

# Nonlinear Trends, Long-Range Dependence, and Climate Noise Properties of Surface Temperature

CHRISTIAN FRANZKE

*British Antarctic Survey, Cambridge, United Kingdom*

(Manuscript received 26 May 2011, in final form 5 December 2011)

## ABSTRACT

This study investigates the significance of trends of four temperature time series—Central England Temperature (CET), Stockholm, Faraday-Vernadsky, and Alert. First the robustness and accuracy of various trend detection methods are examined: ordinary least squares, robust and generalized linear model regression, Ensemble Empirical Mode Decomposition (EEMD), and wavelets. It is found in tests with surrogate data that these trend detection methods are robust for nonlinear trends, superposed autocorrelated fluctuations, and non-Gaussian fluctuations. An analysis of the four temperature time series reveals evidence of long-range dependence (LRD) and nonlinear warming trends. The significance of these trends is tested against climate noise. Three different methods are used to generate climate noise: (i) a short-range-dependent autoregressive process of first order [AR(1)], (ii) an LRD model, and (iii) phase scrambling. It is found that the ability to distinguish the observed warming trend from stochastic trends depends on the model representing the background climate variability. Strong evidence is found of a significant warming trend at Faraday-Vernadsky that cannot be explained by any of the three null models. The authors find moderate evidence of warming trends for the Stockholm and CET time series that are significant against AR(1) and phase scrambling but not the LRD model. This suggests that the degree of significance of climate trends depends on the null model used to represent intrinsic climate variability. This study highlights that in statistical trend tests, more than just one simple null model of intrinsic climate variability should be used. This allows one to better gauge the degree of confidence to have in the significance of trends.

## 1. Introduction

An important issue in climate science is the identification of significant trends in historical climate time series. The investigation of past climate data is not only important for understanding the climate system but also to test and constrain climate prediction models for future climate projections. While there is unequivocal evidence for global warming as measured by the global mean temperature, it is much harder to identify significant warming signals on regional and local scales. On these scales intrinsic climate variability plays a much larger role than on global scales (Hawkins and Sutton 2009). Thus, it will take a longer time period until the warming signal emerges and can be distinguished from intrinsic climate variability.

In practice it is important to decide if an observed warming is due to external forcing (e.g., greenhouse gas

emissions) or if it is due to intrinsic climate variability, which can arise because climate variables exhibit temporal correlations. If the observed warming is due to external forcing then we will call this a *deterministic trend*; on the other hand, if the warming is due to intrinsic climate variability we will call this a *stochastic trend* (Alexandrov et al. 2008; Fatichi et al. 2009).

The amplitude and duration of stochastic trends depend on the autocorrelation structure of the time series. The more strongly consecutive and far apart values are correlated, the longer and larger the stochastic trends can be. There are two paradigmatic models for the correlation structure of time series: short- and long-range-dependent models (Granger 1980; Hosking 1981; Beran 1994; Stephenson et al. 2000; Percival et al. 2001; Franzke 2010; Franzke et al. 2012). A short-range-dependent model is typically used in climate science studies. The most used paradigmatic model for climate variability is an autoregressive process of first order [AR(1)] (von Storch and Zwiers 1999). This is a short-range-dependent process; that is, to predict the next state one only needs knowledge of the present state. Predictions of a long-range-dependent

---

Corresponding author address: Dr. C. Franzke, British Antarctic Survey, High Cross, Madingley Road, Cambridge CB3 0ET, United Kingdom.  
E-mail: chan1@bas.ac.uk

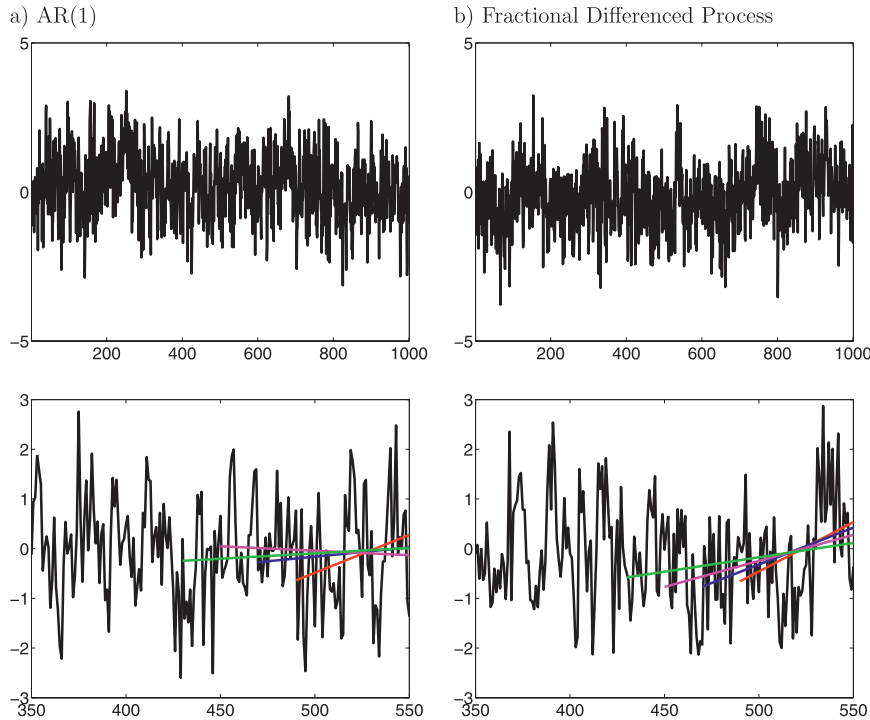


FIG. 1. Examples of two possible models for the atmospheric temperature record and their stochastic trends over different time ranges. (a) A short-range-dependent model [AR(1)]. (b) A long-range-dependent model [ARFIMA(0,  $d$ , 0)] with  $d = 0.28$ . Let us assume that the time units represent years and the stochastic process time series (black solid line) denotes annual mean temperature. Both processes are stationary with zero mean in the asymptotic limit. (top) Thus, there is no trend on long time scales (e.g., over 1000 yr). (bottom) However, over shorter time ranges (e.g., 200 yr) stochastic trends may be found, as illustrated by the colored straight lines, which denote trends derived from linear least squares fits over 30 (red), 50 (blue), 70 (magenta), and 90 (green) yr. Notice how stochastic trends exist over relatively longer time ranges for the long-range-dependent model. In keeping with the way climate studies fit trends, the least squares trends are fitted to the end of the randomly chosen time series chunks. For reference, a 30-yr mean is traditionally viewed as representative of climate.

process need the knowledge of the whole past to predict the next state (Beran 1994). This means that in a long-range-dependent process, even far-apart values are still correlated with each other. The integral over the autocorrelation function of a long-range-dependent process is

$$\lim_{T \rightarrow \infty} \int_{-T}^T \rho(s) ds = \infty, \quad (1)$$

where  $\rho(s)$  is the autocorrelation function at lag  $s$ :

$$\rho(s) = \frac{\int_0^T x(t)x(t+s) dt}{\text{var}(x)}. \quad (2)$$

For a short-range-dependent process the integral in Eq. (1) has a finite value (e.g., Beran 1994; Robinson 2003). These differences in the temporal dependence

structure also lead to different stochastic trend characteristics. This is illustrated in Fig. 1 for AR(1) and a paradigmatic long-range-dependent process. While neither time series realizations have any sizeable trends over the longest time period shown, they can exhibit trends over shorter time periods, as can be seen for the zoomed-in time periods. But these shorter time periods can still be rather long, especially when compared with the length of the observational record of many climate variables. There is increasing evidence that surface temperatures are long-range dependent (Koscielny-Bunde et al. 1998; Gil-Alana 2005; Huybers and Curry 2006; Rybski et al. 2006; Fatichi et al. 2009; Franzke 2010). This makes it more pressing to also test all trends against a long-range-dependent null model so as not to falsely assign a significant deterministic trend to a time series with a stochastic trend (e.g., Fatichi et al. 2009; Franzke 2010).

The statistical identification of significant trends from historical data raises several important issues: (i) how to

define a trend, (ii) how to extract the trend from noisy data, and (iii) how to test its statistical significance.

There is no unique definition of a trend (e.g., Wu et al. 2007; Barbosa 2011). Here we require that a trend is a very smooth monotonic function that has only one extremum within the given observational data span that is near the end points of the time series. A trend also always depends on the method one uses to define and extract the trend. Because of that we will compare different trend detection methods in this study to examine how important this choice is. Furthermore, while the above definition of trend seems to be generally accepted in the climate science community, other communities have a different understanding of trend. For example, in economics the annual cycle is usually referred to as a trend (Alexandrov et al. 2008). In the climate community this is seen as being an oscillatory mode and part of the intrinsic climate variability.

How can we decide if an identified trend is significant? A trend can be considered significant if this trend is unlikely to have occurred by chance. This is determined by comparing the observed trend with stochastic trends generated by a simple stationary stochastic process. The duration and magnitude of these stochastic trends depend on the stochastic process one uses to test the trend against. The stochastic process has to take into account the temporal correlation structure of the background climate variability. Of course, the presence of a deterministic trend will likely impact the temporal correlation structure. But subtracting a stochastic trend from a time series also will influence the temporal correlation structure (Diggle 1990). Because of this we will not detrend the time series before fitting the stochastic processes. This will likely lead to an overestimate of the autocorrelation time scale and increase the magnitude and duration of stochastic trends if a deterministic trend is indeed present. Consequently, our claims about finding a significant trend will be much more conservative.

Because of nonlinear interactions, climate variability operates on a multitude of vastly different temporal and spatial scales. The climate variability on long time scales is usually investigated by examining averaged data (monthly or seasonal means). This is despite the fact that the most dominant time scale of most climate variables is either an annual time scale (i.e., annual cycle; Qian et al. 2011) or on a daily time scale (Feldstein 2000; Franzke 2009; Franzke and Woollings 2011) when considering seasonal data. Thus, part of the observed climate variability on monthly and seasonal time scales stems from the fast weather fluctuations as a result of the averaging. This part of the variability is called climate noise (Leith 1973; Madden 1976; Feldstein 2000; Czaja et al. 2003; Franzke 2009; Franzke and Woollings 2011).

Climate noise is also able to produce stochastic trends (Feldstein 2002; Franzke 2009, 2010). Thus, in this study we will test the observed warming trends against climate noise. To generate surrogate time series of climate noise we will use different paradigmatic models representing short- and long-range-dependent processes and we will also use a nonparametric method to generate surrogate data.

In this study we want to discuss the principal aspects of trend identification and significance testing of trends and the role played by the null model of the background climate variability. In a previous study (Franzke 2010) we focused more on the temperature trends in a particular region (Antarctica) and on just one trend estimation method. Therefore, we examine four exemplary temperature time series, which are presented in section 2. In section 3 we describe various methods for trend identification and in section 4 we discuss significance testing. In section 5 we discuss trend properties of the temperature time series, and we summarize our results in section 6.

## 2. Data

In this study we use daily temperature data from (i) central England for the period 1 January 1772–31 December 2009 (Parker et al. 1992), (ii) Stockholm, Sweden, for the period 1 January 1756–31 December 2009 (Moberg et al. 2002, 2003), (iii) Faraday-Vernadsky (Antarctic Peninsula) for the period 1 January 1951–28 February 2007 (Turner et al. 2004, 2005), and (iv) Alert, Canada, for the period 1 July 1950–30 September 2006. The Faraday-Vernadsky and Alert data are currently only available for these periods. The reason that we focus just on these four time series is that Central England Temperature (CET) and Stockholm are two very long weather station-based temperature time series and Faraday-Vernadsky and Alert are temperature station time series from two polar regions that have experienced some of the most dramatic environmental changes in the last few decades. Hence, these time series constitute a good representation of general temperature variability and trend behavior. All these temperature time series are quality controlled and the CET and Stockholm time series are homogenized. Before the analysis of the temperature data, a mean annual cycle is subtracted by averaging over each 1 January, and so forth. The monthly anomaly temperature time series evaluated here are formed via the monthly averaging of the daily temperature anomaly time series [see Franzke (2010) for more details].

In the following we will estimate the parameters for the statistical null models from the daily data. As

alluded to in the introduction, climate noise can cause apparent trends over finite periods of time. To test the trends against a climate noise null hypothesis we pursue the following approach (Feldstein 2002; Franzke 2009, 2010): we first estimate the trends from the four monthly temperature anomaly time series. We then compute realizations of the three different null models representing daily values and then compute the monthly means of the synthetic data. From these monthly means we estimate the stochastic trends. We then use the distribution of the stochastic trends to decide if the observed temperature trends are significantly different from the climate noise null hypothesis.

### 3. Trend estimation methods

We now discuss the methods we use in this study for trend identification. We will use fitting of a low-order polynomial via various regression methods and wavelets, and also the Ensemble Empirical Mode Decomposition (EEMD) method.

#### a. Ordinary least squares

The most used approach for trend estimation in the climate community is ordinary least squares (OLS), which is a linear regression. This method is able to fit polynomial functions of arbitrary order to data. OLS assumes that a time series can be decomposed into a signal and residuals. It is assumed that the residuals come from a distribution that has a finite variance and are serially uncorrelated. In this study we use linear, quadratic, and cubic polynomials to identify trends. A cubic trend would be of the form

$$y = a + bt + ct^2 + dt^3, \quad (3)$$

and the quadratic and linear trends would have correspondingly fewer parameters.

#### b. Robust regression

Robust regression (Draper and Smith 1998) is a method that is less sensitive to large changes in a small subset of the data. It is a form of weighted least squares regression and is done iteratively. At each iteration step a new set of weights is computed based on the residuals, with larger residuals having smaller weights. Thus the weights depend on the residuals, and consequently large deviations and outliers are down-weighted and have less influence on the regression fit. This constitutes an iterative estimation algorithm. Here we use bisquare weighting (Holland and Welsch 1977) and the weighting function  $w$  is given by

$$w = [\text{abs}(r)(1 - r^2)^2] \quad \text{if } \text{abs}(r) < 1, \quad (4)$$

where  $r$  is given by

$$r = R/[Ts\sqrt{(1 - h)}] \quad (5)$$

and  $R$  represents the residuals,  $h$  represents the leverage values,  $s$  is an estimate of the standard deviation of the error term, and  $T = 4.685$  is the default tuning constant (optimal in the case of Gaussian residuals).

#### c. GLM regression

Generalized linear model (GLM) regression generalizes linear regression by allowing the dependent variable to stem from a distribution from the exponential family (Draper and Smith 1998). The exponential family of distributions includes the Gaussian and gamma distributions. Thus GLM regression should be useful in situations when the distribution of the residuals is non-Gaussian and known.

#### d. EEMD

The Ensemble Empirical Mode Decomposition method (Wu and Huang 2009; Huang et al. 1998; Huang and Wu 2008; Wu et al. 2007; Qian et al. 2009; Franzke 2010) decomposes a time series into a finite number of intrinsic mode functions (IMF) and an instantaneous mean,

$$x(t) = \sum_{j=1}^M \psi_j(t) + R(t), \quad (6)$$

where the  $j$ th IMF  $\psi_j$  can be written in polar coordinates  $\psi_j(t) = r_j(t) \sin[\theta_j(t)]$ , where  $r_j$  is the  $j$ th time-dependent amplitude,  $\theta_j$  the  $j$ th time-dependent frequency, and  $R$  the residual. An IMF is different from Fourier modes for which both  $r_j$  and  $\theta_j$  are time independent. An IMF is defined by the following two properties: (i) each IMF  $\psi_j$  has exactly one zero crossing between two consecutive local extrema (i.e., a sequence of maxima and minima), and (ii) the local mean of each IMF  $\psi_j$  is zero.

With the following algorithm one can estimate IMFs from a given time series (Huang et al. 1998):

- (i) Find all maxima and minima of the time series.
- (ii) Fit a cubic spline through all maxima (minima); these splines define the upper  $e_{\text{up}}$  and lower  $e_{\text{lo}}$  envelope of the time series.
- (iii) Calculate the mean of the upper and lower envelope  $m(t) = [e_{\text{up}}(t) + e_{\text{lo}}(t)]/2$ .
- (iv) Subtract  $m$  from the time series. The resulting curve represents the first IMF. Then go to step (i) and repeat the procedure until the residual is not an IMF anymore.

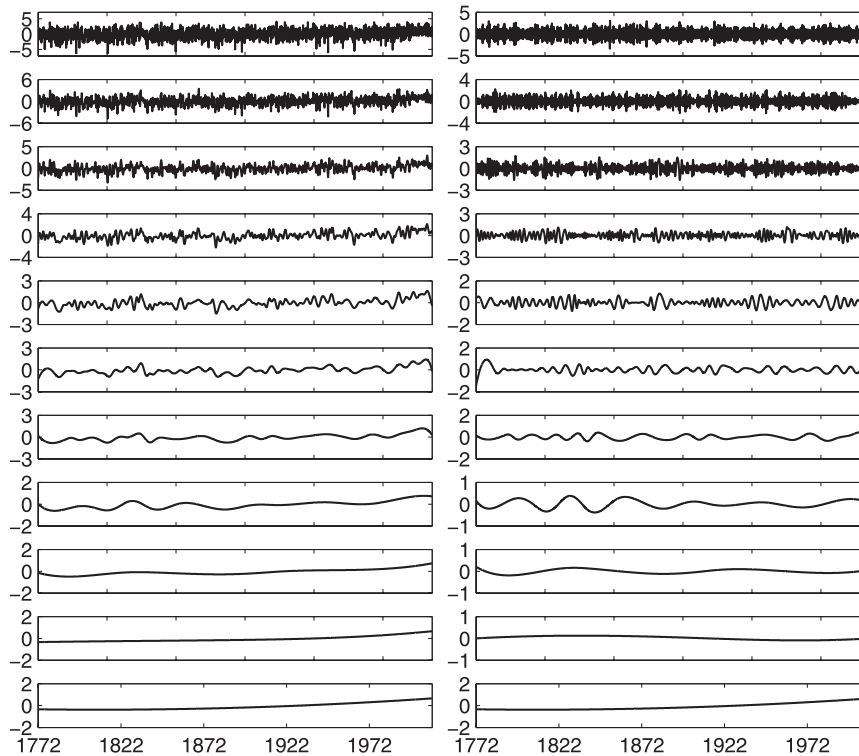


FIG. 2. (top to bottom) (left) Anomalous monthly mean CET time series (mean annual cycle subtracted) and the CET time series with subsequent IMFs subtracted. (right) IMFs of CET time series in descending order.

In practice, the algorithm has to be refined by a so-called sifting process (e.g., Huang et al. 1998), which amounts to iterating steps (i)–(iii) until this can be considered a zero mean to some stopping criterion (Huang et al. 1998; Rilling et al. 2003). Once this is achieved the effective IMF has been determined. The above algorithm is repeated until all IMFs are extracted and the residual is not an IMF anymore; thus, it violates the above two IMF criteria. The residual can now be interpreted as the instantaneous mean of the time series. In case this instantaneous mean is not constant we refer to it as a trend, which is possibly nonlinear (Wu et al. 2007) on the time scale of the time series length.

To avoid mode mixing EEMD adds white noise to the observed time series before the sifting process of the standard Empirical Mode Decomposition (EMD) (Huang et al. 1998; Huang and Wu 2008; Wu et al. 2007; Franzke 2010) and treats the mean of the ensemble as the final IMF. Thus, EEMD is a noise-assisted data analysis method. We utilize 100 ensemble realizations with a noise amplitude of 0.75 standard deviation of the original time series. The results are insensitive to these parameter choices [e.g., using 1000 ensemble realizations does not change the results; see Wu and Huang

(2009) for more details]. EMD and EEMD have been shown to be able to extract nonlinear trends in climatic time series (e.g., Huang et al. 1998; Wu et al. 2007; Wu and Huang 2009; Wu et al. 2011; Franzke 2009, 2010; Franzke and Woollings 2011; Qian et al. 2011).

An example of an EMD decomposition is given in Fig. 2 for the monthly mean CET time series. On the top lhs is displayed the CET time series with the mean annual cycle removed. The IMFs are displayed on the rhs in descending order. IMF1 is displayed on the top rhs and as can be seen contains the highest-frequency fluctuations. The subsequent IMFs contain increasingly lower-frequency fluctuations. The bottom rhs displays the instantaneous mean, which in this study is interpreted as a trend. The panels below the top left display the CET with subsequent IMFs subtracted. Here the filter and smoothing characteristics of the EMD method become visible.

#### e. Wavelets

Wavelets are wavelike functions representing brief oscillations (e.g., Mallat 1999). Here we use the wavelet approach of Andreas and Trevino (1997) to detect linear and quadratic trends. The wavelets used are the inverted Haar wavelet at time  $t$ ,  $I(t, L)$ :

$$A^{-1}I(t, L) = \begin{cases} -1, & -\frac{L}{2} \leq t < 0, \\ 1, & 0 \leq t \leq \frac{L}{2}, \\ 0, & \text{elsewhere} \end{cases} \quad (7)$$

and the elephant wavelet  $P(t, L)$ :

$$B^{-1}I(t, L) = \begin{cases} 1, & -\frac{L}{2} \leq t < -\frac{L}{6}, \\ -2, & -\frac{L}{6} \leq t < \frac{L}{6}, \\ 1, & \frac{L}{6} \leq t < \frac{L}{2}, \\ 0, & \text{elsewhere} \end{cases} \quad (8)$$

where  $A = (2/L)^2$ ,  $B = (3/2L)^3$ , and  $L$  denotes the dilation scale of the wavelet and is the total length of the time series. We use these two particular wavelet functions because the inverted Haar wavelet corresponds to a linear trend and the elephant wavelet to a quadratic trend. The more commonly used Morlet wavelet (Torrence and Compo 1998) does not necessarily extract a smooth trend devoid of oscillations and thus is not used here.

With the help of the above wavelet functions one can estimate the coefficients of a second-order polynomial with less operations than OLS [see Andreas and Trevino (1997) for more details].

*f. Robustness of trend detection*

To test the performance and accuracy of the above trend detection methods we generate synthetic test time series possessing known trend characteristics that were superposed with various types of noise. As trends we use linear, quadratic, cubic, and exponential functions. As noise we use

- independent Gaussian noise;
- independent gamma noise with scale  $\theta = 2$  and shape  $k = 2$  parameters;
- $\alpha$ -stable noise with  $\alpha$  randomly sampled from a uniform distribution  $U(1.75, 2)$ ;
- short-range-dependent noise generated by an autoregressive process of first order with the autoregressive parameter randomly sampled from a uniform distribution  $U(-0.5, 0.5)$  and unit variance white noise;
- long-range-dependent noise generated by an autoregressive fractional integrated moving average process [ARFIMA(0,  $d$ , 0); Stoev and Taqqu 2004; Franzke et al. 2012] with  $d$  randomly sampled from a uniform distribution  $U(0, 0.5)$  and with unit variance white noise.

We use a Monte Carlo approach and generate 100 noise realizations and then calculate the root-mean-square error

(RMSE) between the estimated trends and the true trend. We report here only the results for the quadratic and exponential trends; results for the linear and cubic trends are very similar. Overall, all methods perform reasonably well; the differences in estimated trends are very minor (see Figs. 3, 4). The RMSE is typically one–two orders of magnitude smaller than the amplitude of the superposed noise. As can be seen in Figs. 3, 4 for many cases, the error bounds are overlapping and no method is significantly better than the others. This shows that all methods identify the trends reasonably accurately given the amplitude of the noise.

Unsurprisingly, the parametric regression methods perform slightly better than the EEMD and wavelet methods. The robust regression performs the worst for gamma noise. The ordinary least squares methods are slightly more accurate than the wavelet-based method for linear and quadratic trends. Since the additional computational expense is small, the use of OLS is recommended over the wavelet method. It has to be noted that the regression methods only outperform the nonparametric EEMD if the functional form of the trend is known. Fitting, for example, a straight line to a cubic or exponential trend gives a huge error. This is a potential advantage of the nonparametric EEMD method, which does not a priori assume a functional form of the trend.

Our results show that the influence of deviations from Gaussianity and correlated noise is negligible in our test cases. All methods also work reasonably well in the presence of gamma and  $\alpha$ -stable noise. The gamma variates represent a skewed distribution while  $\alpha$ -stable noise has a power-law decay of its distribution tail and, thus, allows for rare very large values that can be seen as representing outliers. Larger RMSE is produced by  $\alpha$ -stable noise than gamma noise. The  $\alpha$ -stable case is a particularly hard test because of the large number of large values (which in this study will be considered to be outliers), so it is not surprising that it produces the largest RMSE. Robust and GLM regression give similar results as OLS in our experiments. This suggests that deviations from Gaussianity do not overly affect the detection of trends for the methods used here.

**4. Trend significance testing**

In this study we are testing if any detected trends are stochastic trends and in particular if they can be explained as arising from climate noise. This means that we assume a priori that there is no deterministic trend present in the data. Thus, we assume that a priori all detectable trends are stochastic trends. This implies that we fit our stochastic model for the background climate

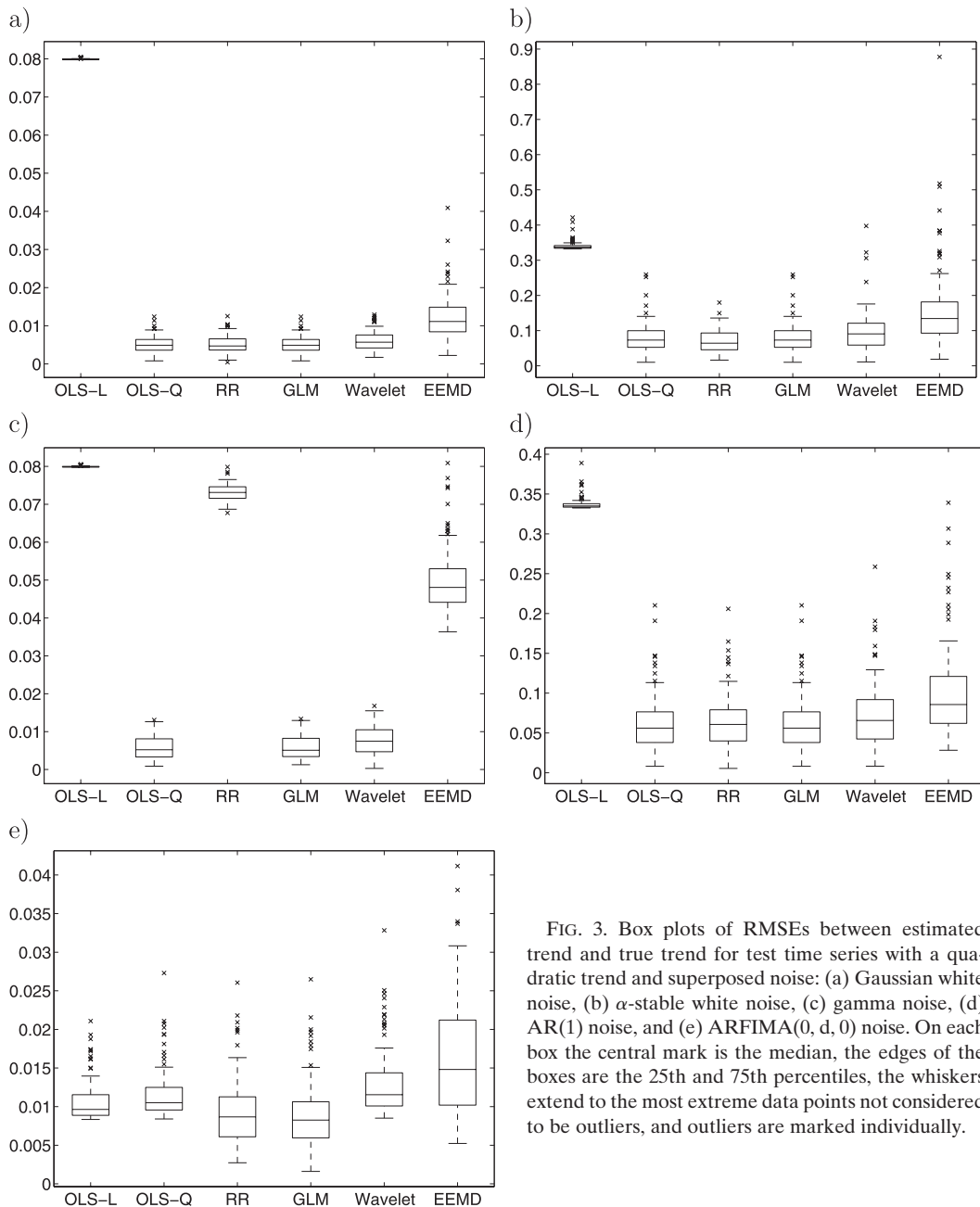


FIG. 3. Box plots of RMSEs between estimated trend and true trend for test time series with a quadratic trend and superposed noise: (a) Gaussian white noise, (b)  $\alpha$ -stable white noise, (c) gamma noise, (d) AR(1) noise, and (e) ARFIMA(0,  $d$ , 0) noise. On each box the central mark is the median, the edges of the boxes are the 25th and 75th percentiles, the whiskers extend to the most extreme data points not considered to be outliers, and outliers are marked individually.

variability to undetrended data because stochastic trends are part of the intrinsic climate variability. If the observed trends are larger than the trends produced by the null models we will claim that the observed trends are significant; that is, they cannot be explained as having arisen from intrinsic climatic fluctuations.

To carry out our trend significance tests we need a model for climate noise. For the purpose of generating surrogate data representing climate noise we use three different approaches: (i) a short-range-dependent model,

(ii) a long-range-dependent model, and (iii) the phase scrambling method. As a short-range-dependent process we use an autoregressive process of first order. This is a standard stochastic process widely used in climate science (von Storch and Zwiers 1999). As a long-range-dependent process we use an autoregressive fractional integrated moving average process (Granger 1980; Hosking 1981; Robinson 2003). The phase scrambling method of Theiler et al. (1992) generates surrogate data that have exactly the same autocorrelation structure as the original

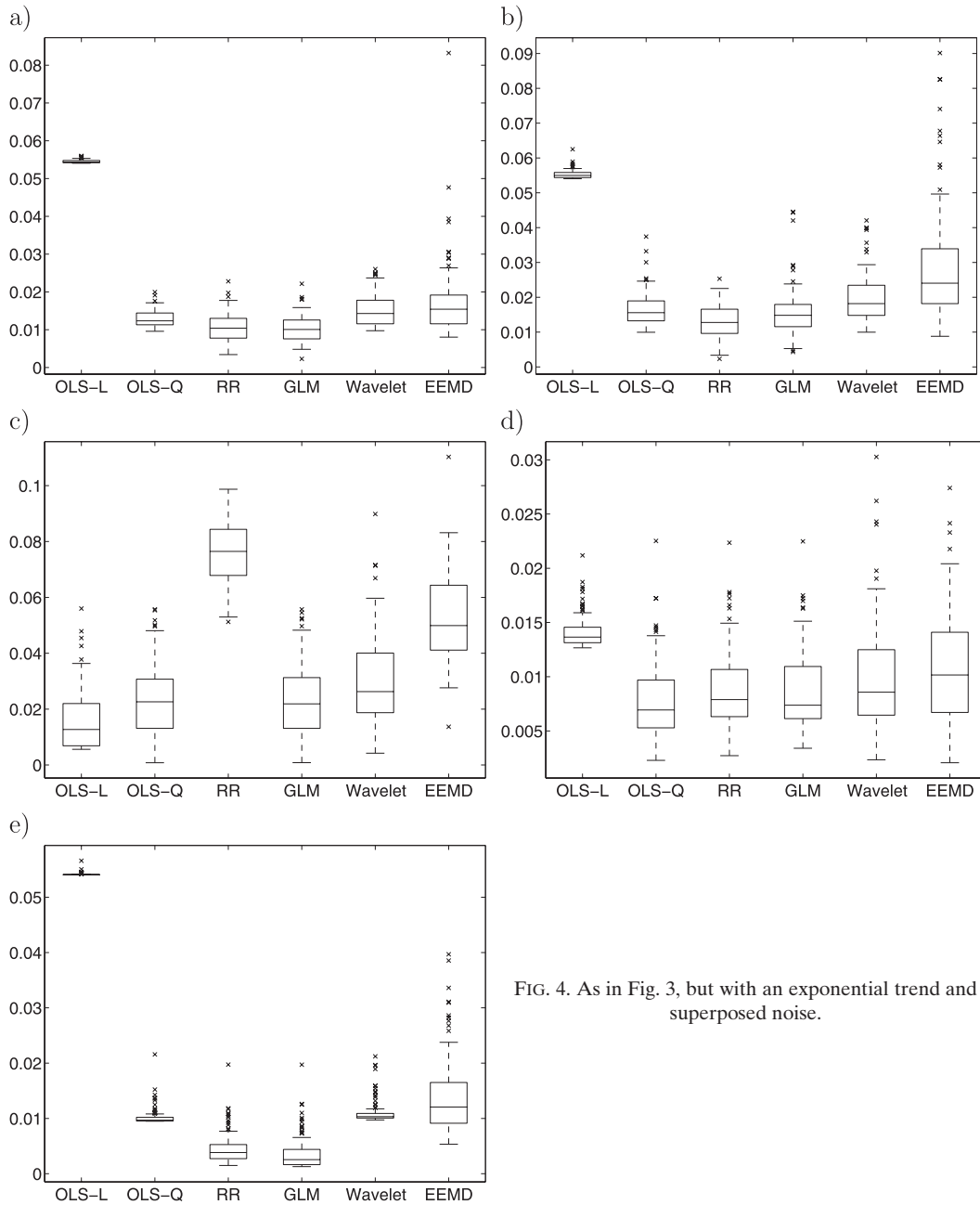


FIG. 4. As in Fig. 3, but with an exponential trend and superposed noise.

data by randomizing the phases. This method first entails a Fourier transformation of the data. The autocorrelation function is uniquely defined by the amplitudes of the Fourier modes. To generate surrogate data, each complex Fourier amplitude is multiplied by  $e^{i\phi}$ , where  $\phi$  is independently chosen from a uniform distribution  $U(0, 2\pi)$  for each frequency. Thus the resulting time series is the sum of randomly phased Fourier components whose amplitudes satisfy the condition that the power spectrum of that time series is identical to the power spectrum of the data time series [see Theiler et al. (1992) for more details].

With these methods we will now perform Monte Carlo experiments by generating 1000 realizations of each of the three surrogate models. The AR(1) estimator (von Storch and Zwiers 1999) and the Geweke-Porter-Hudak (GPH) semiparametric estimator (Geweke and Porter-Hudak 1983; Hurvich and Deo 1999; Franzke 2010; Franzke et al. 2012) are applied to the data to provide parameter estimates and 5% uncertainty bounds for the short- and long-range processes, respectively. The GPH estimator infers the long-range dependence parameter  $d$ , which is a measure of the strength of the temporal dependence or



correlation structure of a time series (e.g., Granger 1980; Hosking 1981; Beran 1994; Franzke 2010; Franzke et al. 2012). A  $d$  value of 0 denotes independent data and larger  $d$  values indicate a strong temporal dependence. A strong dependence means that the past history still influences the future evolution of a system. This characteristic is responsible for long-range-dependent processes exhibiting apparent (stochastic) trends.

For the Monte Carlo experiments we take into account the parameter estimate uncertainty by sampling from a normal distribution with the parameter estimate as the mean and the uncertainty bounds as the variance. The phase scrambling method automatically provides different realizations of surrogate time series for different realizations of the randomly chosen phases.

Then we apply the various trend extraction methods to this ensemble of surrogate data and compute the stochastic trends of the surrogate data. The magnitude of a trend is defined as the range of the trend over the whole time period covered by the corresponding observed time series. If the range of the trend in the observed data is outside the 5th or 95th percentile of the trend ranges computed from the ensemble of surrogate data we claim that the observed trend is statistically significant and is unlikely to have arisen from climate noise based on the used null model.

When comparing the trend significance of different null models one cannot expect that they will always agree. Our null models are also not independent. Trends that are significant against the long-range-dependent models will also be significant against the AR(1) model but not necessarily vice versa. Thus, this requires us to introduce degrees of significance. We will use the following scheme:

- Trend is significant against all three null models: strong evidence of a deterministic trend
- Trend is significant against two null models: moderate evidence of a deterministic trend
- Trend is significant against only one null model: weak evidence of a deterministic trend

In general, one can have more confidence in the significance of an observed trend if this trend is significant against a large number of null models. However, the null models have to be carefully chosen to be suitable and relevant for the problem and the available data.

## 5. Long-range dependence and trend analysis of surface temperature data

First we examine the temperature time series for evidence of long-range dependence. Our analysis finds evidence for long-range dependence for all four temperature time series (Table 1). All long-range-dependence values

TABLE 1. Estimates of the long-range-dependence parameter  $d$  and of the AR(1) coefficient with the corresponding 5% confidence bounds. Also given are the estimated warming values by fitting a cubic polynomial to the data.

	$d$	AR(1)	Trend in °C
CET	$0.19 \pm 0.03$	$0.78 \pm 0.0042$	1.18
Stockholm	$0.28 \pm 0.02$	$0.79 \pm 0.0040$	1.47
Faraday-Vernadsky	$0.28 \pm 0.04$	$0.74 \pm 0.0086$	3.59
Alert	$0.17 \pm 0.05$	$0.79 \pm 0.0084$	0.92

are significantly different from zero at the 5% level and positive. Faraday-Vernadsky and Stockholm have the largest long-range-dependence values at about  $d = 0.28$ . This is consistent with previous studies of surface temperatures that also find evidence of the long-range dependence of temperatures (Koscielny-Bunde et al. 1998; Gil-Alana 2005; Huybers and Curry 2006; Vyushin and Kushner 2009, 2012; Franzke 2010). The fact that all  $d$  values are smaller than 0.5 but positive indicates that all four time series are persistent and stationary. By stationary we mean that they have a constant finite mean and variance. Furthermore, the estimates of the AR(1) parameters give values in the range of 0.74–0.79 (Table 1). It has to be noted that the AR(1) process is better in capturing the initial decay of the autocorrelation function while long-range-dependent processes capture the long-time decay.

Using the methods described above to compute the trends reveals that a least squares cubic polynomial fit has the smallest root-mean-square error for all four time series. This provides evidence for the existence of nonlinear trends in surface temperature. Also, the EEMD trends are nonlinear and very similar to the cubic fits (Fig. 5). Furthermore, all trends correspond to warming trends, with Faraday-Vernadsky exhibiting the largest warming of about 3.59°C (for the period 1951–2007) while the other time series experienced smaller warmings of about 1°–1.5°C (Table 1). The warming at Faraday-Vernadsky is consistent with the warming over the last 120 yr on the Antarctic Peninsula as identified in an ice core (Thomas et al. 2009).

Now that we have found evidence for warming in all four temperature time series we have to check if this warming could have arisen by chance, that is, whether the warming trends are likely deterministic or stochastic trends. Our Monte Carlo experiments show that the warming trend at Faraday-Vernadsky cannot be explained as arising from climate noise for any of the three null models (see Table 2). This provides strong evidence that the trend at Faraday-Vernadsky is a deterministic trend. This is consistent with the findings by Turner et al. (2005) and Franzke (2010).

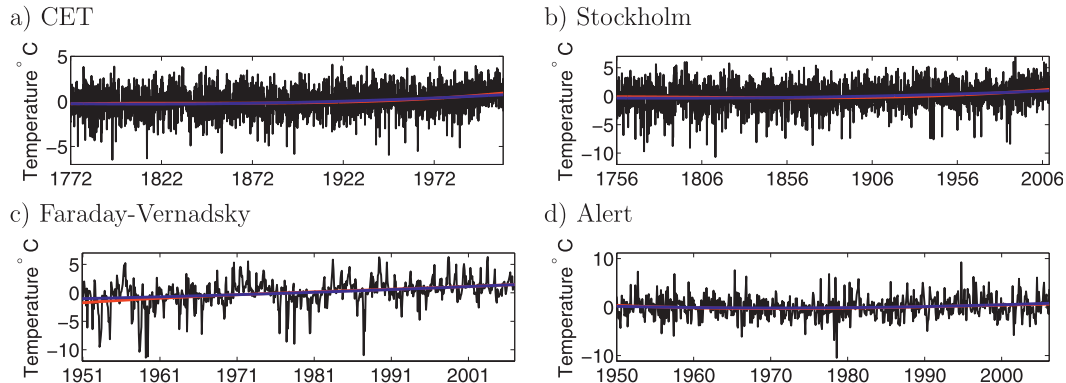


FIG. 5. Monthly mean temperature time series (black line; mean annual cycle subtracted) with cubic trend (red line) and EEMD trend (blue line): (a) CET, (b) Stockholm, (c) Faraday-Vernadsky, and (d) Alert time series.

The warming trends of the CET and Stockholm time series are significant for the AR(1) and phase scrambling tests but not the long-range-dependent model (see Table 2). Thus we find moderate evidence for a significant trend in these two temperature time series. The warming trend at Alert can be explained as having arisen because of climate noise since all three null models are able to produce trends of at least the same magnitude (see Table 2). Thus, we find no evidence for a deterministic trend at Alert.

## 6. Concluding discussion

Our results highlight the importance of the null model to the significance of temperature trends. The used null models have different structures of the decay of the autocorrelation function. This accounts for the fact that highly autocorrelated time series can exhibit stochastic trends over rather long periods of time. This fact has to be taken into account in any significance test of trends. Different paradigmatic null models represent different aspects of the autocorrelation structure and, thus, impact the test of statistical significance of trends. Thus, we recommend using different null models in any significance test of climate trends to see how strong the evidence is for a trend.

We also compared various trend identification methods. Most climate studies use a linear least squares regression fit (e.g., Santer et al. 2000, 2008; Feldstein 2002; Turner et al. 2005), though recently the EMD/EEMD method has become popular in climate research (e.g., Huang et al. 1998; Huang and Wu 2008; Wu et al. 2007; Franzke 2010). We find in tests with synthetic data that various regression methods, EEMD, and wavelets are reliable tools in trend identification and are robust against non-Gaussian and correlated intrinsic fluctuations. We also find evidence for the presence of nonlinear trends. Our results

suggest that one should use low-order polynomials and nonparametric methods like EEMD to compare the results from different trend detection methods.

Here we used three null models representing the background climate variability, two parametric models, and one nonparametric model. We find strong evidence that the observed warming at Faraday-Vernadsky cannot be explained as arising from climate noise. However, the use of a long-range-dependent model for the background climate variability negates the trends of two other temperature time series (CET and Stockholm) when compared with a short-range-dependent model. The short-range-dependent model, an AR(1) process, is the model used in almost all previous temperature trend studies. Using the nonparametric phase scrambling method, which produces surrogates with exactly the same autocorrelation structure as the observed data, provides evidence that CET and Stockholm experience significant warming that cannot be explained as arising from climate noise and, thus, are unlikely to be stochastic trends. Thus, there is moderate evidence for a significant temperature trend in the CET and Stockholm time series.

These results highlight that the correlation and dependence structure of the climate system needs to be much better understood. It is also important to examine if a short- or long-range-dependent model better

TABLE 2. Significance of the trends at the 5% level against the various null models of climate noise. PS denotes the phase scrambling method.

	AR(1)	ARFIMA(0, 1, 0)	PS
CET	X	—	X
Stockholm	X	—	X
Faraday-Vernadsky	X	X	X
Alert	—	—	—

describes internal climate variability. Studies by Percival et al. (2001) and Vyushin and Kushner (2012) addressed this question by comparing the fits of an AR(1) and fractional differenced model; both models contain two parameters that need to be fitted. Percival et al. (2001) conclude that the current climate record is too short to prefer one model over the other, while Vyushin and Kushner (2012) conclude that both paradigmatic models are inadequate for representing internal climate variability. This suggests that either multivariate (von Storch and Zwiers 1999), higher-order autoregressive (von Storch and Zwiers 1999), or other models that combine short- and long-range dependence like ARFIMA( $p, d, q$ ) or nonlinear stochastic models (e.g., Majda et al. 2008, 2009) are needed. This question is part of our ongoing research and will be reported on elsewhere.

These results also call for the investigation and the procurement of new high-resolution, long, climatic time series like ice cores and other climate proxies. The longer the record is, the more unlikely it is to falsely classify a stochastic trend as a deterministic trend or vice versa. An analysis of an ice core from the Antarctic Peninsula at Gomez shows strong evidence of long-range dependence and also of a deterministic trend over the last 120 yr (Thomas et al. 2009). Because of the proximity of Gomez to the Faraday-Vernadsky station, this provides further evidence that the observed warming of the Antarctic Peninsula is a deterministic trend and not due to natural fluctuations.

*Acknowledgments.* I thank Mervyn Freeman for many stimulating discussions and three anonymous reviewers for their valuable comments, which improved an earlier version of this manuscript. This study is part of the British Antarctic Survey Polar Science for Planet Earth Programme. It was funded by the Natural Environment Research Council. I thank the Met Office and the NCAS British Atmospheric Data Centre for providing the CET data, Environment Canada for providing the Alert data, and SMHI for the Stockholm data.

## REFERENCES

- Alexandrov, T., S. Bianconcini, E. B. Dagum, P. Maass and T. S. McElroy, 2008: A review of some modern approaches to the problem of trend extraction. Research Rep. Series (Statistics #2008-3), Statistical Research Division, U.S. Census Bureau, 32 pp.
- Andreas, E. L., and G. Trevino, 1997: Using wavelets to detect trends. *J. Atmos. Oceanic Technol.*, **14**, 554–564.
- Barbosa, S. M., 2011: Testing for deterministic trends in global sea surface temperature. *J. Climate*, **24**, 2516–2522.
- Beran, J., 1994: *Statistics for Long Memory Processes*. Chapman & Hall/CRC, 315 pp.
- Czaja, A., A. W. Robertson, and T. Huck, 2003: The role of Atlantic ocean-atmosphere coupling in affecting North Atlantic Oscillation variability. *The North Atlantic Oscillation: Climatic Significance and Environmental Impact, Geophys. Monogr.*, Vol. 134, Amer. Geophys. Union, 147–172.
- Diggle, P. J., 1990: *Time Series*. Oxford Science Publications, 257 pp.
- Draper, N. R., and H. Smith, 1998: *Applied Regression Analysis*. Wiley Interscience, 706 pp.
- Fatichi, S., S. M. Barbosa, E. Caporali, and M. E. Silva, 2009: Deterministic versus stochastic trends: Detection and challenges. *J. Geophys. Res.*, **114**, D18121, doi:10.1029/2009JD011960.
- Feldstein, S. B., 2000: The timescale, power spectra, and climate noise properties of teleconnection patterns. *J. Climate*, **13**, 4430–4440.
- , 2002: The recent trend and variance increase of the annular mode. *J. Climate*, **15**, 88–94.
- Franzke, C., 2009: Multi-scale analysis of teleconnection indices: Climate noise and nonlinear trends. *Nonlinear Processes Geophys.*, **16**, 65–76.
- , 2010: Long-range dependence and climate noise characteristics of Antarctic temperature data. *J. Climate*, **23**, 6074–6081.
- , and T. Woollings, 2011: On the persistence and predictability properties of North Atlantic variability. *J. Climate*, **24**, 466–472.
- , T. Graves, N. W. Watkins, R. Gramacy, and C. Hughes, 2012: Robustness of estimators of long-range dependence and self-similarity under non-Gaussianity. *Philos. Trans. Roy. Soc. A.*, **370**, 1250–1267.
- Geweke, J., and S. Porter-Hudak, 1983: The estimation and application of long memory time series models. *J. Time Ser. Anal.*, **4**, 221–238.
- Gil-Alana, L. A., 2005: Statistical modeling of the temperatures in the Northern Hemisphere using fractional integration techniques. *J. Climate*, **18**, 5357–5369.
- Granger, C. W. J., 1980: Long memory relationship and the aggregation of dynamic models. *J. Econ.*, **14**, 227–238.
- Hawkins, E., and R. Sutton, 2009: The potential to narrow uncertainty in regional climate predictions. *Bull. Amer. Meteor. Soc.*, **90**, 1095–1107.
- Holland, P. W., and R. E. Welsch, 1977: Robust regression using iteratively reweighted least-squares. *Commun. Stat. Theory Methods*, **6**, 813–827.
- Hosking, J. R. M., 1981: Fractional differencing. *Biometrika*, **68**, 165–176.
- Huang, N. E., and Z. Wu, 2008: A review on Hilbert-Huang Transform: Method and its applications to geophysical studies. *Rev. Geophys.*, **46**, RG2006, doi:10.1029/2007RG000228.
- , and Coauthors, 1998: The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. Roy. Soc. London*, **A454**, 903–995.
- Hurvich, C. M., and R. S. Deo, 1999: Plug-in selection of the number of frequencies in regression estimates of the memory parameter of a long-memory time series. *J. Time Ser. Anal.*, **20**, 331–341.
- Huybers, P., and W. Curry, 2006: Links between annual, Milankovitch and continuum temperature variability. *Nature*, **441**, 329–332, doi:10.1038/nature04745.
- Koscielny-Bunde, E., A. Bunde, S. Havlin, H. Eduardo Roman, Y. Goldreich, and H.-J. Schellnhuber, 1998: Indication of a universal persistence law governing atmospheric variability. *Phys. Rev. Lett.*, **81**, 729–732.
- Leith, C. E., 1973: The standard error of time-average estimates of climatic means. *J. Appl. Meteor.*, **12**, 1066–1069.

- Madden, R. A., 1976: Estimates of the natural variability of time-averaged sea-level pressure. *Mon. Wea. Rev.*, **104**, 942–952.
- Majda, A. J., C. Franzke, and B. Khouider, 2008: An applied mathematics perspective on stochastic modelling for climate. *Philos. Trans. Roy. Soc.*, **A366**, 2427–2453, doi:10.1098/rsta.2008.0012.
- , —, and D. Crommelin, 2009: Normal forms for reduced stochastic climate models. *Proc. Natl. Acad. Sci. USA*, **106**, 3649–3653, doi:10.1073/pnas.0900173106.
- Mallat, S., 1999: *A Wavelet Tour of Signal Processing*. 2nd ed. Academic Press, 637 pp.
- Moberg, A., H. Bergström, J. Ruiz Krigsman, and O. Svanered, 2002: Daily air temperature and pressure series for Stockholm (1756–1998). *Climatic Change*, **53**, 171–212.
- , H. Alexandersson, H. Bergström, and P. D. Jones, 2003: Were southern Swedish summer temperatures before 1860 as warm as measured? *Int. J. Climatol.*, **23**, 1495–1521.
- Parker, D. E., T. P. Legg, and C. K. Folland, 1992: A new daily central England temperature series, 1772–1991. *Int. J. Climatol.*, **12**, 317–342, doi:10.1002/JOC.3370120402.
- 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.
- Qian, C., C. Fu, Z. Wu, and Z. Yan, 2009: On the secular change of spring onset at Stockholm. *Geophys. Res. Lett.*, **36**, L12706, doi:10.1029/2009GL038617.
- , —, and —, 2011: Changes in the amplitude of the temperature annual cycle in China and their implication for climate change research. *J. Climate*, **24**, 5292–5302.
- Rilling, G., P. Flandrin, and P. Goncalves, 2003: On empirical mode decomposition and its algorithms. *Proc. Workshop on Nonlinear Signal and Image Processing NSIP-03*, Grado, Italy, IEEE-EURASIP, 8–11.
- Robinson, P. M., 2003: Long-memory time series. *Time Series with Long Memory*, P. M. Robinson, Ed., Oxford University Press, 4–32.
- Rybski, D., A. Bunde, S. Havlin, and H. von Storch, 2006: Long-term persistence in climate and the detection problem. *Geophys. Res. Lett.*, **33**, L06718, doi:10.1029/2005GL025591.
- Santer, B. D., T. M. L. Wigley, J. S. Boyle, D. J. Gaffen, J. J. Hnilo, D. Nychka, D. E. Parker, and K. E. Taylor, 2000: Statistical significance of trends and trend differences in layer-average atmospheric temperature time series. *J. Geophys. Res.*, **105**, 7337–7356.
- , and Coauthors, 2008: Consistency of modelled and observed temperature trends in the tropical troposphere. *Int. J. Climatol.*, **28**, 1703–1722.
- Stephenson, D. B., V. Pavan, and R. Bojariu, 2000: Is the North Atlantic Oscillation a random walk? *Int. J. Climatol.*, **20**, 1–18.
- Stoev, S., and M. S. Taqqu, 2004: Simulation methods for linear fractional stable motion and FARIMA using the Fast Fourier Transform. *Fractals*, **12**, 95–121, doi:10.1142/S0218348X04002379.
- Theiler, J., S. Eubank, A. Longtin, B. Galdrikian, and J. Doyne Farmer, 1992: Testing for nonlinearity in time series: The method of surrogate data. *Physica D*, **58**, 77–94.
- Thomas, E. R., P. F. Dennis, T. J. Bracegirdle, and C. Franzke, 2009: Ice core evidence for significant 100-year regional warming on the Antarctic Peninsula. *Geophys. Res. Lett.*, **36**, L20704, doi:10.1029/2009GL040104.
- Torrence, C., and G. Compo, 1998: A practical guide to wavelet analysis. *Bull. Amer. Meteor. Soc.*, **79**, 61–78.
- Turner, J., and Coauthors, 2004: The SCAR READER Project: Toward a high-quality database of mean Antarctic meteorological observations. *J. Climate*, **17**, 2890–2898.
- , and Coauthors, 2005: Antarctic climate change during the last 50 years. *Int. J. Climatol.*, **25**, 279–294.
- von Storch, H., and F. W. Zwiers, 1999: *Statistical Analysis in Climate Research*. Cambridge University Press, 484 pp.
- Vyushin, D. I., and P. J. Kushner, 2009: Power-law and long-memory characteristics of the atmospheric general circulation. *J. Climate*, **22**, 2890–2904.
- , and —, 2012: Modelling and understanding persistence of natural climate variability. *J. Geophys. Res.*, in press.
- Wu, Z., and N. E. Huang, 2009: Ensemble empirical mode decomposition: A noise-assisted data analysis method. *Adv. Adapt. Data Anal.*, **1**, 1–41.
- , —, S. R. Long, and C. K. Peng, 2007: On the trend, detrending, and variability of nonlinear and nonstationary time series. *Proc. Natl. Acad. Sci. USA*, **104**, 14 889–14 894.
- , —, J. M. Wallace, B. V. Smoliak, and X. Chen, 2011: On the time-varying trend in global-mean surface temperature. *Climate Dyn.*, **37**, 759–773.