

## Statistics of Regional Surface Temperatures after 1900: Long-Range versus Short-Range Dependence and Significance of Warming Trends

OLA LØVSLETTEN AND MARTIN RYPDAL

*Department of Mathematics and Statistics, UiT Arctic University  
of Norway, Tromsø, Norway*

(Manuscript received 25 June 2015, in final form 19 November 2015)

### ABSTRACT

This paper studies regional climate variability for the time period 1900–2013 using parsimonious stochastic models. Instrumental data records on  $5^\circ \times 5^\circ$ ,  $2^\circ \times 2^\circ$ , and equal-area grids are examined. A long-range dependent (LRD) stochastic process is used as a simplified description of the multitude of response times in the climate system. Fitting a linear trend to the global mean surface temperature (GMST) implies a warming of  $0.08 \text{ decade}^{-1}$ , which is highly significant under an LRD null hypothesis ( $p < 10^{-4}$ ). The regional trends are distributed around the global mean trend, while the fluctuation levels increases when going from global to regional scale. The temperature fluctuations of the tropical oceans are observed to be strongly influenced by El Niño–Southern Oscillation (ENSO) and, therefore, more consistent with autoregressive processes of order 1 [AR(1)]. A likelihood-ratio test is used to systematically determine the best null model [AR(1) or LRD]. About 80% of the regional warming trends are found to be significant (with a 5% significance level).

### 1. Introduction

Given the extensive evidence of global warming, there is now an increased attention to whether trends can be detected on local and/or regional scales and also to the spatiotemporal pattern of climate variability. [Stott et al. \(2010\)](#) and [Knutson et al. \(2013\)](#) have presented such an analysis using control runs of climate models as a null hypothesis for trend detection. An alternative and complementary approach, which we pursue in this paper, is to use stochastic models. The main objective is to test the hypothesis of a linear trend versus the null hypothesis of “stationary climate.” That is, we assume that temperature time series  $Y(t)$  can be modeled as superpositions of

deterministic trend signals  $m(t)$  and stationary, stochastic processes (climate noise)  $X(t)$ :

$$Y(t) = m(t) + X(t), \quad (1)$$

with  $m(t) = a_0 + a_1 t$ . The choice of a linear trend is mainly used to test the hypothesis that stationary climate can explain the last 110 years of warming, without assuming the correctness of this model ([Bloomfield 1992](#)).

For the regional surface temperature series analyzed in this paper we find, except for a small part of the land area, significant positive serial correlation (after detrending) of the residuals, with higher persistence over oceans compared to land. Thus, the stochastic part of the model  $X(t)$  should have built-in memory, consistent with the serial correlations of the observations. For the global mean surface temperature (GMST) there is evidence of long-range dependence (LRD) ([Bloomfield 1992](#); [Rypdal et al. 2013](#)). Similar statistics are found in some grid cells, and it is therefore reasonable to choose stochastic models that exhibit scaling and slowly decaying autocorrelation functions (ACFs).

For the GMST, [Cohn and Lins \(2005\)](#) and [Koutsoyiannis and Montanari \(2007\)](#) have raised doubt about the statistical significance of a warming

Denotes Open Access content.

Supplemental information related to this paper is available at the Journals Online website: <http://dx.doi.org/10.1175/JCLI-D-15-0437.s1>.

*Corresponding author address:* Ola Løvsletten, Department of Mathematics and Statistics, UiT Arctic University of Norway, N-9037 Tromsø, Norway.  
E-mail: [ola.lovsletten@uit.no](mailto:ola.lovsletten@uit.no)

DOI: 10.1175/JCLI-D-15-0437.1

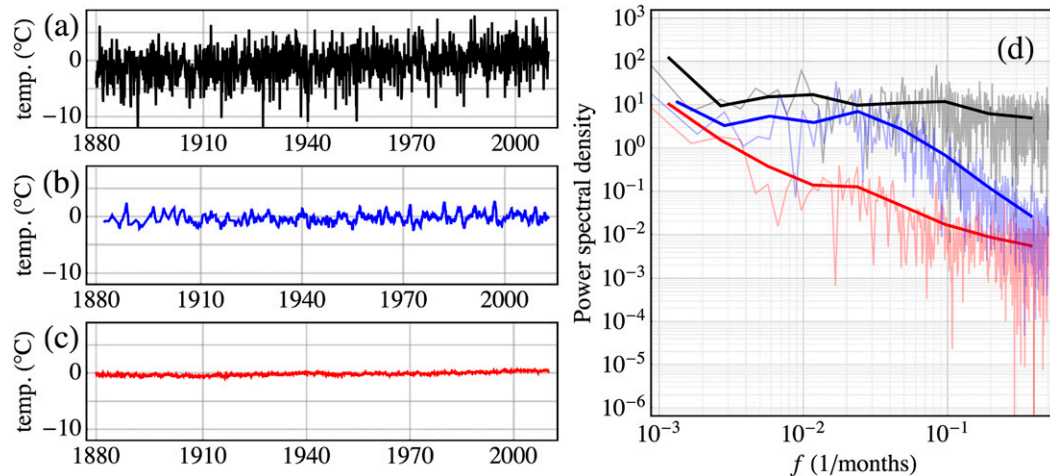


FIG. 1. (a) The black curve is the monthly temperature data for Moscow, Russia. (b) The blue curve is the monthly reconstructed temperature for the  $5^{\circ} \times 5^{\circ}$  grid centered at  $2.5^{\circ}\text{S}$ ,  $142.5^{\circ}\text{W}$ . (c) The red curve is the global mean temperature anomaly plotted with monthly resolution. (d) The PSD of the three time series in (a)–(c). The smooth curves are obtained by averaging over logarithmic bins. The colors of the PSDs are the same as used for the signals in (a)–(c).

trend under an LRD null hypothesis, while Bunde and Lennartz (2012) find that a linear trend is significant at the 5% but not the 1% significance level. We have conducted our own analysis (this is presented in section 4) using standard statistical methods, which shows that a linear trend for the GMST is highly significant ( $p < 10^{-4}$ ). We note that a second-order polynomial trend (with linear term set to zero) is a better model in terms of the explained variation  $R^2$  reflecting that global warming has been accelerating.

On regional scales, the question of statistical significance of trends is not as clear-cut because of the much lower signal-to-noise ratio. This is illustrated in Fig. 1, where we have plotted monthly deseasonalized temperature data for the city of Moscow, Russia, together with the global mean temperature anomaly. While the trend estimates (slopes) are distributed around the GMST trend estimate, the fluctuation level is much higher. However, for many grid cells the persistence parameter (e.g., Hurst exponent in the LRD model) is lower than for the GMST. Thus the result of a detection analysis is not given a priori. A complicating factor is that regions strongly influenced by El Niño–Southern Oscillation (ENSO) have stronger persistence on time scales of 2–5 yr than predicted by an LRD process (Huybers and Curry 2006) and lower persistence than is predicted from an LRD model on time scales longer than a decade. In fact, the estimated power spectral densities (PSDs) of the temperature fluctuations in regions strongly influenced by ENSO are inconsistent with a power law, but fit better with the Lorentzian-shaped PSDs that characterize an

autoregressive process of order 1, the so-called AR(1) model.<sup>1</sup>

We note that in some aspects it is unsatisfactory to use AR(1) models to describe ENSO dynamics, since we know that ENSO is an oscillatory mode in the climate system. The AR(1) models, which can be seen as discretizations of the Ornstein–Uhlenbeck processes, take shape from simple linear first-order equations with dissipation and random forcing, and hence they are incapable of describing oscillating modes. On the other hand, we are not seeking an accurate physical model of ENSO; rather, we need to quantify how the fluctuation levels in the climate noise vary with time scales. More specifically, we need to make an estimate of the natural climate variability on centennial time scales based on the statistical properties of the climate variability on the shorter time scales. The role of the models in trend detection is therefore to correctly prescribe the fluctuation levels on the long time scales using parameters estimated from the statistics on the shorter time scales. If we apply an LRD model in the ENSO regions, we will estimate very large Hurst exponents, which in turn will overestimate the natural variability on the centennial time scales.

For many grid cells it is not clear whether to choose an AR(1) or LRD process. This is an inherent statistical problem given the available sample length of about 110 years of data (Percival et al. 2001). Vyushin et al. (2012) find that climate variability appears to be more

<sup>1</sup> AR(1) models are commonly used to model climate noise [e.g., Fig. SPM.1(b) in IPCC (2013)].

persistent than an AR(1) process and less persistent than a power-law process, and conclude that both representations are potentially useful for statistical applications. Thus, in a first attempt we compute the statistical significance against both null models. A similar approach is taken by [Franzke \(2012\)](#), who classifies the degree of significance based on the fraction of a set of null models that are rejected by the observations. We advance this approach further by selecting the “best” null model based on likelihood-ratio (LR) criteria. The LR test classifies the ENSO regions as significantly (significance level 5%) better described by an AR(1) than the LRD model fractional Gaussian noise (fGn). This is consistent with our empirical analysis and also with the findings of [Huybers and Curry \(2006\)](#). We also observe examples of the opposite [fGn better than AR(1)], while many grid cells are classified as undecided in the sense that the test is unable to discriminate between the two models. By assessing trend significance against the best null model we find that about 80% of the grid points have significant warming trends at the 5% significance level.

The remainder of this paper is organized as follows: In [section 2](#) we review the stochastic models used in this study. An outline of the statistical methods used is given in [section 3](#). In particular, we review the trend detection methodology used in this paper. The main results are presented in [section 4](#). We discuss our findings in [section 5](#).

## 2. Stochastic models

### a. Hurst exponent

As noticed by [Hurst \(1957\)](#), many signals in nature satisfy scaling in the sense that the fluctuations levels of their coarse-grained versions vary as power-law functions of the aggregation scale. For a time series  $X_t$ , this means that the standard deviation of the running mean  $Y_k = t^{-1}(X_{k-t+1} + \dots + X_k)$  scales as  $\propto t^{H-1}$ , so if the signal is stationary we can define the Hurst exponent  $H \in (0, 1)$  by the relation

$$\mathbb{E} \left( \sum_{k=1}^t X_k \right)^2 \propto t^{2H}. \tag{2}$$

From stationarity and Eq. (2) it follows that the autocovariance function  $\gamma(\tau)$  of  $X_t$  decays asymptotically as a power law:

$$\begin{aligned} \gamma(\tau) &= \frac{\sigma^2}{2} (|\tau + 1|^{2H} - 2|\tau|^{2H} + |\tau - 1|^{2H}) \\ &\sim \sigma^{2H(2H-1)} \tau^{2H-2}. \end{aligned} \tag{3}$$

The parameter  $\sigma > 0$  is the standard deviation, while the Hurst exponent  $H \in (0, 1)$  determines the correlation

structure. For  $H = 1/2$ , the stochastic process  $X_t$  is white noise, while  $H > 1/2$  gives persistent (positive correlated) random variables. The case  $H < 1/2$  corresponds to negative correlation and is not relevant here [see [Rypdal and Løvsetten \(2013\)](#) for application of antipersistent stochastic processes with power-law statistics].

One can extend the definition of the Hurst exponent to also include certain nonstationary processes. For instance, if  $X(t)$  is nonstationary with a power-law variogram but has stationary increments, then one can define the Hurst exponent by

$$\mathbb{E}X(t)^2 \propto t^{2(H-1)}. \tag{4}$$

With this (extended) definition a Brownian motion has Hurst exponent  $H = 3/2$  while Gaussian white noise has  $H = 1/2$ .

Two classes of stochastic processes with well-defined Hurst exponents are the self-similar ([Embrechts and Maejima 2002](#)) and the multifractal processes (e.g., [Løvsetten and Rypdal 2012](#)) with finite second moments. The Ornstein–Uhlenbeck process, defined as the solution to the stochastic differential equation (SDE)

$$dX(t) = -\frac{1}{\tau} X(t)dt + dB(t), \tag{5}$$

where  $B(t)$  is a Brownian motion and  $\tau > 0$ , does not satisfy the scaling relation Eq. (2). However, an Ornstein–Uhlenbeck process scales asymptotically. When  $\tau \rightarrow \infty$ ,  $X(t)$  converges to a Brownian motion and as  $\tau \rightarrow 0$  the process  $X(t)$  is a Gaussian white noise.

### b. Fractional Gaussian noise

The LRD model adopted in this paper is the fGn. If we assume that  $X_t$  is a Gaussian and stationary stochastic processes that satisfies the scaling property of Eq. (2), then these properties define the class of fGn. In discrete time fGn can be defined as the increments of a continuous time fractional Brownian motion (fBm) ([Mandelbrot and Van Ness 1968](#)):

$$\begin{aligned} B_H(t) &= \int_0^t (t-s)^{H-1/2} dB_s \\ &\quad - \int_{-\infty}^t [(t-s)^{H-1/2} - (-s)^{H-1/2}] dB_s. \end{aligned}$$

In continuous time fGn is not well defined as a (finite variance) process, but rather as a random signed measure. However, using the definition of fBm one can write a formal (but divergent) integral representation of fGn:

$$X_t = \int_{-\infty}^t (t-s)^{H-3/2} dB_s. \quad (6)$$

### c. Ornstein–Uhlenbeck and AR(1) processes

An Ornstein–Uhlenbeck (OU) process is defined by replacing the power law  $(t-s)^{H-3/2}$  in Eq. (6) with an exponential kernel  $\propto e^{-(t-s)/\tau}$ . This introduces a characteristic time scale  $\tau > 0$ , and the formulation is equivalent to the SDE in Eq. (5). Straightforward discretization of this equation gives an AR(1) process:

$$X_{t+1} = \phi X_t + \varepsilon_t, \quad (7)$$

where  $\phi = 1 - \Delta t/\tau$ , and  $\varepsilon_t$  are independent and identically distributed Gaussian random variables. The power spectral density of an OU process is Lorentzian, with  $S(f) \sim f^{-2}$  for  $f \gg 1/\tau$  and  $S(f) \sim f^0$  for  $f \ll 1/\tau$ . Hence we have two scaling regimes, one corresponding to Brownian motion (i.e.,  $H = 3/2$ ) on short time scales, and one corresponding to white noise (i.e.,  $H = 1/2$ ) on long time scales. The transition between these time scales is given by the characteristic time  $\tau$ , which is also the  $e$ -folding time for the ACF.

### 3. Statistical methods

In this section we present theory for trend significance testing for linear models where the noise is an LRD process. Many of these results can be found in [Ko et al. \(2008\)](#) and the references therein, but we will also present some extensions and modifications of the existing theory. We note that the statistical methods used in this paper have been tested and validated in the supplementary material.

Consider  $n$  observations from the linear trend model in Eq. (1) where the climate variability  $X_t$  is represented by an fGn with scale parameter  $\text{var}(X_1) = \sigma^2$  and Hurst exponent  $H$ . From the definition of an fGn it follows that the random vector  $\mathbf{X} = (X_1, \dots, X_n)^T$  is multivariate normal distributed

$$\mathbf{X} \sim \mathcal{N}(0, \mathbf{\Gamma}),$$

where the  $n \times n$  covariance matrix  $\mathbf{\Gamma}$  is the Toeplitz matrix of the autocovariances  $[\gamma(0), \dots, \gamma(n-1)]$ ; that is, elements  $(i, j)$  of  $\mathbf{\Gamma}$  are in the form  $\gamma(|i-j|)$ , with  $\gamma(\cdot)$  defined in Eq. (3). Denote by  $\mathbf{R}_H$  the correlation matrix of  $\mathbf{X}$  and note that  $\mathbf{\Gamma} = \sigma^2 \mathbf{R}_H$ . It is convenient to write the linear trend model in vector form:

$$\mathbf{Y} = \mathbf{B}^T \mathbf{a} + \mathbf{X}, \quad (8)$$

where  $\mathbf{a} = (a_0, a_1)^T$  and  $\mathbf{B}$  is the  $2 \times n$  design matrix with ones on the first row and the sampling times  $(1, 2, \dots, n)$

as the second row. The ordinary least squares (OLS) estimator of  $\mathbf{a}$  can then be written as

$$\hat{\mathbf{A}} \stackrel{\text{def}}{=} (\hat{A}_0, \hat{A}_1)^T = (\mathbf{B}\mathbf{B}^T)^{-1} \mathbf{B}\mathbf{Y}. \quad (9)$$

This estimator has a bivariate normal distribution with mean  $\mathbf{a}$  and covariance matrix

$$\text{Cov}(\hat{\mathbf{A}}) = \mathbf{C}(H)\sigma^2, \quad (10)$$

with

$$\mathbf{C}(H) = (\mathbf{B}\mathbf{B}^T)^{-1} \mathbf{B}\mathbf{R}_H \mathbf{B}^T (\mathbf{B}\mathbf{B}^T)^{-1}.$$

If we define  $c(H)$  to be element  $(2, 2)$  of the correlation matrix  $\mathbf{C}(H)$ , then the estimator for the slope is distributed as  $\hat{A}_1 \sim \mathcal{N}[a_1, \sigma^2 c(H)]$ ; that is,

$$T_{a_1}(H, \sigma, \hat{A}_1) \stackrel{\text{def}}{=} \frac{\hat{A}_1 - a_1}{\sigma c(H)^{1/2}} \sim \mathcal{N}(0, 1), \quad (11)$$

where  $c(H)^{1/2} \sim n^{H-2}$ . A closed-form expression for the variance factor  $c(H)$  can be found in [Lee and Lund \(2004\)](#). By setting  $a_1 = 0$ , Eq. (11) gives the distribution of trend estimates under the null hypothesis of no trend. It follows that a  $(1 - \alpha) \times 100\%$  confidence interval is given by

$$\hat{a}_1 \pm \text{sd}(\hat{A}_1 | \sigma, H) z_{\alpha/2}, \quad (12)$$

with  $\hat{a}_1$  the (OLS) estimated slope and  $z_\alpha$  the  $\alpha$  upper quantile of the standard normal distribution. The corresponding  $p$  value (probability of an fGn producing a larger trend estimate than the observed estimate) is given by

$$p_{\text{asym}} = 2 \left\{ 1 - \Phi \left[ \left| \frac{\hat{a}_1}{\text{sd}(\hat{A}_1 | \sigma, H)} \right| \right] \right\}, \quad (13)$$

where  $\Phi$  is the cumulative distribution function of a standard normal random variable. Equations (12) and (13) come with the tacit assumption of known noise parameters. For most practical applications, one only has access to a set of parameter estimates. To assess trend significance, in a first attempt, one can just plug in the estimates of the noise parameters. For consistent estimators this approach results in an asymptotically (i.e., as the sample size goes to infinity) valid significance test. The advantage of this approach is that analytical formulas are available.

To estimate the Hurst exponent we use the maximum likelihood (ML) method (e.g., [McLeod et al. 2007](#)). As noted by [Koutsoyiannis and Montanari \(2007\)](#), the usual white-noise estimator for the scale parameter  $\sigma$  is

severely biased for LRD processes. A better alternative is to use the ML estimator, adjusted such that the sample length  $n$  in the denominator is replaced with  $n - 2$ :

$$\hat{\sigma}^2 = \frac{1}{n-2} \mathbf{x}^T \mathbf{R}_{\hat{h}}^{-1} \mathbf{x}, \quad (14)$$

where  $\hat{h}$  is the maximum likelihood estimate (MLE) of the Hurst exponent, and  $\mathbf{R}_{\hat{h}}$  the corresponding Toeplitz matrix formed from the ACF of  $n$  observations. In Eq. (14) the matrix  $\mathbf{R}_{\hat{h}}$  has the effect of decorrelating an fGn vector  $\mathbf{X}$ , in the sense that  $\mathbf{L}_H^{-1} \mathbf{X}$ , with  $\mathbf{R}_H = \mathbf{L}_H \mathbf{L}_H^T$ , is independent, standard normal variables.

The noise parameters are estimated from the residuals  $\mathbf{x}$ , found by subtracting the OLS linear trend. Several authors (e.g., Koutsoyiannis and Montanari 2007; Franzke 2012) have argued that, to reflect the null hypothesis, these estimates should be calculated directly from the data. This gives a very weak significance test, since only the null hypothesis, and not the null and alternative hypothesis, is taken into account. Indeed, if we have a trend, this approach will lead to an erroneous high estimate of the scale parameter and also the Hurst exponent. If we instead subtract an estimated trend, given the null hypothesis, we introduce a small bias in the estimates. A similar bias is also introduced by just subtracting the sample mean (see Table S2 in the supplementary material). However, this inherent bias can be accounted for by adopting the small-sample correction proposed by Ko et al. (2008), and the details of this procedure can be found in the supplementary material.

While uncertainties in the estimates of the Hurst exponent and the scale parameter are taken into account with this small-sample correction method, the significance test still depends crucially on the estimated Hurst exponent. To add robustness to our results, we consider ML estimates on several time scales, and also detrended fluctuation analysis of order 2 and simple variograms. The advantage of these methods is that one can visually inspect the scaling properties (taking into account the well-known error bars). In addition we have inspected the ACFs for detrended data. From these nonparametric methods we identify a lack of scaling for the temperature fluctuations in some grid cells, most notably in the ENSO region.

Trend detection under an AR(1) model follows along the same lines with an explicit description given by Lee and Lund (2008).

#### 4. Analysis of surface temperature data

##### a. Data

Four datasets are analyzed in this project. These are the HadCRUT4 surface temperature anomalies

(Morice et al. 2012), which combine the land temperatures from the CRU surface temperature data version 4 (CRUTEM4; Jones et al. 2012) and the sea surface temperatures (SSTs) from the Hadley Centre SST data version 3 (HadSST3; Kennedy et al. 2011). We also use the NOAA Merged Land–Ocean Surface Temperature Analysis (MLOST, V3.5.4) data developed by Smith and Reynolds (2005). In both of these datasets the mean temperature in  $5^\circ \times 5^\circ$  grids are provided with monthly time resolution. In addition to these we use Berkeley Earth’s 15984 equal-area dataset, and the GISS Surface Temperature Analysis (hereafter GISS; Hansen et al. 2010), with 1200-km smoothing, which is given on  $2^\circ \times 2^\circ$  grids. Possible sources of differences between the GISS, HadCRUT4, and NOAA MLOST data products have been briefly discussed by Libardoni and Forest (2011). The majority of land surface data [which comes from the Global Historical Climatology Network (GHCN)] are treated differently in construction of the different datasets. For instance, in the construction of the HadCRUT4 data there is a requirement that stations should have a certain number of observations in their normal period 1960–90, while in the construction of the GISS data (with 1200-km smoothing) a station is only included if there are other stations within a 1200-km radius with a period of overlap that is at least 20 years. In addition, each data product uses different SSTs, and there are differences in the way that data are extrapolated, or not extrapolated. The Berkeley land temperatures are constructed from 16 preexisting data archives. The current archive uses over 39 000 unique stations which is roughly 5 times the number of stations used in GHCN. The Berkeley SST is a modified version of the HadSST3.

All four datasets were downloaded on 1 October 2015 from the web pages listed in the supplementary material. The time period analyzed is January 1900–December 2013.

##### b. Sampling scale

For the regional surface temperature series we observe that direct application of the ML method tends to give higher estimates of the Hurst exponent compared with the detrended fluctuation analysis of order 2 (DFA2). For the latter we have control over which time scales contribute to the estimate. We also observe that the discrepancy between the two methods disappears if the signals are coarse grained over 4-month windows prior to the ML estimation (i.e., if a new, coarser time series is produced by dividing the series into 4-month segments and averaging the data points within each segment). Which time scales that should be emphasized in the parameter estimation is always a trade-off between the improved statistics achieved when focusing on



TABLE 1. Linear trend model with fGn errors. The first column labeled “trend” is the OLS estimate of the slope, with standard deviation in parentheses. The bias-corrected ML estimate of the Hurst exponent is  $H_{BC}$ ;  $\sigma$  and  $\sigma_{wn}$  are estimates, adjusted by ML [Eq. (14) with  $H_{BC}$ ] and OLS, respectively, of the standard deviation around the slope. The  $p$  value of the trend and standard deviation of  $\hat{a}_1$  are computed from the small-sample correction explained in section 3.

Trend (K decade <sup>-1</sup> )	$p$ value trend	$H_{BC}$	$\sigma_{BC}$	$\sigma_{wn}$
0.083 (0.015)	GISS <10 <sup>-6</sup>	0.99	0.44	0.14
0.079 (0.016)	Berkeley Earth <10 <sup>-5</sup>	0.99	0.45	0.15
0.075 (0.016)	HadCRUT4 <10 <sup>-4</sup>	0.99	0.45	0.15
0.081 (0.014)	NOAA MLOST <10 <sup>-6</sup>	0.99	0.40	0.13

the shorter scales and the increased relevance and importance of the longer scales. The choice to apply a 4-month coarse graining is based on this type of consideration, and it is meant to ensure that distinctive features of the month-to-month fluctuations do not have too large an impact on the predicted centennial-scale fluctuation level.

### c. GMST trend significance

In Table 1 we present the results of a trend detection analysis for the four GMST time series. We see that there is very little variation between the four data products, with linear trends  $\approx 0.08$  K decade<sup>-1</sup> and

fluctuation levels  $\sigma_{wn} \approx 0.15$  K (4 months)<sup>-1</sup>. Here  $\sigma_{wn}$  denotes the white-noise estimator, which is defined in Eq. (14), with  $\mathbf{R}_h$  replaced by the identity matrix. The MLEs of the Hurst exponents are  $H = 0.97$  for the GISS data and  $H = 0.98$  for the other three GMST time series (not shown in the table). Since the methods we apply are restricted to the case  $H < 1$ , we should be attentive to the fact that the high estimates for  $H$  could simply be a result of the upper bound  $H = 1$ . This would be the case if the GMST scales with an exponent  $H > 1$ . However, this can be tested using the DFA2 estimator, which can be used both in the cases  $H < 1$  and  $H > 1$ . The results of the DFA2 estimator to the GMST data are in the range from  $H = 0.87$  to  $H = 0.96$  for all the four data products. The bias-corrected ML estimates are  $H_{BC} = 0.99$ , and the resulting adjusted ML estimator for the fluctuation level [see Eq. (14)] is  $\sigma \approx 0.45$  K (4 months)<sup>-1</sup>. The rather large discrepancy between the estimates for the fluctuation level is caused by Hurst exponents close to one.

The statistical significance of the trend estimates are computed using  $H_{BC}$  and  $\sigma$  with the small-sample correction outlined in section 3 (details of this method are given in the supplementary material). The  $p$  values for the OLS slopes are less than 10<sup>-4</sup> and thus highly significant. The 95% confidence intervals for the trends are  $\approx 0.08 \pm 0.03$  K decade<sup>-1</sup>.

### d. Regional results

We start the discussion of regional statistics by first considering the GISS dataset. Figure 2a shows the

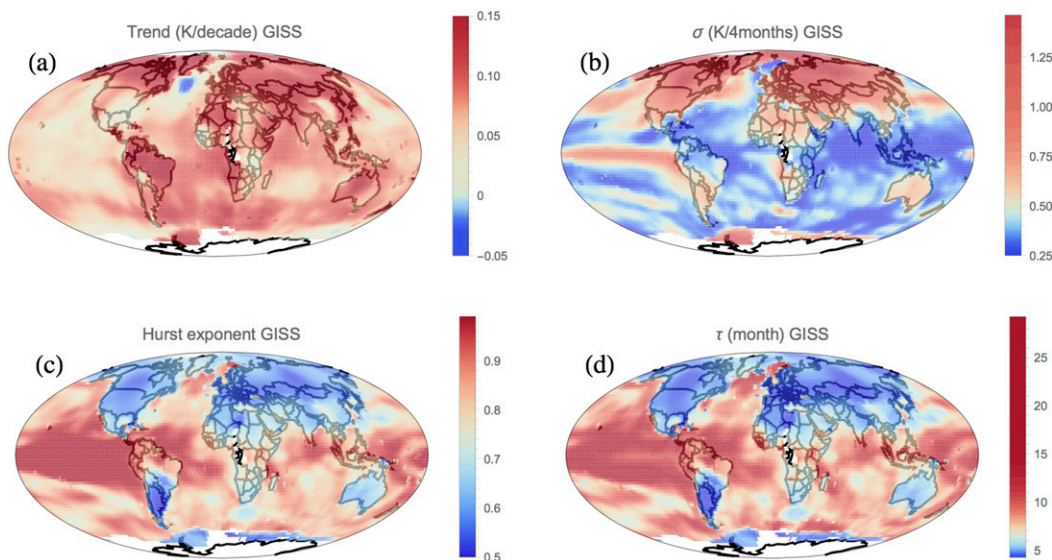


FIG. 2. (a) Linear trend for the period 1900–2013 in each  $2^\circ \times 2^\circ$  grid of the GISS dataset. (b) Standard deviation  $\sigma$  around the regression line. (c) Hurst exponents in the fGn model. (d) Correlation time  $\tau$  in the AR(1) process. All estimates are performed subsequent to a 4-month coarse graining.

TABLE 2. Summary of regional trends and standard deviations, with GMST values in the last column.

	Min	1st quartile	Median	Mean	3rd quartile	Max	Global
Trends ( $\text{K decade}^{-1}$ )							
GISS	-0.04	0.06	0.08	0.09	0.11	0.28	0.08
Berkeley Earth	-0.25	0.05	0.08	0.08	0.10	0.52	0.08
HadCRUT4	-0.13	0.05	0.07	0.08	0.10	0.36	0.08
NOAA MLOST	-0.04	0.06	0.08	0.08	0.10	0.24	0.08
Standard deviation [ $\text{K (4 months)}^{-1}$ ]							
GISS	0.16	0.35	0.47	0.69	0.93	2.04	0.14
Berkeley Earth	0.20	0.41	0.50	0.63	0.69	2.69	0.15
HadCRUT4	0.25	0.50	0.64	0.77	0.87	2.44	0.15
NOAA MLOST	0.16	0.33	0.41	0.56	0.65	2.19	0.13

estimated trends, and as can be seen in Table 2, the regional trends are distributed around the GMST trend. We observe warming over all of Earth's surface, except for a small region in the North Atlantic. The warming trends are generally weaker in the SST compared to surface air temperature (SAT) over land; in particular, we observe weaker trends in the Pacific Ocean.

Figure 2b shows the (white noise) fluctuation levels of the temperature signal (i.e., standard deviation around the regression line). A summary of these estimates can be found in Table 2. The MLEs of the fluctuation levels based on an AR(1) model and a fGn model yield similar results. Very large fluctuation levels are observed over land compared to the oceans, and hence it is not a priori clear that the stronger trend over land is more significant than the weaker trend in the oceans. There are also large fluctuation levels around the equator in the Pacific

Ocean. This is a region that is colder than average during the La Niña cold phase and warmer than average in the El Niño warm phase. In this region, the standard deviations are influenced by the ENSO, and not only the year-to-year variability. As discussed in the introduction, this is one of the reasons why an AR(1) process is a better null model in this region.

The estimated Hurst exponents are shown in Fig. 2c, and we observe stronger persistence in SST than in land temperatures. In North America and in Eurasia the estimated model is close to a white-noise process (i.e.,  $H \approx 0.5$ ), while we apparently have strong LRD in the oceans, in particular in the tropical Pacific. A similar picture is seen in Fig. 2d. Here we have plotted the estimated correlation length in an AR(1) process. We observe that the estimated correlation time varies from a few months over much of Earth's land areas to a

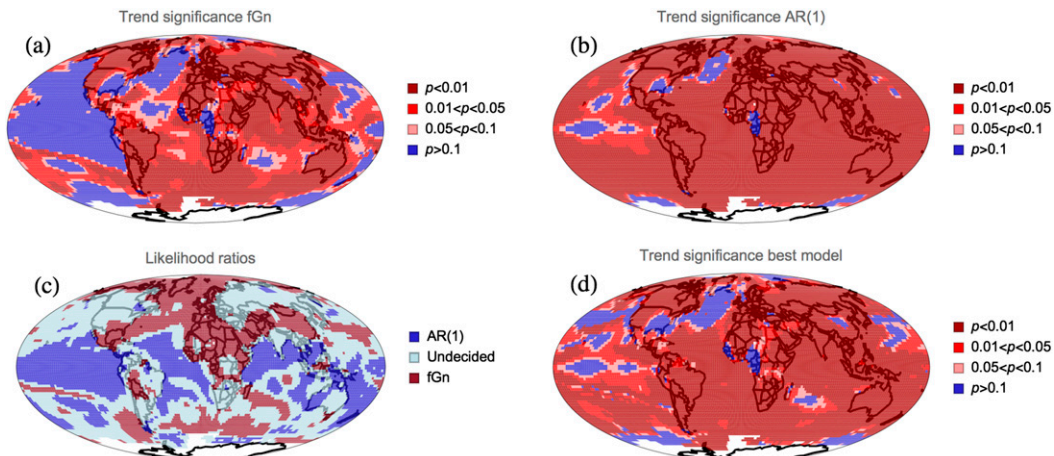


FIG. 3. GISS dataset: (a) The distribution of  $p$  values based on an fGn null model. (b) The distribution of  $p$  values based on an AR(1) null model. (c) The results of the likelihood ratio model selection test. In the grid points marked as red the data are more consistent with a fGn error model, and in the grid points marked as blue the data are more consistent with an AR(1) error model. In the grid points marked as light blue, one model is not significantly preferred over the other. (d) The distribution of  $p$  values when the model with the highest likelihood is chosen as the null model in each grid point. The  $p$  values are adjusted for multiple testing using the false discovery rate (FDR) method.

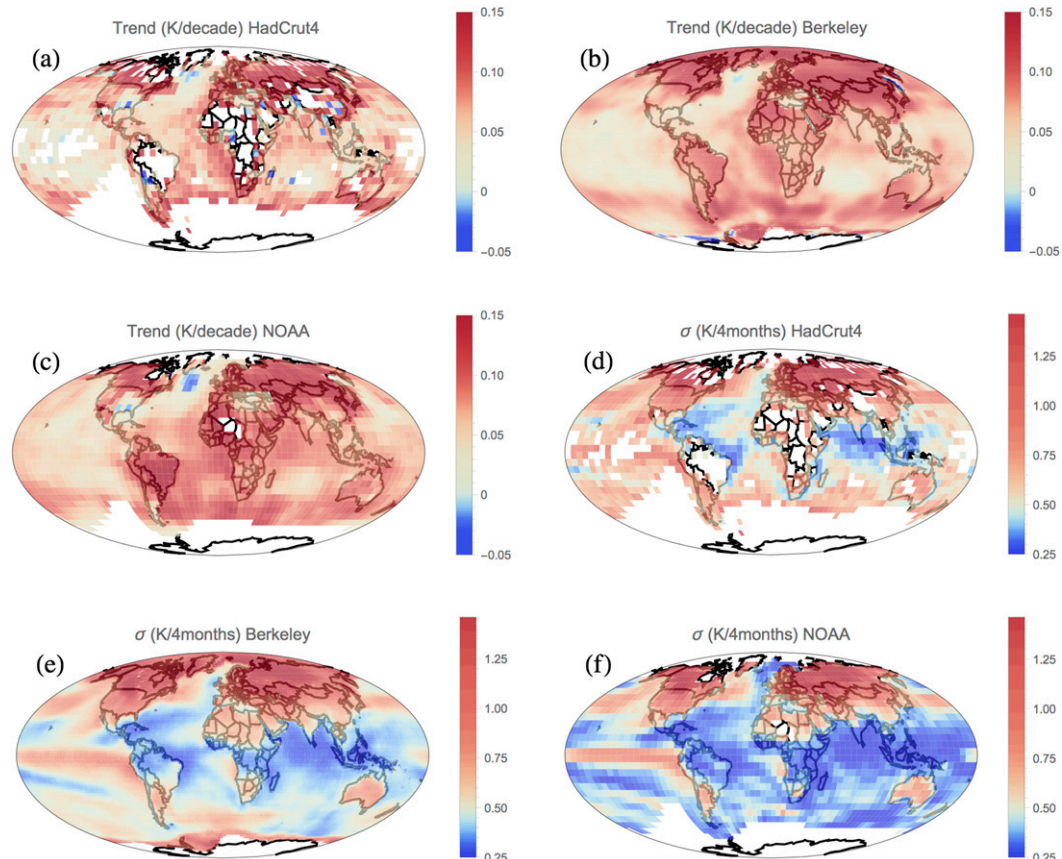


FIG. 4. (a)–(c) Linear trend for the period 1900–2013. (d)–(f) Standard deviation  $\sigma$  around the regression line. All estimates are performed subsequent to a 4-month coarse graining. Data product shown in the titles.

couple of years in the tropical Pacific and tropical Atlantic.

Based on the parameter estimates presented in Fig. 2 we can compute the  $p$  values for the estimated trends. As illustrated in Figs. 3a and 3b, these  $p$  values depend crucially on the chosen null model. In Fig. 3a we have shown a map of the  $p$  values computed with respect to the fGn model, and in Fig. 3b we have shown the corresponding  $p$  values computed with respect to the AR(1) model. A striking feature in these plots is that the SST trends for cell points in the Pacific Ocean are determined as significant with respect to an AR(1) model, but cannot be concluded as significant if we apply an LRD model. Hence, our interpretation of the significance of the local warming trends in the Pacific Ocean depends on which model is best suited to describe the correlation structure in these data.

As discussed in the introduction, we observe that many of the time series in this region (see, e.g., Figs. 1b,d) have statistical properties that are strongly influenced by ENSO. That is, the PSDs are not power laws, but rather have strong persistence on the shortest time

scales and white-noise characteristics on longer scales. In contrast, many of the SST series in the North Atlantic basin, the statistical properties of which are influenced by the Atlantic multidecadal oscillation (AMO), are consistent with a scaling model. It is important to realize that a persistent ( $H > 0.5$ ) scaling description of the climate noise is a parsimonious way of stating that there are natural oscillations on all scales, and the parameter  $H$  determines the relative fluctuation levels of the slow oscillations compared to the faster modes. However, as the PSD reveals, the ENSO is too strong to be consistent with an LRD model and must be seen as an anomalous oscillation in this description. Whether or not the AMO is anomalous with respect to an LRD description is difficult to determine from the instrumental record due to insufficient statistics. In any case, it is evident that the persistent multidecadal SST variability in the North Atlantic and SAT variability over adjacent continents is related to the AMO and the North Atlantic Oscillation (NAO) (Li et al. 2013).

To systematically determine if an AR(1) null model or LRD null model is best suited at a given geographic



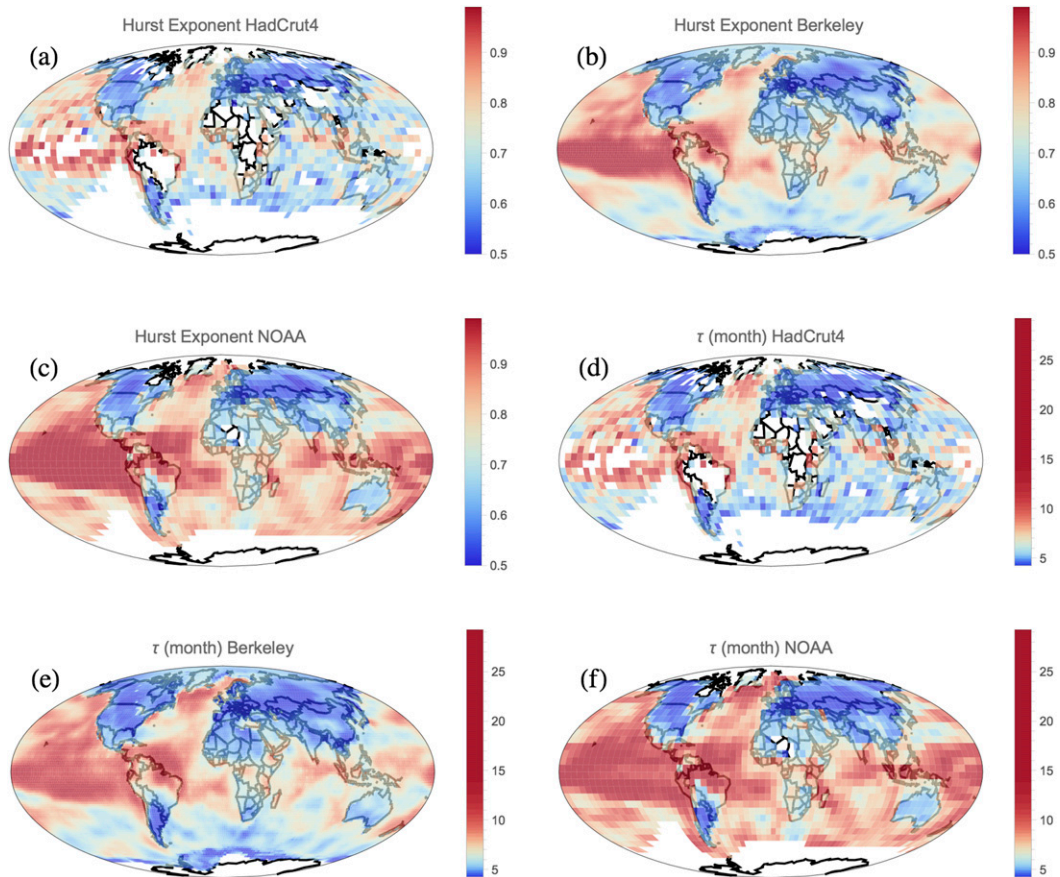


FIG. 5. (a)–(c) Hurst exponents in the fGn model. (d)–(f) Correlation time  $\tau$  in the AR(1) process. All estimates are performed subsequent to a 4-month coarse graining. Data product shown in the titles.

location, we apply the likelihood ratio (LR) model selection test (see Fig. 3c). We observe that AR(1) processes are preferred over an fGn in much of the Pacific Ocean, while fGn models are preferred in the North Atlantic and over the adjacent continents.

In Fig. 3d we have combined Figs. 3a and 3b so that the  $p$  value for the preferred model is plotted in each grid point. When combining the two models we have more grid points with significant warming than what is obtained using the fGn null hypothesis, but less than inferred from the AR(1) null model.

#### e. Comparisons of the datasets

To add robustness to the results presented in the previous section, we have repeated the same regional statistical analysis on the datasets from HadCRUT4, Berkeley Earth, and NOAA MLOST. The trends and standard deviations are shown in Fig. 4 and summarized in Table 2. The persistence parameters are shown in Fig. 5. For the GISS dataset, these estimates are shown in Fig. 2. The most notable difference between the four

data products is in the southern oceans. This can be seen by comparing the persistence parameters, and also the standard deviations.

In Fig. 6d the statistical significance of the trends, based on the best null model, are shown for HadCRUT4, Berkeley Earth, and NOAA MLOST data. The patterns are similar to what we found for the GISS data, where the largest domains of insignificant trends are found in the Pacific and North Atlantic Oceans. Table 3 shows the percentages of trends that are significant. The relative frequency of significant trends, at the 5% significance level tested against the best null model, is approximately 80% for all the data products. The HadCRUT4 data shows the smallest percentage (70%) of significant trends, but this can be understood from the difference in spatial coverage. See Fig. 6d.

## 5. Summary and discussion

This paper studies climate variability after 1900 using simple stochastic models and four different data

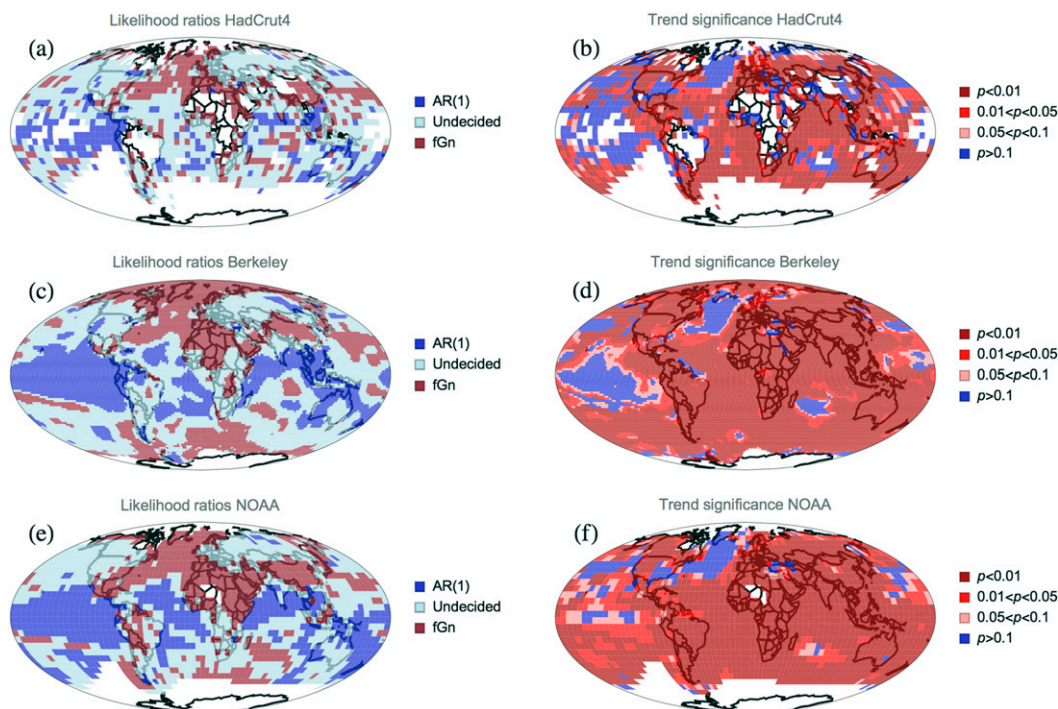


FIG. 6. (a),(c),(e) The results of the likelihood ratio model selection test. In the grid points marked as red (blue) the data are more consistent with an fGn error model [AR(1) error model]. In the grid points marked as light blue, one model is not significantly preferred over the other. (b),(d),(f) The distribution of  $p$  values when the model with the highest likelihood is chosen as the null model in each grid point. The  $p$  values are adjusted for multiple testing using the FDR method. Data product shown in the titles.

products. The results are in many respects similar for the four data products, although there are some differences that are discussed in section 4e.

One of our main focuses has been statistical significance testing of regional temperature trends in this time period with an LRD representation of the internal climate variability. Several studies have presented such detection analysis for a few selected locations, and an advantage of this study is that we get a global overview of local and regional climate variability.

Bloomfield (1992) has shown that the GMST trend is significantly different from zero. Our study confirms this conclusion with an updated estimate of the GMST trend

of  $0.08 \pm 0.03 \text{ K decade}^{-1}$ . Here, the error bars indicate the 95% confidence interval under the assumption of a linear trend superposed on long-range dependent (LRD) stationary fluctuations, which in this work is represented by the fGn model. Under the same assumption we have shown that the  $p$  value (the probability of a fGn producing pseudotrends larger than the observed warming) is less than  $10^{-4}$ .

For regional surface temperatures we find that approximately 80% of the analyzed grid cells have significant warming trends. This number is obtained from first choosing the best null model [fGn or AR(1)] based on a likelihood-ratio criteria, and subsequently applying a

TABLE 3. Percentage significant trends at the 1% ( $p < 0.01$ ) and 5% ( $p < 0.05$ ) significance level assuming an fGn null hypothesis and AR(1) null hypothesis. In the last column (preferred model) the trend significance is tested against the model selected by the likelihood-ratio criteria. The  $p$  values are adjusted for multiple testing using the FDR method.

	fGn		AR(1)		Preferred model	
	$p < 0.01$	$p < 0.05$	$p < 0.01$	$p < 0.05$	$p < 0.01$	$p < 0.05$
GISS	47%	67%	85%	92%	68%	81%
Berkeley Earth	51%	68%	88%	92%	73%	84%
HadCRUT4	36%	55%	78%	85%	56%	70%
NOAA MLOST	47%	67%	89%	94%	72%	85%

trend detection using the most appropriate model. This approach is preferable compared to the standard method, which is to restrict the analysis to a single class of models (e.g., fGn). The main reason for this is that some regions, in particular those strongly influenced by ENSO, show a lack of scaling, while other regions are more consistent with LRD processes.

A similar fraction of grid cells with significant warming trends (about 80%) was also found by Karoly and Wu (2005) for trends over 1903–2002, although a one-sided test was used there. The results of our study, as well as those of Karoly and Wu (2005), Stott et al. (2010), and Knutson et al. (2013), are evidence that global warming is observable on regional scales.

The regions where we do not have warming trends, or where we cannot establish significance of the warming trends, can be identified with feedback mechanisms in the ocean dynamics. In fact, the lack of warming trends in the North Atlantic basin can partly be explained by the 60-yr periodicity in the AMO. The AMO began a negative phase around the year 1900, and in the time period 1900–2013 (the period we have analyzed) it had not quite completed two full cycles. Consequently, the AMO has a negative contribution to the SST trends over the period.

Another region where we cannot establish significant warming trends is the in the equatorial Pacific Ocean, specifically its eastern part (see, e.g., Fig. 3d). This is related to the so-called Pacific cold tongue, which is a region around the equator west of South America that experiences cooling relative to the other regions of the Pacific Ocean. The phenomenon is produced by upwelling of cold water in the eastern Pacific and its amplification by the trade winds. Our results for this region are consistent with a study of Zhang et al. (2010), where principal component analysis is used to discern a spatial pattern for the variations in the SST over the last century, and where the Pacific cold tongue is identified in the second orthogonal function mode. Climate models show that the cooling mode is not observed in the pre-industrial period, and therefore it might be seen as a negative dynamical feedback to global warming (Zhang et al. 2010).

In a wider perspective, this paper presents a simple methodology for accurately quantifying the local and regional temperature variability on centennial time scales. Several authors have used climate models to determine the relative role of natural variations to the overall uncertainty in the climate predictions for the next century (see, e.g., Monier et al. 2015; Deser et al. 2012, 2014). In these studies, the natural variability is defined as the variations of the individual runs around the ensemble means. The obvious advantage of climate

models in this respect is the availability of a large number of runs, which makes it possible to construct ensemble means. When analyzing the instrumental temperature records, we only have a single realization at each location, and we have to apply different methods in order to separate internal climate variability from the climate system's response to the anthropogenic changes in radiative forcing. This separation of signals into noise terms (internal variability) and trends is exactly what is done in trend significance testing, and hence this paper contains a description of natural climate variability, including its dependence on geographic location and how its fluctuation levels depend on time scale. Our study can be seen as a complement to the ongoing efforts of using climate models to quantify uncertainty in future climate projections.

*Acknowledgments.* This work has received support from the Norwegian Research Council under Contract 229754/E10. We thank the referees for useful comments that helped improve the paper. The authors also acknowledge useful discussions with Kristoffer Rypdal and Hege-Beate Fredriksen.

#### REFERENCES

- Bloomfield, P., 1992: Trends in global temperature. *Climatic Change*, **21**, 1–16, doi:10.1007/BF00143250.
- Bunde, A., and S. Lennartz, 2012: Long-term correlations in earth sciences. *Acta Geophys.*, **60**, 562–588, doi:10.2478/s11600-012-0034-8.
- Cohn, T. A., and H. F. Lins, 2005: Nature's style: Naturally trendy. *Geophys. Res. Lett.*, **32**, L23402, doi:10.1029/2005GL024476.
- Deser, C., A. Phillips, V. Bourdette, and H. Teng, 2012: Uncertainty in climate change projections: The role of internal variability. *Climate Dyn.*, **38**, 527–546, doi:10.1007/s00382-010-0977-x.
- , —, M. A. Alexander, and B. V. Smoliak, 2014: Projecting North American climate over the next 50 years: Uncertainty due to internal variability. *J. Climate*, **27**, 2271–2296, doi:10.1175/JCLI-D-13-00451.1.
- Embrechts, P., and M. Maejima, 2002: *Self-Similar Processes*. Princeton University Press, 152 pp.
- Franzke, C., 2012: On the statistical significance of surface air temperature trends in the Eurasian Arctic region. *Geophys. Res. Lett.*, **39**, L23705, doi:10.1029/2012GL054244.
- Hansen, J., R. Ruedy, M. Sato, and K. Lo, 2010: Global surface temperature change. *Rev. Geophys.*, **48**, RG4004, doi:10.1029/2010RG000345.
- Hurst, H. E., 1957: A suggested statistical model of some time series which occur in nature. *Nature*, **180**, 494, doi:10.1038/180494a0.
- Huybers, P., and W. Curry, 2006: Links between annual, Milankovitch, and continuum temperature variability. *Nature*, **441**, 329–332, doi:10.1038/nature04745.
- IPCC, 2013: Summary for policymakers. *Climate Change 2013: The Physical Science Basis*, T. F. Stocker et al., Eds., Cambridge University Press, 3–29. [Available online at [https://www.ipcc.ch/pdf/assessment-report/ar5/wg1/WG1AR5\\_SPM\\_FINAL.pdf](https://www.ipcc.ch/pdf/assessment-report/ar5/wg1/WG1AR5_SPM_FINAL.pdf).]

- 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 an update to 2010. *J. Geophys. Res.*, **117**, D05127, doi:10.1029/2011JD017139.
- Karoly, D. J., and Q. Wu, 2005: Detection of regional surface temperature trends. *J. Climate*, **18**, 4337–4343, doi:10.1175/JCLI3565.1.
- Kennedy, J. J., N. A. Rayner, R. O. Smith, D. E. Parker, and M. Saunby, 2011: 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.
- Knutson, T. R., F. Zeng, and A. T. Wittenberg, 2013: Multimodel assessment of regional surface temperature trends: CMIP3 and CMIP5 twentieth-century simulations. *J. Climate*, **26**, 8709–8743, doi:10.1175/JCLI-D-12-00567.1.
- Ko, K., J. Lee, and R. Lund, 2008: Confidence intervals for long memory regressions. *Stat. Probab. Lett.*, **78**, 1894–1902, doi:10.1016/j.spl.2008.01.057.
- Koutsoyiannis, D., and A. Montanari, 2007: Statistical analysis of hydroclimatic time series: Uncertainty and insights. *Water Resour. Res.*, **43**, W05429, doi:10.1029/2006WR005592.
- Lee, J., and R. Lund, 2004: Revisiting simple linear regression with autocorrelated errors. *Biometrika*, **91**, 240–245, doi:10.1093/biomet/91.1.240.
- , and —, 2008: Equivalent sample sizes in time series regressions. *J. Stat. Comput. Simul.*, **78**, 285–297, doi:10.1080/10629360600758484.
- Li, J., C. Sun, and F.-F. Jin, 2013: NAO implicated as a predictor of Northern Hemisphere mean temperature multidecadal variability. *Geophys. Res. Lett.*, **40**, 5497–5502, doi:10.1002/2013GL057877.
- Libardoni, A. G., and C. E. Forest, 2011: Sensitivity of distributions of climate system properties to the surface temperature dataset. *Geophys. Res. Lett.*, **38**, L22705, doi:10.1029/2011GL049431.
- Løvstletten, O., and M. Rypdal, 2012: Approximated maximum likelihood estimation in multifractal random walks. *Phys. Rev. E*, **85**, 046705, doi:10.1103/PhysRevE.85.046705.
- Mandelbrot, B. B., and J. W. Van Ness, 1968: Fractional Brownian motions, fractional noises and applications. *SIAM Rev.*, **10**, 422–437, doi:10.1137/1010093.
- McLeod, I. A., H. Yu, and Z. L. Krougly, 2007: Algorithms for linear time series analysis: With R package. *J. Stat. Softw.*, **23**, 1–26, doi:10.18637/jss.v023.i05.
- Monier, E., X. Gao, J. R. Scott, A. P. Sokolov, and C. A. Schlosser, 2015: A framework for modeling uncertainty in regional climate change. *Climatic Change*, **131**, 51–66, doi:10.1007/s10584-014-1112-5.
- 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.
- Percival, D. B., J. E. Overland, and H. O. Mofjeld, 2001: Interpretation of North Pacific variability as a short- and long-memory process. *J. Climate*, **14**, 4545–4559, doi:10.1175/1520-0442(2001)014<4545:IONPVA>2.0.CO;2.
- Rypdal, K., L. Østvand, and M. Rypdal, 2013: Long-range memory in Earth's surface temperature on time scales from months to centuries. *J. Geophys. Res. Atmos.*, **118**, 7046–7062, doi:10.1002/jgrd.50399.
- Rypdal, M., and O. Løvstletten, 2013: Modeling electricity spot prices using mean-reverting multifractal processes. *J. Phys.*, **392A**, 194–207, doi:10.1016/j.physa.2012.08.004.
- Smith, T. M., and R. W. Reynolds, 2005: A global merged land–air–sea surface temperature reconstruction based on historical observations (1880–1997). *J. Climate*, **18**, 2021–2036, doi:10.1175/JCLI3362.1.
- Stott, P. A., N. P. Gillett, G. C. Hegerl, D. J. Karoly, D. A. Stone, X. Zhang, and F. Zwiers, 2010: Detection and attribution of climate change: A regional perspective. *Wiley Interdiscip. Rev.: Climate Change*, **1**, 192–211, doi:10.1002/wcc.34.
- Vyushin, D. I., P. J. Kushner, and F. Zwiers, 2012: Modeling and understanding persistence of climate variability. *J. Geophys. Res.*, **117**, D21106, doi:10.1029/2012JD018240.
- Zhang, W., J. Li, and X. Zhao, 2010: Sea surface temperature cooling mode in the Pacific cold tongue. *J. Geophys. Res.*, **115**, C12042, doi:10.1029/2010JC006501.