



POTSDAM-INSTITUT FÜR
KLIMAFOLGENFORSCHUNG

Originally published as:

Mann, M. E., Miller, S. K., [Rahmstorf, S.](#), Steinman, B. A., Tingley, M. (2017): Record temperature streak bears anthropogenic fingerprint. - Geophysical Research Letters, 44, 15, 7936-7944.

DOI: <https://doi.org/10.1002/2017GL074056>



RESEARCH LETTER

10.1002/2017GL074056

Key Points:

- The sequence of record-breaking temperatures in 2014–2016 had a negligible likelihood of occurrence in the absence of anthropogenic warming
- It was still a rare but not implausible event (roughly 1–3% likelihood) taking anthropogenic warming into effect
- The probability that three consecutive records would have been observed at some point since 2000 is ~40–80% given anthropogenic warming

Supporting Information:

- Supporting Information S1

Correspondence to:

M. E. Mann,
mann@psu.edu

Citation:

Mann, M. E., S. K. Miller, S. Rahmstorf, B. A. Steinman, and M. Tingley (2017), Record temperature streak bears anthropogenic fingerprint, *Geophys. Res. Lett.*, *44*, 7936–7944, doi:10.1002/2017GL074056.

Received 10 JAN 2017

Accepted 14 JUL 2017

Accepted article online 10 AUG 2017

Published online 12 AUG 2017

Record temperature streak bears anthropogenic fingerprint

Michael E. Mann¹ , Sonya K. Miller¹ , Stefan Rahmstorf², Byron A. Steinman³ , and Martin Tingley¹

¹Department of Meteorology, Pennsylvania State University, University Park, Pennsylvania, USA, ²Earth System Analysis, Potsdam Institute for Climate Impact Research, Potsdam, Germany, ³Department of Earth and Environmental Sciences and Large Lakes Observatory, University of Minnesota, Duluth, Minnesota, USA

Abstract We use a previously developed semiempirical approach to assess the likelihood of the sequence of consecutive record-breaking temperatures in 2014–2016. This approach combines information from historical temperature data and state-of-the-art historical climate model simulations from the Coupled Model Intercomparison Project Phase 5 (CMIP5). We find that this sequence of record-breaking temperatures had a negligible (<0.03%) likelihood of occurrence in the absence of anthropogenic warming. It was still a rare but not implausible event (roughly 1–3% likelihood) taking anthropogenic warming into effect. The probability that three consecutive records would have been observed *at some point* since 2000 is estimated as ~30–50% given anthropogenic warming and <0.7% in its absence. The likelihood of observing the specific level of record warmth recorded during 2016 is no more than ~one-in-a-million neglecting anthropogenic warming, but as high as 27%, i.e., a nearly one-in-three chance of occurrence taking anthropogenic warming into account.

1. Introduction

The year 2016, aided in part by a historically large El Niño event, set a new global temperature record. We have thus witnessed three consecutive record breaking annual mean temperatures (2014, 2015, and 2016) in most global and/or hemispheric surface temperature series for the first time since historical observations began in the nineteenth century. It is reasonable to suspect that such an event would be extremely unlikely in the absence of anthropogenic warming, but it is worthwhile to ask just how unlikely such events actually are both with and without anthropogenic influence on climate.

In previous work [Mann *et al.*, 2016a; henceforth M16] we analyzed the likelihood of the recent run of record warm years (9 of the 10 warmest and 13 of the 15 warmest years through 2015 having occurred since 2000), finding that both runs were highly unlikely (likelihood of less than one in 10,000) in the absence of anthropogenic warming and relatively likely in its presence. The study was motivated by press reports that greatly exaggerated how unlikely these events were (e.g., as small as one in 650 million) based on the flawed assumption of statistical independence of individual years. Nonetheless, the underlying scientific issues addressed in the study are of intrinsic scientific merit, highlighting the problem of how one properly accounts for statistical dependence in assessing the statistics of record temperature streaks.

In reality, neighboring years tend to be correlated with each other. Such *autocorrelation* leads to a greater likelihood of streaks of warm years. M16 employed a combination of model simulations and observations to account for the autocorrelation structure in surface temperature series in more rigorously estimating the likelihood of runs of record-warm temperatures. This approach differs from past studies [e.g., Zorita *et al.*, 2008; Craigmile *et al.*, 2014] which have used a purely empirical approach to address the statistics of record temperature streaks by using model simulations to explicitly account for the forced component of temperature change.

In this article, we use the same analysis method as M16 to investigate the likelihood of the unprecedented now-observed run of three consecutive global temperature records in 2014, 2015, and 2016. We assess the likelihood of this sequence of events both with and without the impact of anthropogenic warming. We also assess the likelihood of the more general phenomenon of three consecutive record-breaking years during the recent (post-1970) era, as well as the conditional likelihoods of a 2016 record given consecutive 2014 and 2015 records. Finally, we evaluate the likelihoods of the specific annual temperatures reached during 2014, 2015, and 2016.

2. Data and Methods

We employ the semiempirical approach of M16 which combines information from climate model simulations and observational temperature data in conjunction with a Monte Carlo resampling scheme to produce ensembles of surrogate global and hemispheric mean temperature series both with and without the influence of anthropogenic warming. While the reader is referred to M16 for methodological details, the methods and data used are also summarized below.

All raw data, ©Matlab code, and results from our analysis are available at the supplementary website: <http://www.meteo.psu.edu/~mann/supplements/GRLRecordTemps16>.

2.1. Observational Data

As in M16, we make use of two different observational surface temperature assessments: HadCRUT4 [Morice *et al.*, 2012] and GISTEMP [Hansen *et al.*, 2006] and analyze both the global and Northern Hemisphere (NH) annual mean temperature anomaly series through 2016. We focus on the 1880–2016 interval of overlap between the two data sets and center the anomalies on the long-term (1880–2016) mean.

These two widely used surface temperature data sets differ substantially in how missing data is treated. GISTEMP, which spatially interpolates across data gaps in the Arctic—the fastest warming region of the globe, shows the most warming in recent decades. HadCRUT4, by contrast, simply leaves out the missing regions (arguably, therefore, underestimating large-scale warming) and shows the least warming. As discussed in M16, the use of these two end-member data sets consequently allows us to assess the sensitivity of our results to basic spatial sampling considerations.

The years 2014, 2015, and 2016 constituted three consecutive records for both the global and NH mean and for both (GISTEMP and HadCRUT4) data sets with the exception of the HadCRUT4 NH mean for which 2015 and 2016 were statistically tied for warmest year.

2.2. Model Simulations

The model simulations analyzed in this study are taken from the Coupled Model Intercomparison Project Phase 5 (CMIP5) historical simulation experiments [Stocker *et al.*, 2013]. Estimates of the total forced component of global and Northern Hemisphere (NH) mean temperature are derived by averaging over the full ensemble of CMIP5 multimodel all-forcing historical experiments ($N = 163$ total realizations, $M = 40$ models), while estimates of the anthropogenic component are obtained by averaging over a separate ensemble of anthropogenic-only forcing experiments ($N = 40$ total realizations, $M = 8$ models) (see Table S1 in the supporting information). The natural forcing-only component is estimated by subtracting the anthropogenic series from the all-forcing series.

The multimodel mean series, as in M16, are computed as a simple arithmetic mean of ensemble members. The series were extended to 2016 using the protocol of Mann *et al.* [2016b] which involves alignment of the anthropogenic-only and all-forcing simulations during a common interval of zero mean natural forcing followed by linear forward extension of the anthropogenic-only component from 2005 to the terminal (2016) boundary (a type of “business as usual” extension) and persistence extension of the natural-only component. As with the observational temperature series, we focused on the interval 1880–2016 and defined anomalies relative to the long-term (1880–2016) mean.

There is a disparity between the surface temperatures actually measured by the observations (i.e., surface air temperature or “SAT” over land and sea surface temperature or “SST” over oceans) and the quantity that is instead often calculated from the CMIP simulations (i.e., SAT everywhere) [e.g., Stocker *et al.*, 2013]. The distinction is critical because SSTs are expected to warm less and more slowly than SATs. We thus instead employ a blended estimate of the CMIP5 model fields of SAT over land and SST over ocean. Although beyond the scope of our current analysis, a more sophisticated approach involves applying a time-dependent data mask that accounts for the time-evolving sea ice/open water boundary [Cowtan *et al.*, 2015]. This approach would be a useful extension of the methods presented here and should be pursued in future work.

Finally, we employ a recent set of corrections that account for revisions to both the anthropogenic and natural radiative forcings prescribed in the CMIP5 historical experiments for the post-1966 period [Schmidt

et al., 2014]. The corrections lead to less simulated warming over the past two decades and better agreement between simulated and observed temperature changes over that time frame. M16 adopted these corrections in all comparisons. In the present study, we have chosen to employ alternative time series that both do and do not impose these corrections in an effort to assess the robustness of our findings with respect to the uncertainties in the simulated forced temperature response over the past decade.

2.3. Estimating Internal Variability

As in M16, we represent global and hemispheric mean temperature variations (measured as anomalies relative to an 1880–2016 base period) through a statistical model of the form

$$T = F + I \quad (1)$$

where $F = A + N$ is the total forced component of temperature change, with A representing the anthropogenic forced component and N representing the natural (volcanic + solar) forced component of temperature change. I is the random internal variability (i.e., noise) component. The observations provide us with T , while climate model simulations provide us with an estimate of F , A , and N , as described in the preceding section. The difference $T - (A + N)$, i.e., the residual, can be interpreted as representing a single (i.e., the actual observed) realization of the pure internal variability component (I). It is appropriate to define a stationary stochastic time series model to represent I using parameters estimated from the residual series, as described below.

2.4. Monte Carlo Simulations

The internal variability component, I , can be modeled under a variety of null hypotheses, the simplest of which is the first-order autoregressive, i.e., AR(1), “red noise” process. For such a process, the effective degrees of freedom in tests of the mean are $N' = N [(1 - \rho)/(1 + \rho)]$, where N is the number of years of data, and $0 \leq \rho \leq 1$ is the lag 1 autocorrelation coefficient of the annual data series. For $\rho = 0$ we recover the result for an uncorrelated series $N' = N$, while as $\rho \rightarrow 1$, the effective degrees of freedom in the sample approaches $N' = 0$ regardless of the size of the sample.

Such a statistical model may not adequately account for additional sources of natural variability, e.g., oscillatory temperature variations due to phenomena such as El Niño–Southern Oscillation. We thus instead employ the more general linear autoregressive/moving average ARMA(p, q) process, which accommodates considerably more statistical structure, to characterize the natural variability (noise) in the temperature series. In our analysis, we take the true observational anomalies as fixed, but one could indeed generalize the procedure [e.g., Guttorp and Kim, 2013; Arquez *et al.*, 2013] to account for the uncertainty in the ranking/ordering of observed anomalies due to measurement/sampling uncertainty.

The orders p, q are selected by the conservative Bayesian Information Criterion (BIC) as in M16. Note that in some cases (i.e., when $p = 1, q = 0$ is selected), the chosen ARMA model reduces to AR(1). Tests of model adequacy were performed as described in M16 (see Figure S1).

We performed additional experiments based on the alternative, overly conservative model of “persistent red noise.” This model takes the form of an AR(1) process, but the autocorrelation coefficient ρ is instead in this case estimated not from the residual series but from the raw temperature series itself. Such a model produces higher levels of apparent “persistence” (i.e., considerably larger low-frequency fluctuations), because long-term anthropogenic warming is instead now treated as if it reflects natural low-frequency internal variability, inflating the apparent amplitude of long-term natural variability. The model can be argued to provide an upper bound on the noise amplitude and persistence.

We performed eight different experiments using both (1) NH GISTEMP and (2) NH HadCRUT4 with *no corrections* to CMIP5 forcings, (3) Global GISTEMP and (4) Global HadCRUT4 with *no corrections*, (5) NH GISTEMP and (6) NH HadCRUT4 *with corrections*, and (7) Global GISTEMP and (8) Global HadCRUT4 *with corrections*. We performed additional experiments to assess the sensitivity of the results to the means by which the CMIP5-simulated trends are extended from the 2005 terminal date to 2016, using an alternative assumption of simple persistence of the anthropogenic component beyond 2005.

For each experiment performed we generated a half million (500,000) ARMA Monte Carlo realizations and analyzed results based on both natural-only forcing and all-forcing simulations. For the persistent red

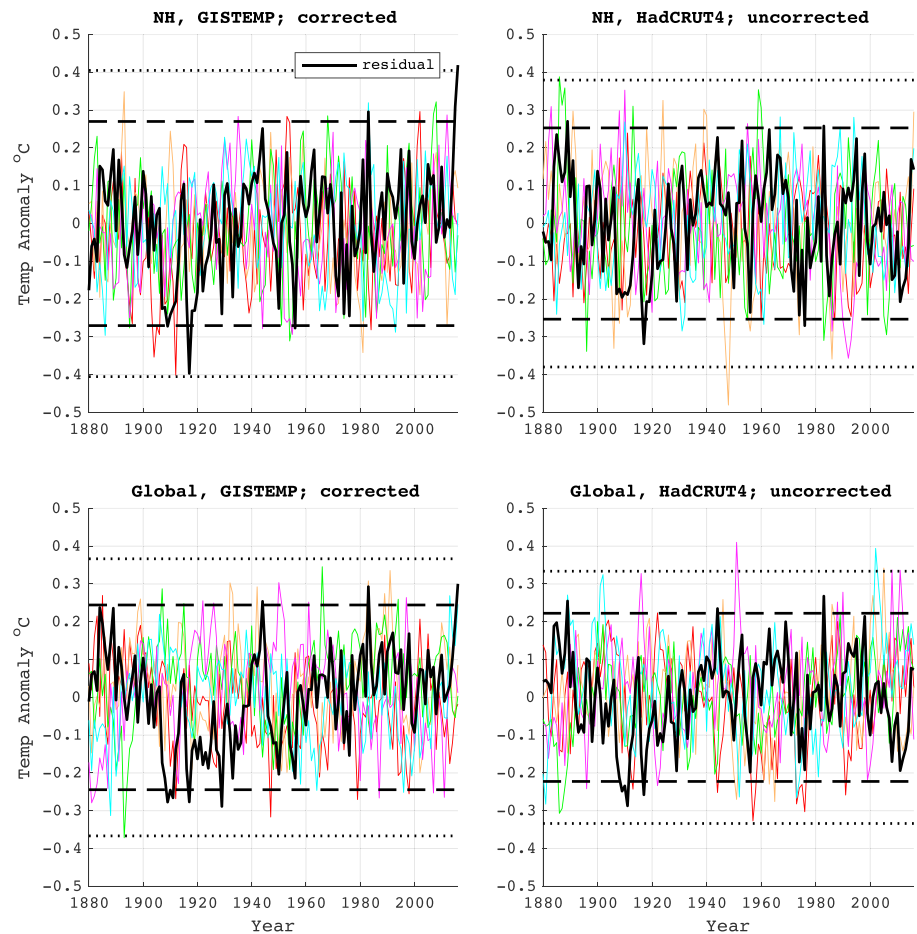


Figure 1. Estimated internal variability component (I) series (black) and five associated Monte Carlo surrogates (colored curves) based on (top left) NH GISTEMP observations and *corrected* CMIP5 multimodel series, (top right) NH HadCRUT4 observations and *uncorrected* CMIP5 multimodel series, (bottom left) as in Figure 1 (top left) but for global mean, (bottom right) as in Figure 1 (top right) but for global mean. Results for both NH and global mean temperature for both these and the other two cases (GISTEMP observations and uncorrected CMIP5 model series, HadCRUT4 series, and corrected CMIP5 model series) are available in the supporting information.

noise case, the statistical model depends only on the attributes of the raw observational temperature series, yielding four distinct experiments (one for each observational series).

Each Monte Carlo surrogate represents an alternative temperature history consistent with historical anthropogenic and natural forcing and a single realization of internal variability I . Figure 1 shows the residual series (i.e., the estimated historical series I) along with a selection of surrogates for I . The residual series is qualitatively similar to the surrogates as we expect it to be given the demonstrated adequacy (Figure S1) of the underlying statistical model. The 2016 value, associated with a major 2015/2016 El Niño event, is however somewhat of an outlier, exceeding $Z = +3$ standard deviations for at least one case (e.g., using the GISS NH series and employing forcing corrections), but within $Z = 2$ standard deviations for other cases (e.g., HadCRUT4 global series and simulations without forcing corrections). In all cases we can infer that a large positive El Niño-related internal temperature fluctuation contributed to the record 2016 warmth.

Using the full ensemble of plausible alternative temperature histories, we can tabulate how often particular records and series or runs of records occur. In each case, we collected statistics (see Table 1) on the frequency of the following events: reaching or exceeding the temperature levels for 2014 (1), 2015 (2), and 2016 (3), occurrence of the warmest year in 2016 (4), occurrence of the warmest two years in 2015 and 2016 (5), occurrence of the three warmest years in 2014, 2015, and 2016 (6), and simultaneous occurrence of conditions 4, 5, and 6, i.e., three consecutive records in 2014, 2015, and 2016.

Table 1. Estimated Likelihoods (in %)

	2014 ^a	2015 ^b	2016 ^c	1/1 ^d	2/2 ^e	3/3 ^f	1,2,3 (2016) ^g	2016(cond) ^h
Corrected forcings								
<i>Natural</i>								
NH GISTEMP [AR(1)]	< ⁱ	<	<	0.20	0.014	0.001	$2 \cdot 10^{-4}$	2.8
NH HadCRUT4 [AR(1)]	<	<	<	0.17	0.014	$6 \cdot 10^{-4}$	<	8.3
Glb GISTEMP [ARMA(1,1)]	<	<	<	0.24	0.026	0.006	0.001	7.2
Glb HadCRUT4 [AR(1)]	<	<	<	0.14	0.010	$8 \cdot 10^{-4}$	$4 \cdot 10^{-4}$	7.7
<i>Anthro</i>								
NH GISTEMP [AR(1)]	16	0.74	0.05	21	10	5.6	1.3	25
NH HadCRUT4 [AR(1)]	39	1.5	2.5	21	11	6.0	1.4	26
Glb GISTEMP [ARMA(1,1)]	19	5.0	0.63	21	11	6.8	1.8	32
Glb HadCRUT4 [AR(1)]	47	5.3	5.6	20	10	5.8	1.4	27
Uncorrected forcings								
<i>Natural</i>								
NH GISTEMP [AR(1)]	<	<	<	1.4	0.21	0.035	0.0062	6.0
NH HadCRUT4 [AR(1)]	<	<	<	1.4	0.19	0.032	0.0046	4.8
Glb GISTEMP [ARMA(1,1)]	<	<	<	1.6	0.38	0.13	0.027	14
Glb HadCRUT4 [AR(1)]	<	<	<	1.4	0.24	0.05	0.0084	6.7
<i>Anthro</i>								
NH GISTEMP [AR(1)]	43	5.5	0.78	25	14	9.7	2.4	29
NH HadCRUT4 [AR(1)]	71	9.4	14	26	15	10	2.5	29
Glb GISTEMP [ARMA(1,1)]	50	22	5.7	27	16	12	3.2	35
Glb HadCRUT4 [AR(1)]	81	26	27	26	15	11	2.9	32
Persistent red noise								
NH GISTEMP	0.18	0.011	$8 \cdot 10^{-4}$	1.21	0.54	0.29	0.074	23
NH HadCRUT4	0.20	0.006	0.007	1.20	0.54	0.27	0.067	22
Glb GISTEMP	0.33	0.086	0.021	1.4	0.75	0.47	0.14	30
Glb HadCRUT4	0.61	0.06	0.05	1.4	0.66	0.39	0.11	26
Constant Forcing/White Noise				0.7%	0.010	$4 \cdot 10^{-4}$	<	0.7

^aLikelihood of exceeding actual 2014 temperature.
^bLikelihood of exceeding actual 2015 temperature.
^cLikelihood of exceeding actual 2016 temperature.
^dLikelihood that 2016 is the warmest year.
^eLikelihood that 2015 and 2016 are the warmest two years.
^fLikelihood that 2014, 2015, and 2016 are the warmest three years.
^gLikelihood that 2014, 2015, and 2016 are consecutive record breakers.
^hConditional likelihood that 2016 is a record breaker given that 2014 and 2015 were both record breakers.
ⁱThe "less than" symbol denotes an estimated likelihood of $<2 \cdot 10^{-4}$ % (i.e., no occurrences in 500,000 realizations).

Finally, we can estimate, as a baseline for comparison, the likelihoods that would be expected given a stationary process with constant mean and statistically independent individual years. The likelihood that 2016 is the warmest year in a 137 year long series in this case is $1/137 = 0.7\%$, the likelihood that 2015 and 2016 are the two warmest years is, e.g., $2 \times 1/(136 \cdot 137) = 0.01\%$, and the likelihood of consecutive 2014, 2015, and 2016 records is, e.g., $1/(135 \cdot 136 \cdot 137) = 4 \times 10^{-5}\%$. These theoretical estimates are consistent with results obtained from Monte Carlo experiments using constant forcing and white noise surrogates.

3. Results and Discussion

The precise results of our analysis depend on the observational surface temperature product (GISTEMP versus HadCRUT4), whether recent forcing corrections are employed, and whether one considers the global mean or NH mean (Table 1). Some general conclusions are nonetheless possible.

Based on the Monte Carlo natural variability-only experiments employing the standard (ARMA) noise model (Figure 2), we conclude that the record warmth in recent years (both the specific temperature levels breached and the sequence of record temperatures) had a negligibly small likelihood of occurrence in the absence of anthropogenic warming. The specific temperature levels reached in 2014, 2015, and 2016 each had no greater than approximately one-in-a-million likelihood of chance occurrence (i.e., due to natural variability) in all of our analyses.

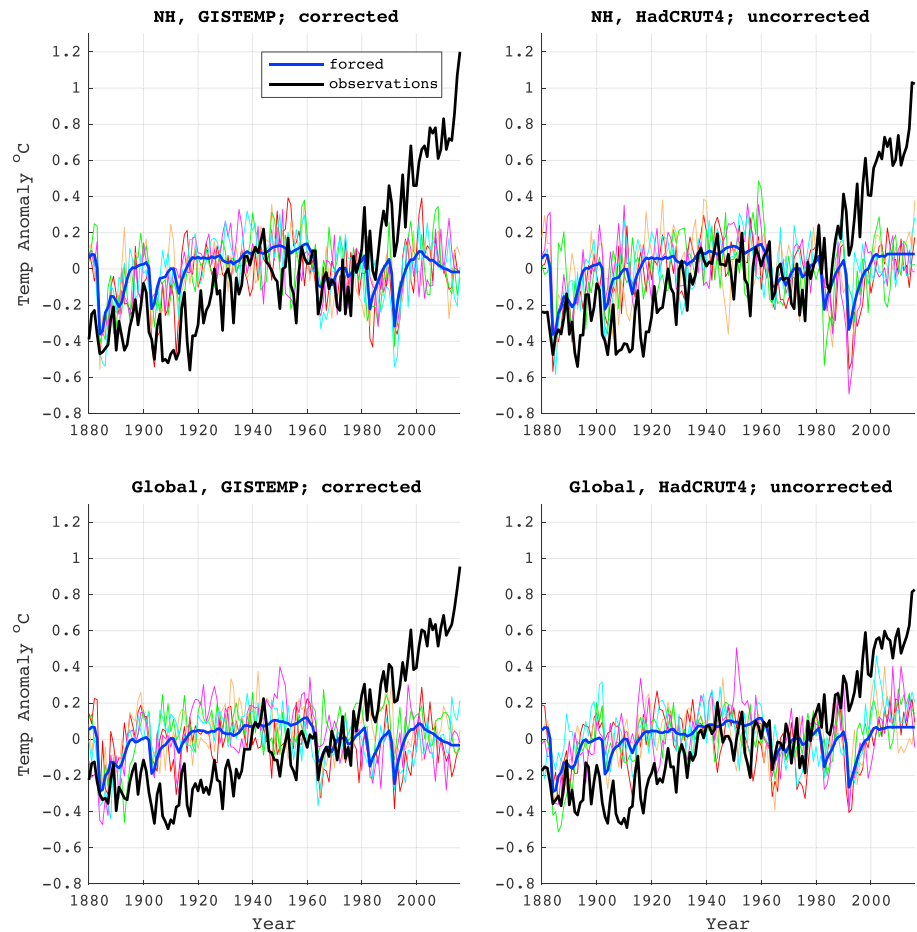


Figure 2. Simulated versus observed temperatures (1880–2016) based on natural forcing-only experiments. Shown are observations (black) and five associated Monte Carlo surrogates (colored curves) for the same cases as in Figure 1. Simulated and observed temperatures here and elsewhere are defined as anomalies relative to an 1880–2016 base period.

The likelihood of a *record* having occurred naturally in 2016 ranges from ~ 0.1 – 0.2% using corrected forcings to 1.4 – 1.6% (an order of magnitude greater) using uncorrected forcings. The lower likelihoods in the corrected forcing case are associated with a substantial dip over the past decade in the natural forced component (see Figure 2) due to negative solar and volcanic radiative forcings in the Schmidt *et al.* [2014] corrections. These likelihoods can be compared against a baseline 0.7% likelihood under the assumption of constant forcing and statistical independence of individual years.

The likelihood of the two warmest years occurring in 2015 and 2016 similarly ranges from ~ 0.01 to 0.03% using corrected forcings to 0.2 – 0.4% using uncorrected forcings, while the likelihood of the three warmest years occurring in 2014, 2015, and 2016 ranges from $<0.006\%$ using corrected forcings to 0.03 – 0.13% using uncorrected forcings. The likelihood of *three consecutive records* in 2014, 2015, and 2016 requires not only that these three years are the warmest but also that each is warmer than the former. The probability of such an occurrence is understandably even lower, varying from $<0.001\%$ (1 in 10,000) using corrected forcings to as high as 0.03% (~ 1 in 3000 likelihood) using uncorrected forcings.

If we instead use the overly conservative persistent red noise null hypothesis which exhibits low-frequency fluctuations that are arguably too large and persistent to be physically plausible (Figure 3), we unsurprisingly find substantially higher likelihoods of chance occurrence of record warmth. The likelihood of three consecutive records in 2014, 2015, and 2016 is nonetheless found to have been at most 0.14% (<1 in 700). The record *level of warmth* observed in 2016 had less than a 0.05% (~ 1 in 2000) probability of having occurred by chance.

When we account for anthropogenic warming (Figure 4) the record warmth of recent years is seen to be substantially more likely. The level of warmth observed in 2014 is found to have a likelihood ranging from 16 to

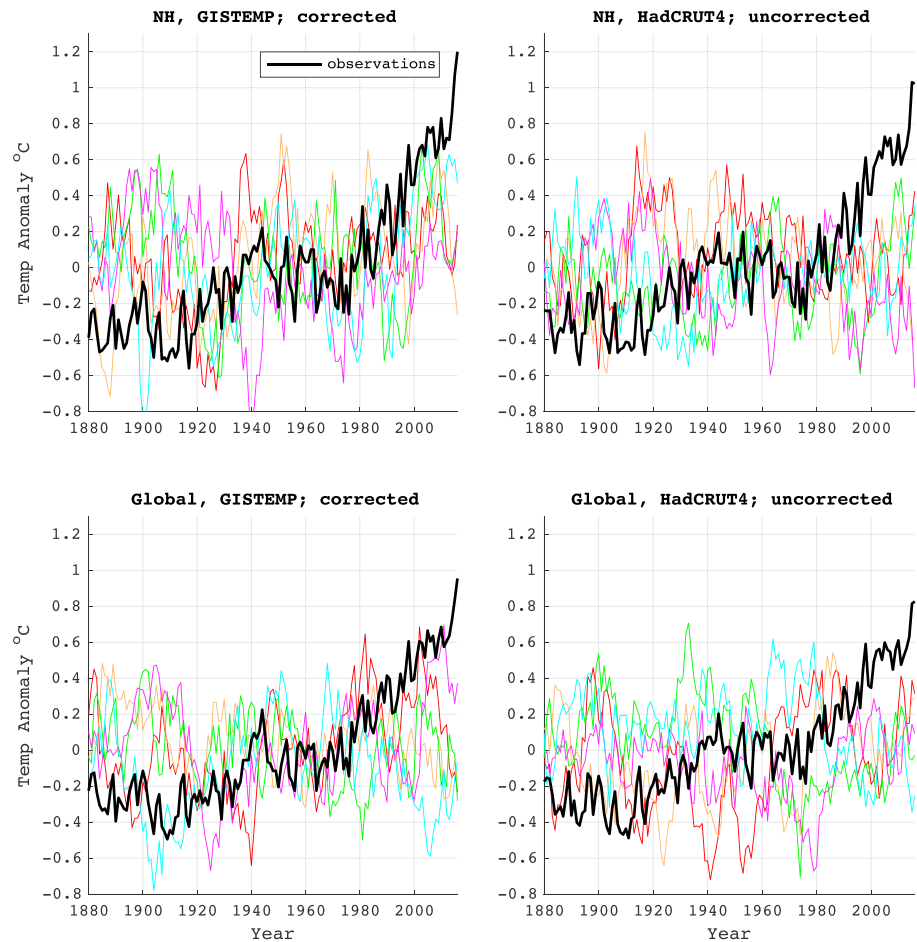


Figure 3. Simulated versus observed temperatures (1880–2016) based on “persistent red noise” experiments. Shown are observations (black) and five associated Monte Carlo surrogates (colored curves) for the same cases as in Figure 1.

81%, with the lower end of the range corresponding to the use of GISTEMP (which shows more recent warming than HadCRUT4), the use of the NH mean (which shows greater warming of the global mean), and employing the corrected CMIP5 simulations (which results in less predicted warming over the past decade). The substantially warmer 2015 ranges in likelihood from 0.7 to 26%, while 2016 ranges from 0.05 to 27%. The 2016 warmth is indeed found to be an extreme outlier if the GISTEMP NH record is used for the observations and the surrogates are based on the corrected CMIP5 simulations, but it is found to be a far more likely ~one-in-four event if the global HadCRUT4 global temperature record is used for the observations and the surrogates are based on the uncorrected CMIP5 simulations. The estimated likelihood of *a record* (but not necessarily one that achieves the observed level of warmth) in 2016 ranges from 20 to 27%, while the likelihood of the two warmest years being 2015 and 2016 ranges from 10 to 16% and the likelihood of 2014, 2015, and 2016 being the three warmest years ranges from 6 to 12%.

We estimate the likelihood of three consecutive records being observed specifically in 2014, 2015, and 2016 as ranging from 1 to 3%, suggesting that it was an unusual event but not an extraordinary one. The likelihood of three consecutive records having occurred at any specific time during the modern (post-1970) era climbs (Figure S2) to as high as 6–9% (depending on the precise series used) in the late 1990s but falls off subsequently as the warming rate in the forced series decreases modestly (see, e.g., Figure 4), dropping to the 1–3% range during the most recent years. Notable sharp decreases in the early 1980s and early 1990s are associated with the forced cooling from the El Chichon and Pinatubo eruptions, respectively.

Arguably more relevant is the question of how likely we were to observe a streak of three consecutive records at *least once* during the recent era, which amounts to a normalized sum of over time over successful

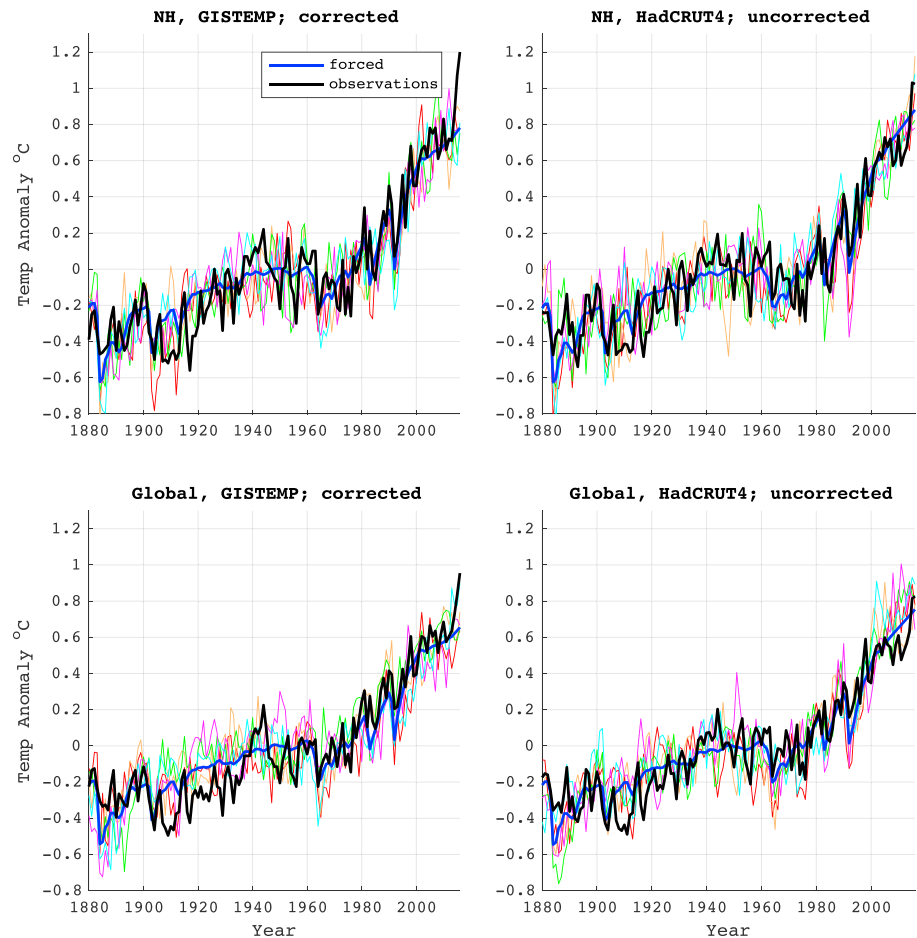


Figure 4. Simulated versus observed temperatures (1880–2016) based on all-forcing experiments. Shown are observations (black) and five associated Monte Carlo surrogates (colored curves) for the same cases as in Figure 1.

realizations (see the curves shown in Figure S2 for a depiction of the time dependence of this likelihood). We find that this ranges from 46% to 60% since 1970 and from 29 to 48% since 2000. For comparison, the probability in the absence of anthropogenic warming using ARMA noise is 0.05–0.67% and 0.04–0.46%, respectively. Using persistent red noise, it is still only 3.6–5.9% and 1.1–1.8%, respectively.

The *conditional* likelihood of a 2016 record given consecutive 2014 and 2015 (Table 1) is considerably higher than the unconditional likelihood of three consecutive records owing to the effect of persistence. When accounting for anthropogenic warming, it ranges from 25 to 35%. Even in the absence of anthropogenic warming, the persistence of neighboring years yields sizeable conditional likelihoods (5–14%). The persistent red noise model gives a conditional likelihood (22–30%) nearly as high as the anthropogenic case.

Experiments using persistence for the post-2005 extension (Figure S3 and Table S2) yields similar likelihoods for the natural-only case, but notably reduced likelihoods for the anthropogenic case. This is unsurprising since substantial (0.1–0.2°C depending on the domain and whether forcing corrections are used) additional warming takes place over the final decade, elevating the likelihood of record temperatures relative to a decade earlier.

4. Conclusions

We have employed a previously described semiempirical approach, based on a combination of historical (CMIP5) climate model simulations and observational surface temperature data, to assess the likelihood of the record warmth observed during 2014, 2015, and 2016. We find that the sequence of consecutive

record-breaking temperatures had a negligibly small ($<0.03\%$) likelihood of occurrence in the absence of anthropogenic warming based on experiments employing the most defensible (ARMA) null hypothesis for internal climate variability. Even employing a model of persistent red noise which exhibits unphysically large and persistent low-frequency internal temperature fluctuations, the likelihood of three consecutive records in 2014, 2015, and 2016 was at most $\sim 0.1\%$ (less than ~ 1 in 1000). The conclusion that this series of consecutive record-breaking temperatures was extremely unlikely to arise in the absence of anthropogenic warming (i.e., from natural variability alone) thus appears to be robust.

This series of consecutive record-breaking temperatures was still relatively unexpected, but not extraordinary (roughly 1–3% likelihood), taking anthropogenic warming into effect. There was a substantially higher, roughly 30–50% likelihood of three consecutive records occurring *sometime* since 2000, an event which was relatively unlikely ($<0.5\%$) in the absence of anthropogenic warming.

Finally, the absolute level of warmth observed during 2016 had a likelihood no greater than \sim one in a million, neglecting anthropogenic warming. However, the record had as much as a 27% likelihood, taking anthropogenic warming into account, consistent with expectations of the combined impacts on surface temperatures of anthropogenic warming and a large positive internal fluctuation associated with a historic El Niño event.

Acknowledgments

We acknowledge the World Climate Research Programme's Working Group on Coupled Modeling, which is responsible for CMIP, and we thank the climate modeling groups for producing and making available their model output. We thank Gavin A. Schmidt for providing natural and anthropogenic radiative forcing response updates. Observational surface temperature data were obtained from http://www.metoffice.gov.uk/hadobs/hadcrut4/data/current/time_series/HadCRUT.4.5.0.0.monthly_ns_avg.txt; http://www.metoffice.gov.uk/hadobs/hadcrut4/data/current/time_series/HadCRUT.4.5.0.0.monthly_nh.txt http://data.giss.nasa.gov/gistemp/tabledata_v3/GLB.Ts+dSST.txt; and http://data.giss.nasa.gov/gistemp/tabledata_v3/NH.Ts+dSST.txt.

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, doi:10.1002/2013GL057999.
- Cowtan, K., Z. Hausfather, E. Hawkins, P. Jacobs, M. E. Mann, S. K. Miller, B. A. Steinman, M. B. Stolpe, and R. G. Way (2015), Robust comparison of climate models with observations using blended land air and ocean sea surface temperatures, *Geophys. Res. Lett.*, *42*, 6526–6534, doi:10.1002/2015GL064888.
- Craigmile, P. F., P. Guttorp, R. Lund, R. L. Smith, P. W. Thorne, and D. Arndt (2014), Warm streaks in the U.S. temperature record: what are the chances?, *J. Geophys. Res. Atmos.*, *119*, 5757–5766, doi:10.1002/2013JD021446.
- Guttorp, P., and T. Y. Kim (2013), Uncertainty in ranking the hottest years of U.S. surface temperatures, *J. Clim.*, *26*, 6323–6328.
- Hansen, J., M. Sato, R. Ruedy, K. Lo, D. W. Lea, and M. Medina-Elizade (2006), Global temperature change, *Proc. Natl. Acad. Sci. U.S.A.*, *103*, 14,288–14,293, doi:10.1073/pnas.0606291103.
- Mann, M. E., S. Rahmstorf, B. A. Steinman, and S. K. Miller (2016a), The likelihood of recent record warmth, *Nat. Sci. Rep.*, *6*, 19,831.
- Mann, M. E., B. Steinman, S. K. Miller, L. Frankcombe, M. England, and A. H. Cheung (2016b), Predictability of the recent slowdown and subsequent recovery of large-scale surface warming using statistical methods, *Geophys. Res. Lett.*, *43*, 3459–3467, doi:10.1002/2016GL068159.
- 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 data set, *J. Geophys. Res.*, *117*, D08101, doi:10.1029/2011JD017187.
- Schmidt, G. A., D. T. Shindell, and K. Tsigaridis (2014), Reconciling warming trends, *Nat. Geosci.*, *7*, 158–160.
- Stocker, T. F., et al. (2013), Climate change 2013: The physical science basis—Summary for policymakers, in *Climate Change 2013: The Physical Science Basis. Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by T. F. Stocker et al., pp. 1–29, Intergovernmental Panel on Clim. Change, Switzerland.
- Zorita, E., T. F. Stocker, and H. von Storch (2008), How unusual is the recent series of warm years?, *Geophys. Res. Lett.*, *35*, L24706, doi:10.1029/2008GL036228.