.*, **116**, D19121, https://doi.org/10.1029/2011JD016187.

Lewandowsky, S., J. S. Risbey, and N. Oreskes, 2015: On the definition and identifiability of the alleged "hiatus" in global warming. *Sci. Rep.*, **5**, 16 784, https://doi.org/10.1038/srep16784.

Liu, W., and Coauthors, 2015: Extended Reconstructed Sea Surface Temperature version 4 (ERSST.v4): Part II. Parametric and structural uncertainty estimations. *J. Climate*, **28**, 931–951, https://doi.org/10.1175/JCLI-D-14-00007.1.

Menne, M. J., C. N. Williams, and B. E. Gleason, 2018: The Global Historical Climatology Network monthly temperature dataset, version 4. *J. Climate*, **31**, 9835–9854, https://doi.org/10.1175/JCLI-D-18-0094.1.

NOAA, 2019: State of the climate: Global climate report for March 2019. National Centers for Environmental Information, accessed 6 May 2019, www.ncdc.noaa.gov/sotc/global/201903.

Santer, B. D., and Coauthors, 2011: Separating signal and noise in atmospheric temperature changes: The importance of timescale. *J. Geophys. Res.*, **116**, D22105, https://doi.org/10.1029/2011JD016263.

Schwartz, J., and N. Popovich, 2019: It's official: 2018 was the fourth-warmest year on record. *New York Times*, 6 May, www.nytimes.com/interactive/2019/02/06/climate/fourth-hottest-year.html.

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

Wilks, D. S., 2006: *Statistical Methods in the Atmospheric Sciences*. 2nd ed. International Geophysics Series, Vol. 59, Academic Press, 627 pp.

Unauthenticated | Downloaded 12/18/20 10:59 AM UTC