

Statistical significance of rising and oscillatory trends in global ocean and land temperature in the past 160 years

L. Østvand¹, K. Rypdal², and M. Rypdal²

¹Department of Physics and Technology, UiT The Arctic University of Norway, Norway

²Department of Mathematics and Statistics, UiT The Arctic University of Norway, Norway

Correspondence to: Lene Østvand
(lene.ostvand@uit.no)

Abstract. Various interpretations of the notion of a trend in the context of global warming are discussed, contrasting the difference between viewing a trend as the deterministic response to an external forcing and viewing it as a slow variation which can be separated from the background spectral continuum of long-range persistent climate noise. The emphasis in this paper is on the latter notion, and a general scheme is presented for testing a multi-parameter trend model against a null hypothesis which models the observed climate record as an autocorrelated noise. The scheme is employed to the instrumental global sea-surface temperature record and the global land temperature record. A trend model comprising a linear plus an oscillatory trend with period of approximately 70 yr, and the statistical significance of the trends, are tested against three different null models: first-order autoregressive process, fractional Gaussian noise, and fractional Brownian motion. The parameters of the null models are estimated from the instrumental record, but are also checked to be consistent with a Northern Hemisphere temperature reconstruction prior to 1750 for which an anthropogenic trend is negligible. The linear trend in the period 1850-2010 AD is significant in all cases, but the oscillatory trend is insignificant for ocean data and barely significant for land data. However, by using the significance of the linear trend to constrain the null hypothesis, the oscillatory trend in the land record appears to be statistically significant. The results suggest that the global land record may be better suited for detection of the global warming signal than the ocean record.

1 Introduction

At the surface of things, the conceptually simplest approach to detection of anthropogenic global warming should be the estimation of trends in global surface temperature throughout the instrumental observation era starting in the mid-nineteenth century. These kinds of estimates, however, are subject to deep controversy and confusion originating from disagreement about how the notion of a trend should be understood. In this paper we adopt the view that there are several, equally valid, trend definitions. Which one that will prove most useful depends on the purpose of the analysis and the availability and quality of observation data.

At the core of the global change debate is how to distinguish anthropogenically forced warming from natural variability. A complicating factor is that natural variability has forced as well as internal components. Power spectra of climatic time series also suggest to separate internal dynamics into quasi-coherent oscillatory modes and a continuous and essentially scale-invariant spectral background. Over a vast range of time scales this background takes the form of a persistent, fractional noise or motion (*Lovejoy and Schertzer, 2013; Markonis and Koutsoyannis, 2013*). Hence, the issue is threefold: (i) to distinguish the climate response to anthropogenic forcing from the response to natural forcing, (ii) to distinguish internal dynamics from forced responses, and (iii) to distinguish quasi-coherent, oscillatory modes from the persistent noise background. This conceptual structure is illustrated by the Venn diagram in Figure 1a. Figure 1b illustrates three possible trend notions based on this picture. Fundamental for all is the separation of the observed climate record into a trend component (also termed the *signal*) and a *climate noise* component. The essential difference between these notions is how to make this separation.

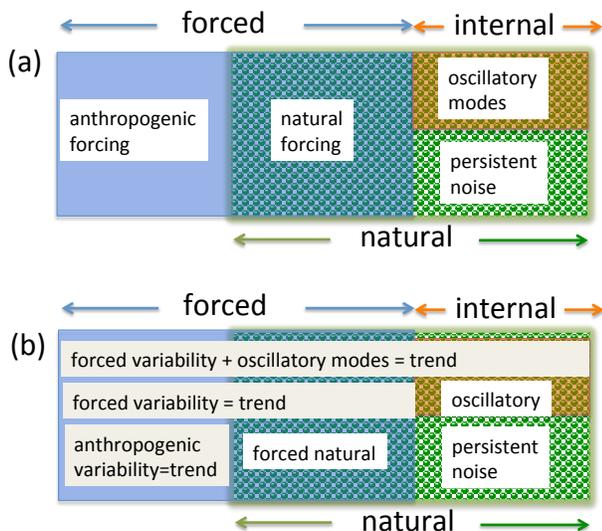


Fig. 1. Venn diagrams illustrating the interplay between forced, internal, and natural variability and various definitions of trend. (a): Natural variability can be both forced and internal. Forced variability can be both anthropogenic and natural. Internal variability is natural, but can consist of quasiperiodic oscillatory modes as well as a continuum of persistent noise. (b): The three different trend notions discussed in the text.

The widest definition of the trend is to associate it with all forced variability and oscillatory modes as illustrated by the upper row in Figure 1b. With this notion the methodological challenge will be to develop a systematic approach to extract the trend from the observed record, and then to subtract this component to establish the persistent noise component. The physical relevance of this separation will depend on to what extent we can justify to interpret the extracted trend as a forced response with internally generated oscillatory modes superposed. If detailed information on the time evolution of the climate forcing is not used or is unavailable such a justification is quite difficult. In this case we will first construct a parametrized model for the trend based on the appearance of the climate record at hand and our physical insight about the forcing and the nature of the dynamics. The next step will be to estimate the parameters of the trend model by conventional regression analysis utilizing the observed climate record. The justification of interpreting this trend as something forced and/or coherent different from background noise will be done through a test of the null hypothesis which states that the climate record can be modeled as a long-range memory (LRM) stochastic process. Examples of such processes are persistent fractional Gaussian noises (fGns) or fractional Brownian motions (fBms). LRM processes exhibit stronger random fluctuations on long time scales than short-memory processes and hence a null model based on LRM-noise will

make it more difficult to reject the null hypothesis for a given estimated trend. For comparison we will also test the null hypothesis against a conventional short-memory notion of climate noise, the first-order autoregressive process (AR(1)). In general, rejection of the null hypothesis will be taken as an acceptance of the hypothesis that the estimated trend is significant, and will strengthen our confidence that these trends represent identifiable dynamical features of the climate system.

A trend can be rendered significant under the AR(1) null hypothesis, but insignificant under an LRM-hypothesis, and then it could of course be argued that the value of this kind of analysis of statistical significance is of little interest, unless one can establish evidence that favors one null model over another. One can, however, test the null models against the observation data, and here analysis seems to favor the fGn/fBm models over short-memory models. There are dozens of papers that demonstrate scaling properties consistent with fGn or fBm properties in instrumental temperature data (see *Rypdal et al. (2013)* for a short review and some references). But, since the instrumental records may be strongly influenced by the increasing trend in anthropogenic forcing, it is difficult to disentangle LRM introduced by the forcing from that arising from internal, unforced variability. Detrending methods such as the detrended fluctuation analysis (*Kantelhardt et al., 2001*) are supposed to do this, but the short duration of the instrumental records does not seem to allow us to make an undisputable distinction between AR(1) and fGn/fBm. We analyze this issue in section 3.3, where we also comment the methods and conclusions in a recent study by (*Vyushin et al., 2012*).

There are also other approaches that favor the LRM models for description of random internal variability in global data on time scales from months to centuries. One is based on analysis of temperature reconstructions for the last millennium prior to the anthropocene (*Rybski et al., 2006; Rypdal et al., 2013*). These temperature data are not influenced by an anthropogenic trend, but exhibit self-similar scaling properties with spectral exponent $\beta \approx 1$ (to be explained in section 2) on time scales at least up to a century. Short-memory processes like the AR(1) will typically exhibit scaling with $\beta \sim 2$ up to the autocorrelation time, and a flat ($\beta \sim 0$) spectrum on time scales longer than this, but this is not observed in these data. Another line of investigation has been to use available time-series information about climate forcing in a parametrized, linear, dynamic-stochastic model for the climate response (*Rypdal and Rypdal, 2013*). The trend then corresponds to the deterministic solution to this model, i.e., the solution with the known (deterministic) component of the forcing. In this model the persistent noise component of the temperature record is the response to a white noise stochastic forcing. In (*Rypdal and Rypdal, 2013*) analysis of the residual obtained by subtracting the deterministic forced solution from the observed instrumental global temperature record

shows scaling properties consistent with an fGn model and inconsistent with an AR(1) model.

The method is described in *Rypdal and Rypdal (2013)*, where only exponential and scale-free long-range persistent responses are modeled, without allowing for quasi-coherent oscillations. The approach in that paper adopts the trend definition described in the second row of Figure 1b. Here the trend is the forced variability, while all unforced variability is relegated to the realm of climate noise. It is possible, however, to incorporate forced and natural oscillatory dynamics into such a response model.

The lower row in Figure 1b depicts the trend notion of foremost societal relevance; the forced response to anthropogenic forcing. Once we have estimated the parameters of the forced response model, we can also compute the deterministic response to the anthropogenic forcing separately. One of the greatest advantages of the forced-response methodology is that it allows estimation of this anthropogenic trend/response and prediction of future trends under given forcing scenarios, subject to rigorous estimates of uncertainty. On the other hand, that method is based on the assumption that the forcing data employed are correct. The construction of forcing time series relies heavily on uncertain observations and modeling, hence there is an obvious case for complementary approaches to trend estimation that do not rely on this kind of information. This is the approach that will be explored in the present paper.

2 Trend Detection Methodology

2.1 The null models

The noise modeling in this paper makes use of the concept of long-range memory (LRM), or (equivalently) long-term persistence (LTP) (*Beran, 1994*). In global temperature records this has been studied in e.g., *Pelletier and Turcotte (1999)*; *Lennartz and Bunde (2009)*; *Rybski et al. (2006)*; *Rypdal and Rypdal (2010)*; *Efstathiou et al. (2011)*; *Rypdal et al. (2013)*; *Rypdal and Rypdal (2013)*. Emanating from these studies is the recognition that ocean temperature is more persistent than land temperature and that the 20th century rising trend is stronger for land than for ocean. LRM is characterized by a time-asymptotic ($t \rightarrow \infty$) autocorrelation function (ACF) of power-law form $C(t) \sim t^{\beta-1}$ for which the integral $\int_0^\infty C(t)dt$ diverges. Here β is a power-law exponent indicating the degree of persistence. The corresponding asymptotic ($f \rightarrow 0$) power spectral density (PSD) has the form $S(f) \sim f^{-\beta}$, hence β is also called the spectral index of the LRM process. For $0 < \beta < 1$ the process is stationary and is termed a persistent fGn. For $1 < \beta < 3$ the process is non-stationary and termed an fBm. As a short-memory alternative we shall also consider the AR(1) process which has an exponentially decaying ACF and is completely character-

ized by the lag-one autocorrelation ϕ (*von Storch and Zwiers, 1999*).

2.2 Previous work using LRM null models

Bloomfield and Nychka (1992) studied the significance of a linear trend in 128 years of global temperature assuming different stochastic models, including fractionally integrated white noise. They found that the trend in the record could not be explained as natural variability by any of the models.

Significance of linear trends under various null models, some exhibiting LRM, was also studied by *Cohn and Lins (2005)*. One of their main points was that trends classified as statistically significant under a short-memory null hypothesis might end up as insignificant under an LRM hypothesis. The paper is a theoretical study of trend significance and is motivated by the strong persistence which is known to exist in hydroclimatic records. As an example they study the Northern Hemisphere (NH) temperature record and find that their test renders the trend insignificant under the LRM null hypothesis. They conclude that the trend *might* be due to natural dynamics. Analyses with similar and other methodologies on other records indicate that the global trend signal is significant in spite of LRM (*Gil-Alana, 2005*; *Rybski et al., 2006*; *Lennartz and Bunde, 2009*; *Halley and Kugiumtzis, 2011*; *Rypdal et al., 2013*). We show in the present paper that the global land temperature record turns out to exhibit a stronger trend and weaker LRM than the NH temperature which is sufficient to establish trend significance. In contrast, the weaker trend and stronger LRM of global ocean temperature yield a less significant trend for this signal.

Some recent papers on LRM and trends are *Fatichi et al. (2009)*; *Rybski et al. (2009)*; *Franzke (2009, 2010)*; *Franzke and Woollings (2011)*; *Franzke (2012a,b)*; *Franzke et al. (2012)*. *Fatichi et al. (2009)* and *Rybski et al. (2009)* study station temperatures under different LRM null hypotheses, and find significant linear trends in some, but not all, of the records. *Franzke (2012b)* applies a methodology similar to that of *Cohn and Lins (2005)* to single-station temperature records in the Arctic Eurasian region. He emphasizes that almost all stations show a positive trend, and that the melting of Arctic sea ice leaves no doubt about the reality of an anthropogenic warming signal in the Arctic. By evaluating all station data together, for instance by analysing the regional averaged temperature, one would most likely arrive at a significant trend. His point is that the natural variability for single stations is so large and long-range correlated that it may mask the warming signal at the majority of individual stations at the present stage of global warming. This is an important message to convey to those policymakers who have got the impression that local climate projections universally are sufficiently reliable to implement adaptive measures.

2.3 Hypothesis testing methodology

In the present paper our main objective is to establish beyond doubt the significance of the global warming signal, and if possible also the multidecadal oscillation. From the studies discussed above, we know that there are many temperature records from which this significance cannot be established under an LRM null hypothesis, so we should search for a signal that is optimal for trend detection. Such an optimal signal seems to be the instrumental global land temperature record HadCRUT3 (Jones *et al.*, 2012). We will contrast this with analysis of the global ocean record (Kennedy *et al.*, 2011). These records are land-air and sea-surface temperature anomalies relative to the period 1961-90, with monthly resolution from 1850 to date. The analysis is made using a trend model which contains a linear plus a sinusoidal trend, although the methodology developed works for any parametrized trend model. We test this model against the null model that the full temperature record is a realization of an AR(1) process, an fGn, or an fBm (the fBm model is of interest only for the strongly persistent ocean data).

The significance tests are based on generation of an ensemble of synthetic realizations of the null models; AR(1) processes ($\phi < 1$), fGns ($0 < \beta < 1$), and fBms ($1 < \beta < 3$). Each realization is fully characterized by a pair of parameters; $\theta \equiv (\sigma, \phi)$ for AR(1) and $\theta \equiv (\sigma, \beta)$ for fGn and fBm, where σ is the standard deviation of the stationary AR(1) and fGn processes and the standard deviation of the differenced fBm. For an LRM null model the estimated value of $\hat{\beta}$ depends on which null model (fGn or fBm) one adopts. As we will show below, for ocean data, it is not so clear whether an fGn or an fBm is the most proper model (Lennartz and Bunde, 2009; Rypdal *et al.*, 2013), so we will test the significance of the trends under both hypotheses.

Technically, we make use of the R package by McLeod *et al.* (2007) to generate synthetic fGns and to perform a maximum-likelihood estimation of β . Since generation of fBms is not included in this package, synthetic fBms with memory exponent $1 < \beta < 3$ are produced by generating an fGn with exponent $\beta - 2$ and then forming the cumulative sum of that process. This is justified because the one-step differenced fBm with $1 < \beta < 3$ is an fGn with memory exponent $\beta - 2$ (Beran, 1994). Maximum-likelihood estimation of β for synthetic fBms and observed data records modeled as an fBm is done by forming the one-time-step increment (differentiation) process, estimate the memory exponent β_{incr} for that process and find $\beta = \beta_{\text{incr}} + 2$. There are some problems with this method when $\beta \approx 1$. Suppose we have a data record (like the global ocean record) and we don't know whether $\beta < 1$ or $\beta > 1$. For all estimation methods there are large errors and biases for short data records of fGns/fBms for $\beta \approx 1$ (Rypdal *et al.*, 2013). This means that there is an ambiguity as to whether a record is a realization of an fGn or an fBm when we obtain estimates of β in the vicinity of 1. For the MLE method this ambiguity becomes apparent from Figure

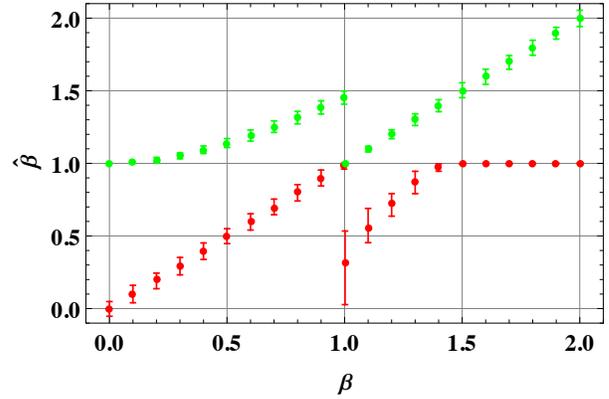


Fig. 2. The red symbols and 95% confidence intervals represent the maximum-likelihood estimate $\hat{\beta}$ for realizations of fGns/fBms with memory parameter β by adopting an fGn model. Hence, for $\beta > 1$ we find the estimate $\hat{\beta}$ from a realization of an fBm with a model that assumes that it is an fGn. The green symbols represent the corresponding estimate by adopting an fBm model, i.e., for $\beta < 1$ we find the estimate $\hat{\beta}$ from a realization of an fGn with a model that assumes that it is an fBm. “Adopting an fBm model” means that the synthetic record is differentiated, then analyzed as an fGn by the methods of McLeod *et al.* (2007) to obtain $\hat{\beta}_{\text{incr}}$, and then finally $\beta = \hat{\beta}_{\text{incr}} + 2$.

2. Here we have plotted the MLE estimate $\hat{\beta}$ with error bars for an ensemble of realizations of fGns (for $0 < \beta < 1$) and of fBms ($1 < \beta < 2$) with 2000 data points. The red symbols are obtained by adopting an fGn model when β is estimated. Hence, for $\beta > 1$ we find the estimate $\hat{\beta}$ from a realization of an fBm with a model that assumes that it is an fGn. It would be expected that the analysis would give $\hat{\beta} \approx 1$ for an fBm, but we observe that it gives $\hat{\beta}$ considerably less than 1 in the range $1 < \beta < 1.4$, so if we observe a β in the vicinity of 1 by this analysis we cannot know whether it is an fGn or an fBm. The ambiguity remains by estimating with a model that assumes that the record is an fBm, because this yields a corresponding positive bias as shown by the green symbols when the record is an fGn. This ambiguity seems difficult to resolve for ocean data as short as the monthly instrumental record.

The standard method for establishing a trend in time-series data is to adopt a parametrized model $T(A;t)$ for the trend, e.g., a linear model $A_1 + A_2 t$ with parameters $A = (A_1, A_2)$, and estimate the model parameters by a least-square fit of the model to the data. Another method, which brings along additional meaning to the trend concept, is the MLE method. This method adopts a model for the stochastic process; $x(t) = T(A;t) + \sigma w(t)$, where $w(t)$ is a correlated or uncorrelated random process and establishes the set of model parameters A for which the likelihood of the stochastic model to produce

the observed data attains its maximum. The method applied to uncorrelated and Gaussian noise models is described in many standard statistics texts (*von Storch and Zwiers*, 1999), and its application to fGns is described in *McLeod et al.* (2007). If $w(t)$ is a Gaussian, independent and identically distributed (i.i.d.) random process, the MLE is equivalent to the least-square fit. If $w(t)$ is a strongly correlated (e.g., LRM) process, and the trend model provides a poor description of the large-scale structures in the data, MLE may assign more weight to the random process (greater σ) than the least-square method. On the other hand, if the trend model is chosen such that it can be fitted to yield a good description of the large-scale structure, the parameters estimated by the two methods are quite similar, even if $w(t)$ used in the MLE method is an LRM process. In this case we can use least-square fit to establish the trend parameters without worrying about whether the residual noise obtained after subtracting the estimated trend can be modeled as a Gaussian, i.i.d. random process.

In the following, we make some definitions and outline the methodology we adopt to assess the significance of the estimated trend. Concepts defined are named with bold-face fonts. Our methodology is based on standard hypothesis testing, where the trend hypothesis (termed the “alternative hypothesis”) is accepted (although not verified, which is stronger) by rejection of a “null hypothesis.” Failure of rejection of the null hypothesis implies failure of acceptance of the alternative hypothesis, and hence the trend will be characterized as insignificant under this null hypothesis. Hence, it is clear that the outcome of the significance test will depend on the choice of alternative hypothesis (trend model) as well as on the null hypothesis (noise model).

Let us take the pragmatic point of view that a trend is a simple and slowly varying (compared with a predefined time scale τ) function $T(A; t)$ of t , parametrized by the trend coefficients $A = (A_1, \dots, A_n)$. It is also required that for the optimal choice of parameters, $A = \hat{A}_{\text{obs}}$ the trend $T(\hat{A}_{\text{obs}}; t)$ makes a good fit to the large-scale structure of the data record. In practice, this means that the trend should be close to a low-pass filtered version of the signal, for instance a moving average over time-scale τ . The trend is significant with respect to a particular null model if the fitted $T(\hat{A}_{\text{obs}}; t)$ is very unlikely to be realized in an ensemble of fits $T(\hat{A}; t)$ to realizations of the null model.

The alternative hypothesis can be formulated as follows: The observed record $x(t)$ is a realization of the stochastic process

$$T(A; t) + \sigma w(t), \quad (1)$$

where the trend $T(A; t)$ is a specified function of t depending on the trend coefficients $A = (A_1, \dots, A_n)$, and $w(t)$ is a Gaussian stationary random process of unit variance. These coefficients are estimated from a least-square fit to $x(t)$ and have the values \hat{A}_{obs} . We assume that the trend model can be

fitted so well to the data that MLE-estimates of A with different noise models (white noise vs. strongly persistent fGn) give approximately the same \hat{A}_{obs} .

The null hypothesis states that the record $x(t)$ is a realization of a stochastic process

$$\varepsilon(\theta; t), \quad (2)$$

with certain properties to be specified (e.g., the process is AR(1), fGn, or fBm). Like for the alternative hypothesis, the parameters θ should be restricted to be close to the values $\hat{\theta}_{\text{obs}}$ found from estimating it from fitting the null model (2) to the data record by means of MLE.

The Monte Carlo null ensemble is the collection of realizations $x_i(\theta)$, $i = 1, 2, \dots$, of the null model process (2).

The best choice of null model would be to utilize all our possible knowledge about the true parameter set θ . This implies considering θ as a random variable, and hence a Bayesian approach (*Gelman et al.*, 2004). We generate the null ensemble by drawing θ from the conditional distribution $P(\theta|\hat{\theta}_{\text{obs}})$, i.e., the probability that the “real” parameters of the observed process are θ given that the estimated parameters from the observed data are $\hat{\theta}_{\text{obs}}$. One way of establishing this distribution is to generate an ensemble of realizations of the noise process with θ varied in a range around $\theta \approx \hat{\theta}_{\text{obs}}$ and establish the conditional distribution $P(\hat{\theta}|\theta)$. From Bayes’ theorem one then has $P(\theta|\hat{\theta}) = P(\hat{\theta}|\theta)P(\theta)/P(\hat{\theta})$. By setting $\hat{\theta} = \hat{\theta}_{\text{obs}}$, and assuming a flat prior distribution $P(\theta)$ in the range in the vicinity of θ_{obs} corresponding to the width of the distribution we find $P(\theta|\hat{\theta}_{\text{obs}}) = P(\hat{\theta}_{\text{obs}}|\theta)$.

As an alternative to the Bayesian ideas described above one could employ a frequentist approach. This means that we assume that the null model has a fixed true parameter value θ . This parameter value is unknown, and the strategy is to create the Monte Carlo null ensemble $x_i(\hat{\theta}_{\text{obs}})$, $i = 1, 2, \dots$, using the θ -values estimated from the observed data. We must then take the uncertainty in the θ -estimates into account, since $\hat{\theta}_{\text{obs}}$ may deviate from the true θ . This estimation error can be quantified using the bootstrap method, which assumes that the error in the parameter estimates in the null model with parameters θ can be well approximated by the corresponding errors for the null model with parameters $\hat{\theta}_{\text{obs}}$. When estimation errors are quantified one can easily adjust for these in the hypothesis tests.

Pseudotrend estimates $\hat{A}^{(i)}$ are the coefficients obtained by least-square fit of the trend model $T(A; t)$ to the realizations $x_i(\theta; t)$ of the null ensemble.

Pseudotrend distribution is the n -dimensional PDF $P(\hat{A})$ over the null ensemble.

Null-hypothesis confidence region is the region Ω in n -dimensional A -space for which $P(A) > P_{\text{thr}}$, where P_{thr} is chosen such that $\int_{\Omega} P(A) dA = 0.95$.

Significance of the trend model is established if the null hypothesis is rejected, e.g., the full n -dimensional trend is 95% significant if $\hat{A}_{\text{obs}} \notin \Omega$.

If the null hypothesis is rejected by this procedure, we are rejecting only those aspects of the null model that are relevant to the full trend model, i.e., the trend model (alternative hypothesis) produces trend coefficients \hat{A}_{obs} that give a good fit to the large-scale structure of the data, while it is very improbable that the null model can produce \hat{A} in the vicinity of \hat{A}_{obs} .

2.4 The trend model explored in this work

We will apply the method described in the previous subsection to global temperature record using the following trend model:

$$T(A; t) = \delta + A_1 t + A_2 \sin(2\pi f t + \varphi). \quad (3)$$

This is a simplified version of the models used in several works by N. Scafetta (e.g., *Scafetta*, 2011, 2012) and the oscillation is supposed to model the 60-yr cycle observed in the instrumental record (*Schlesinger and Ramankutty*, 1994). The frequency f is not to be considered as a free model parameter to be estimated from the observed record and from realizations of the null ensemble. When estimating pseudotrends it has little meaning to let f be a free parameter, since the synthetic noise records contain no preferred frequencies. We rather treat f as a fixed quantity which is an inherent part of the alternative hypothesis. In practice we select f from a least-square fit of the trend model to the observed record varying all five parameters including f , but this is not essential. We could just as well have hypothesized a reasonable value of f by inspection of the record or from other evidence of this oscillation presented in the literature. The important thing to keep in mind is that the value of f is part of the hypothesis. Of the estimated pseudotrend coefficients $(\hat{A}_1, \hat{A}_2, \hat{\delta}, \hat{\varphi})$ only (\hat{A}_1, \hat{A}_2) quantify the strength of the trend, so the relevant pseudotrend distribution to establish is $P(\hat{A}_1, \hat{A}_2)$ irrespective of the values of irrelevant parameters $(\hat{\delta}, \hat{\varphi})$. Table 1 shows the estimated $\hat{\theta}_{\text{obs}}$ according to the null model in (2) using AR(1), fGn and fBm as the stochastic process $\varepsilon(\theta; t)$. Also in this table are the estimated trend parameters (\hat{A}_1, \hat{A}_2) from applying the trend model in (3) and the period $T = 1/f$ of the oscillatory trend. Since, as mentioned above, this period has been selected from a fitting procedure it has slightly different values for the ocean and land records.

2.5 Results

The results of the analysis are shown in Figure 3. We observe that the trend parameters $(\hat{A}_1, \hat{A}_2)_{\text{obs}}$ are outside the null-hypothesis 95% confidence region for all three noise models and for ocean as well as land records. But we also observe that the significance is more evident for land than for ocean, and is reduced as more strongly persistent noise models are employed. For the fBm model applied to ocean data the trend is barely outside the 95% confidence region.

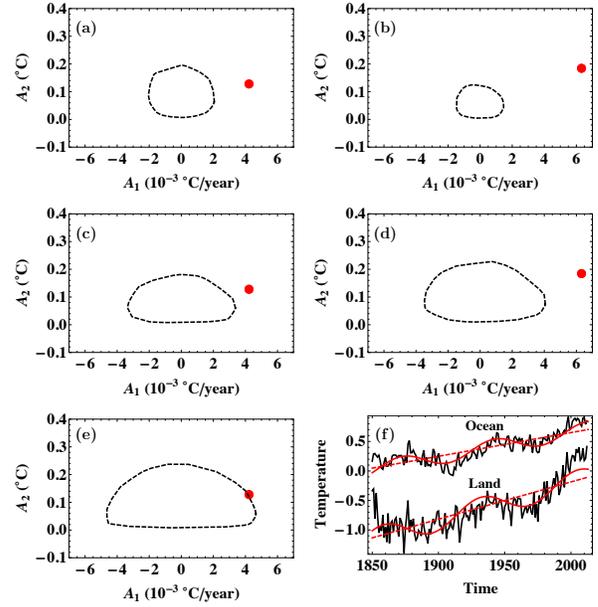


Fig. 3. In panels (a-e) the red dots represent the estimated trend coefficients $(\hat{A}_1, \hat{A}_2)_{\text{obs}}$ and the dashed, closed curve the 95% confidence contour of the distribution $P(\hat{A}_1, \hat{A}_2)$. (a): ocean data and AR(1) null model. (b): land data and AR(1) null model. (c): ocean data and fGn null model. (d): land data and fGn null model. (e): ocean data and fBm null model. (f): Black curves: The global ocean and land temperature records. Red curves: the linear and sinusoidal trends.

It is the full trend model equation (3) that is accepted by this test, but something can also be said about the separate significance of the individual trends represented by the individual trend coefficients from the pseudotrend distribution $P(\hat{A}_1, \hat{A}_2)$. For the AR(1) and fGn null models it is apparent from Figure 3a-d that the linear trend is highly significant since $\hat{A}_{1,\text{obs}}$ is located far to the right of the confidence region. On the other hand, except for the AR(1) model applied to land data in Figure 3b, $\hat{A}_{2,\text{obs}}$ is not totally above the confidence region. This means that the linear pseudotrends observed in the null ensemble has negligible chance of getting near the observed trend, while there is some chance to find oscillatory trends in the null ensemble which are as large as $\hat{A}_{2,\text{obs}}$. The significance of those separate trends against these null models is determined by forming the separate one-dimensional PDFs, $P(\hat{A}_1) \equiv \int P(\hat{A}_1, \hat{A}_2) d\hat{A}_2$ and $P(\hat{A}_2) \equiv \int P(\hat{A}_1, \hat{A}_2) d\hat{A}_1$ and form the confidence intervals in the standard way. In Figure 4 we have formed the corresponding one-dimensional cumulative distribution functions (CDFs) from the two-dimensional PDFs for ocean data shown in Figure 3a, c, and e. We observe that the linear trend is significant for the AR(1) and fGn null models, but barely significant for

Table 1. Estimated noise parameters $\hat{\theta}_{\text{obs}}$ from the null hypotheses in (2) and trend parameters \hat{A}_{obs} estimated from the trend model (3). The units for the trend estimation are months for $\hat{\tau}_{\text{obs}}$, 10^{-3} °C/yr for $\hat{A}_{1,2,\text{obs}}$, and yr for the oscillation period T .

| | AR(1) | fGn | | fBm | | Trend | | |
|-------|---------------------------|----------------------------|-----------------------------|----------------------------|-----------------------------|--------------------------|--------------------------|------|
| | $\hat{\tau}_{\text{obs}}$ | $\hat{\beta}_{\text{obs}}$ | $\hat{\sigma}_{\text{obs}}$ | $\hat{\beta}_{\text{obs}}$ | $\hat{\sigma}_{\text{obs}}$ | $\hat{A}_{1,\text{obs}}$ | $\hat{A}_{2,\text{obs}}$ | T |
| Ocean | 21.3 | 0.994 | 0.25 | 1.45 | 0.086 | 4.21 | 0.128 | 69.7 |
| Land | 3.43 | 0.654 | 0.49 | | | 6.34 | 0.186 | 73.4 |

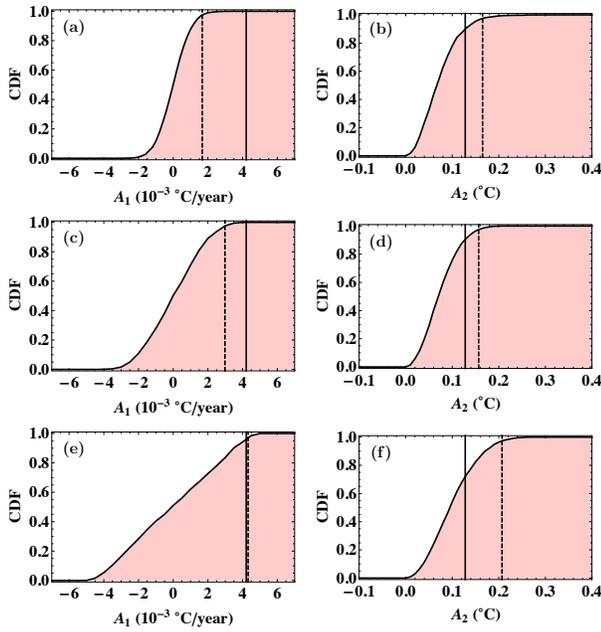


Fig. 4. Curved lines are CDFs for trend coefficients \hat{A}_1 and \hat{A}_2 established from the null model ensemble for ocean data. Vertical dashed line marks the upper 95% confidence limit. Vertical solid line marks $\hat{A}_{1,2,\text{obs}}$. (a) and (b): AR(1) null model. (c) and (d): fGn null model. (e) and (f): fBm null model.

the fBm model. The oscillatory trend is insignificant for all models.

The corresponding CDFs for land data are shown in Figure 5. The linear trend is even more significant than for ocean data, while the oscillatory trend is significant for the AR(1) model, but barely significant for the fGn model.

One important lesson to learn from this analysis is that the stronger persistence in the ocean temperature record makes it harder to detect significant trends as compared to the land record. This effect outweighs the increased trend significance from the lower noise levels in the ocean record compared to the land record. Another is that the land record analysis establishes beyond doubt that there is a significant global linear trend throughout the last century, and that the reality of an

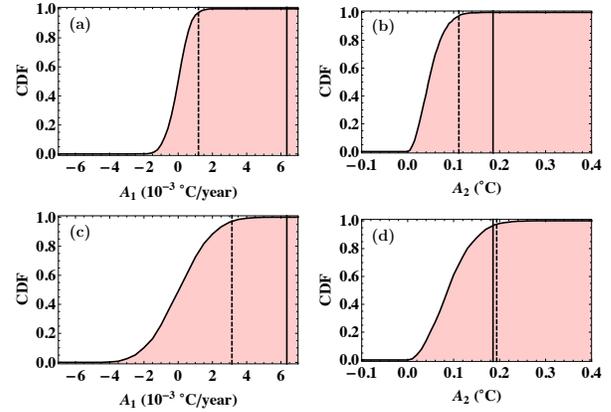


Fig. 5. Curved lines are CDFs for trend coefficients \hat{A}_1 and \hat{A}_2 established from the null model ensemble for land data. Vertical dashed line marks the upper 95% confidence limit. Vertical solid lines mark $\hat{A}_{1,2,\text{obs}}$. (a) and (b): AR(1) null model. (c) and (d): fGn null model.

oscillatory trend is probable, but not beyond the 95% confidence limit.

3 Constraining and Evaluating the Null Hypothesis

By estimating the parameters for the null model from the full observed record (without detrending), and allowing this model to be an LRM noise, we have selected the fractional noise model that is most likely to explain the variance of the full record. Hence it can be considered as the null model for the climate noise that is least likely to be rejected by the observed trend. If this null model is rejected, i.e., if the trend is found to be significant under this null, it is very unlikely that it will be found insignificant under other reasonable null hypotheses. Since we have found that the linear trend in the global land record *is* significant under this null, we should have very high confidence in this result. The non-significance of the oscillatory trend, however, deserves a reassessment in the light of the established significance of the linear trend. In a Bayesian spirit, it would be appropriate to investigate the oscillatory trend further by including the linear trend as

Table 2. Estimated noise parameters $\hat{\theta}_{\text{obs}}$ from the new null hypotheses in (4). The units are same as in Table 1.

| | AR(1) | | fGn |
|------|---------------------------|----------------------------|-----------------------------|
| | $\hat{\tau}_{\text{obs}}$ | $\hat{\beta}_{\text{obs}}$ | $\hat{\sigma}_{\text{obs}}$ |
| Land | 2.04 | 0.584 | 0.391 |

an established fact and construct a null model constrained to accept the existence of the linear trend;

3.1 A constrained null model yields significant oscillation

$$\hat{\delta}_{\text{obs}} + \hat{A}_{1,\text{obs}}t + \varepsilon(\theta; t). \quad (4)$$

We now first estimate a new $\hat{\theta}_{\text{obs}}$ by fitting the new null model (4) to the observed land record. The new estimated noise parameters are shown in Table 2. Then we produce a new null ensemble of records from the null model by drawing θ from the conditional distribution $P(\theta|\hat{\theta}_{\text{obs}})$. Finally we fit the trend model (3) to each realization in the ensemble and form $P(\hat{A}_1, \hat{A}_2)$. The result is shown for land data and $\varepsilon(\theta; t)$ modeled as an fGn in Figure 6a. The inclusion of the linear trend in the null model will imply that we shall fit $\varepsilon(\theta; t)$ to the record $\tilde{x}(t) \equiv x(t) - (\hat{\delta}_{\text{obs}} + \hat{A}_{1,\text{obs}}t)$ rather than to $x(t)$. Since we already have established that $x(t)$ contains a significant linear trend, the variability of $\tilde{x}(t)$ may be considerably less than the variability of $x(t)$ and hence the new estimated noise parameters $\hat{\theta}_{\text{obs}}$ may correspond to smaller $\hat{\sigma}_{\text{obs}}$ and $\hat{\beta}_{\text{obs}}$ than we obtained for the original null model. This reduction in noise parameters leads to narrowing of $P(\hat{A}_1, \hat{A}_2)$, and a narrower CDF for the oscillation trend parameter \hat{A}_2 , as shown in Figure 6b. The result is that this constrained test establishes that the oscillatory trend is also significant.

3.2 Evaluation of the null model

The long-range memory associated with fractional noises and motions gives rise to larger fluctuations on long time scales that allows description of such variability as part of the noise background rather as trends. The implication is that variability which has to be described as significant trends under white noise or short-memory noise hypotheses may have to be classified as insignificant trends under an LRM null hypothesis. The issue of the most proper choice of null hypothesis was touched upon in section 2, but let us re-examine the issue in the light of the results we have obtained so far.

One way to deal with this issue is to apply an estimator that characterizes the correlation structure of the observed record and compare the outcome with those arising from applying the same estimator to different models for the climate-noise background. There are several estimators, for instance wavelet variances and detrended fluctuation analysis, that are well suited for extracting the scaling properties of a time se-

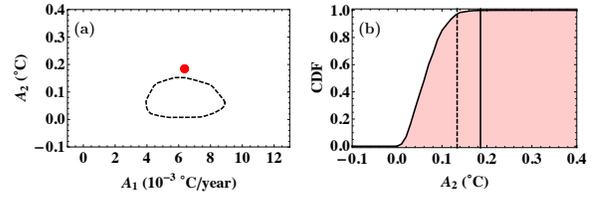


Fig. 6. (a): The 95% confidence contour of the distribution $P(\hat{A}_1, \hat{A}_2)$ for land data obtained by the new null model (4) with $\varepsilon(\theta; t)$ an fGn process. (b): The CDF derived from $P(\hat{A}_2)$ for this null model, with upper 95% confidence limit marked as dotted vertical line.

ries and estimating a β -exponent. For LRM processes such as fBm and fGn (which are respectively self-similar processes and the differences of self-similar processes) the fluctuation level of a time series varies as a power law versus time scale τ , and one can therefore analyze data using double-logarithmic plots of the so-called fluctuation functions. For processes with a characteristic time scale τ_c , such as the AR(1) processes, the fluctuation functions will not be power laws, and this can be seen from the estimated fluctuation functions. For an AR(1) process, which has an autocorrelation function on the form e^{-t/τ_c} , the time series behaves like a Brownian motion ($\beta = 2$) for time scales $t \ll \tau_c$ and a white noise process ($\beta = 0$) for $t \gg \tau_c$. If a time series is sufficiently long, the crossover between these two scaling regimes is clearly visible in the estimated fluctuation functions, and since we do not observe such crossovers in global temperature records, we can use fluctuation functions to illustrate that LRM processes are better suited than AR(1) processes as models for the global temperature. This idea is pursued in (Rypdal and Rypdal, 2013), where detrended fluctuation analysis is employed to show that a residual signal (constructed by subtracting the deterministic response to the external forcing) is inconsistent with an AR(1) process, but consistent with an LRM process.

The test described above utilizes a method designed to estimate the scaling exponent β in LRM processes. As an alternative, we can use a test based on an estimator for the correlation time τ_c in an AR(1) process. For this test we should think of our time series as a discrete-time sampling of a continuous-time stochastic process. The continuous-time analog of an AR(1) process is the Ornstein-Uhlenbeck (OU) process. If a time series T_k is obtained from an OU process by sampling it at times $t_k = k\Delta t$, then the one-lag autocorrelation of T_k is $\phi^{(\Delta t)} = e^{-\Delta t/\tau_c}$. We can obtain a standard sample estimate $\hat{\phi}^{(\Delta t)}$ of the lag-one autocorrelation, and from this we obtain an estimate of the correlation time:

$$\hat{\tau}_c = \frac{\Delta t}{-\log \hat{\phi}^{(\Delta t)}}. \quad (5)$$

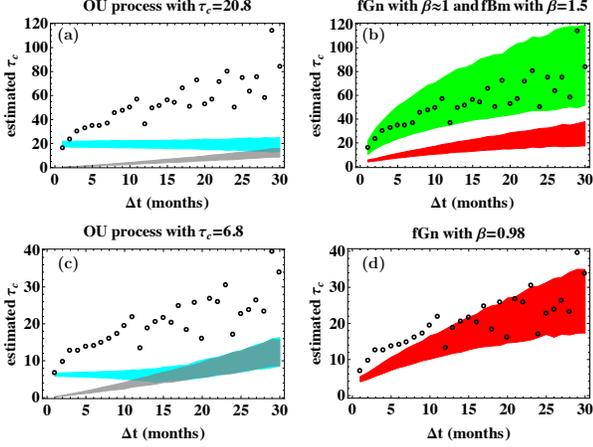


Fig. 7. Panels (a) and (b) show the estimated decorrelation time τ_c as a function of sampling time Δt for the ocean temperature (black circles) and for ensembles of synthetic realizations of three different stochastic processes: An OU process (cyan) in panel (a), and fGns (red) and fBms (green) in panel (b). The synthetic processes are generated with parameters estimated from the observed record by the MLE method, and the colored areas are the 95% confidence regions for these estimates. The gray area in panel (a) is the confidence region for τ_c for a white noise process. Panels (c) and (d) show the decorrelation time of the linearly detrended ocean temperature and for the synthetic realizations of the processes generated from the new null model; equation (4).

Monte Carlo simulations show that this estimate is independent of Δt , as long as $\Delta t < \tau_c$. However, if the process is an fGn rather than an OU process, then the autocorrelation function of the time series T_k is approximated well by $(\beta + 1)\beta(k\Delta t)^{\beta-1}$, and hence the lag-one autocorrelation is

$$\phi^{(\Delta t)} \approx (\beta + 1)\beta\Delta t^{\beta-1}.$$

If τ_c is defined via $\tau_c = \Delta t / (-\log \phi^{(\Delta t)})$, then

$$\tau_c = \frac{\Delta t}{-\log(\beta + 1)\beta - (\beta - 1)\log \Delta t}.$$

This shows that OU processes and fGns can be distinguished by how an estimator of the correlation length depends on the sampling rate for the time series: For an OU process the estimate of τ_c is independent of Δt as long as $\Delta t < \tau_c$, and for fGns the estimates of τ_c grow with Δt . In Figures 7 and 8 we have plotted the estimates of τ_c according to equation (5) for ocean and land temperatures respectively, with and without linear detrending. For the land temperature, full detrending (removing the trend (3)) is also included. The estimates are shown as the circular plot markers in the figures. There is a clear increase in the τ_c estimate as Δt varies from 1 to 30 months. We have compared the results with Monte Carlo simulation of a white noise process, OU processes, fGns and

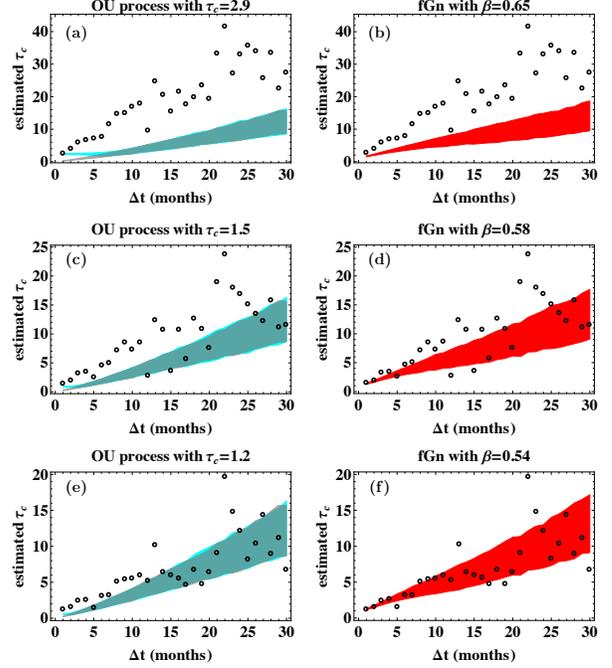


Fig. 8. Panels (a) and (b) show the estimated decorrelation time τ_c as a function of sampling time Δt for the land temperature (black circles) and for ensembles of synthetic realizations of three different stochastic processes: An OU process (cyan) in panel (a), and fGns (red) in panel (b). The synthetic processes are generated with parameters estimated from the observed record by the MLE method, and the colored areas are the 95% confidence regions for these estimates. The gray area in panel (a) is the confidence region for τ_c from a white noise process. Panels (c) and (d) show the decorrelation time of the linearly detrended land temperature and for the synthetic realizations of the processes generated from the new null model; equation (4). Panels (e) and (f) show the decorrelation time of the land temperature after removing the full trend; equation (3), and for the synthetic realizations of the processes generated from the detrended record by the MLE method.

fBms. Here the synthetic temperature series are constructed using parameters obtained by MLE. For the ocean temperature without detrending the test shows that the data is most consistent with a nonstationary fBm, and after linear detrending it is more consistent with an fGn than with an OU process. For the land temperature we observe that neither of the processes fit the data unless we perform a detrending, and for the detrended data there are only small differences between a white noise process, an OU process and the fGn with $\beta = 0.54$. The reason for this is that the ML estimate of τ_c is so small (close to the monthly time resolution of the temperature record) that the model OU process is effectively reduced to a white noise on all resolved time scales. The white noise process is a special case of an fGn, so the fGn class of pro-

cesses is clearly preferred in this case as well, although the test presented here is not suitable for estimating the β exponent. There are other tests that are better suited for accurate estimation of β , and if we apply these we will see that a persistent process ($\beta > 0$) is a better model for detrended land temperatures than white noise ($\beta = 0$) (Rypdal *et al.*, 2013).

The model selection test we have described here illustrates the important point that if one decides to model global temperature fluctuations as OU processes, then the choice of optimal model depends strongly on the time resolution of the time series. The same is not true for fGns and fBms, and this reflects the fact that global temperature data to a good approximation are scale invariant.

The method presented here can be seen as a generalization of the method presented by Vyushin *et al.* (2012), who attempt to distinguish between scale-free processes and AR(1) processes by considering estimates of $\phi^{(\Delta t)}$ for two different time resolutions Δt (monthly and annual). However, our results show that this test fails if the estimated τ_c is less than a year, which turns out to be the case for the land record. Vyushin *et al.* (2012) analyze a large number of local and regional time series and find that some are consistent with fGns, other with AR(1)s, but most are inconsistent with both. It is reasonable to expect that many of these records are in the category for which the test fails.

4 Conclusions

In this paper we have attempted to classify the various possible ways to understand the notion of a trend in the climate context, and then we have focused on the detection of a combination of a rising and oscillatory trend in global ocean and land instrumental data when no information about the climate forcing is used. It is well known that the statistical significance of the trends depends on the degree of autocorrelation (memory) assumed for the random noise component of the climate record (Cohn and Lins, 2005; Rybski *et al.*, 2006, 2009). It is also known that the linear trends are easier to detect and appear to be more significant in global than in local data (Lennartz and Bunde, 2009), although local records exhibit weaker long-term persistence than global records. Despite this fact, much effort is spent on establishing trends and their significance in data from local stations (e.g., Franzke, 2012b) with variable results. The failure of detecting consistent trends in local data records reflects the tendency of internal spatiotemporal variability to mask the trend that signals global warming, and we believe therefore that investigation of such trends should be performed on globally averaged data. For global data records our study demonstrates very clearly that the long-range memory observed in sea-surface temperature data leads to lower significance of detected trends compared to land data. This does not mean, of course, that the global warming signal and internal oscillations are not present in all of those records. It is just

not possible to establish the statistical significance of these trends from these records alone, since the large short-range weather noise in local temperatures and the slower fluctuations in ocean temperature both reduce the possibilities of trend detection. Hence, one needs to search for the optimal climate record to analyze for detection of the global warming signal, and our results suggest that the global land temperature signal may be the best candidate for such trend studies.

While a linear trend is only marginally significant under the long-range memory null hypothesis in ocean data, it is clearly significant in land data. Hence, there should be no doubt about the significance of a global warming signal over the last 160 years even under null hypotheses presuming strong long-range persistence of the climate noise.

Assessment of the statistical significance of a linear trend is of course not the only way to detect the global warming signal in temperature records. An alternative hypothesis in the form of a second- or third-order polynomial trend would give a more precise, but more technically complex assessment. Other approaches are not based on trend estimates at all. Some methods compare spatiotemporal observations to patterns of natural variability obtained from global climate models. These patterns represent the null model, and the detection is typically performed through “fingerprint methods” rather than using just single observable such as the global temperature (Hasselmann, 1993; Hegerl, 1996). The validity of the method depends, of course, on the assumption that the climate model correctly describes the relevant aspects of the pattern of natural variability, e.g., the long-range correlation structure in space and time. This is not an obvious assumption, since there are significant differences between climate models in this respect (Govindan *et al.*, 2001; Blender and Fraedrich, 2003).

Other methods are based on null models like those considered in the present paper, but rather than estimating trends one estimates the probability of observing the recent clustering of record-breaking temperatures at the end of the instrumental record (Zorita *et al.*, 2008). The method is conceptually and technically simpler than the trend assessment, but it depends crucially on the assumption that the null model is strictly true on the shortest inter-annual time scales, since it assumes that the probability of variation from one year to the next is determined by this model. In contrast, the trend assessment emphasizes the properties of the null model on time scales up to a century, so it rather assumes the null model is strictly true on multi-decadal to century scales. The two approaches are complementary, but we believe the trend approach is better designed to detect the smooth, monotonic global warming signal, since it will be insensitive to particular interannual to decadal variability such as ENSO, or variability due to forcing from clusters of volcanic eruptions or solar-cycle variations. The elimination of these variabilities may be important for detection of the anthropogenic trend, as was shown by multiple regression techniques by Foster and Rahmstorf (2011) and Lean and Rind (2009). Moreover,

in the approach of *Zorita et al.* (2008) inclusion of the 70 yr oscillation in the null model would lead to enhanced probability of clustering of record-breaking temperatures at the end of the twentieth century, and hence a reduction of the significance of the warming signal. These are examples illustrating that one may arrive at misleading results without careful selection of the alternative as well as null models based on the data at hand and existing knowledge. In a Bayesian framework this is obvious.

Our initial analysis leaves some doubt about the significance of the 70 yr oscillatory mode in the global signal, as shown in Figure 4d and 4f and Figure 5d. By means of a Bayesian iteration, however, utilizing the established significance of a linear trend to formulate a constrained null hypothesis, we are able to establish statistical significance of the oscillatory trend in the land data record. We believe this is an important result, because it means that we cannot dismiss this oscillation as a spontaneous random fluctuation in the climate noise background. By the analysis presented here we cannot decide whether this oscillation is an internal mode in the climate system or an oscillation forced by some external influence. Such insights can be obtained from a generalization of the response model of *Rypdal and Rypdal* (2013) by employing information about the climate forcing, and will be the subject of a forthcoming paper. There are various published hypotheses about the nature of this oscillation. The least controversial is that this is a global manifestation of the Atlantic Multidecadal Oscillation (AMO) which is essentially an internal climate mode (*Schlesinger and Ramankutty*, 1994). Some authors go further and suggest that this oscillation is synchronized and phase locked with some astronomical influence (*Scafetta*, 2011, 2012). Although some of these suggestions seem very speculative, there are some quite well-documented connections between periodic tidal effects on the Sun from the motion of the giant planets and radioisotope paleorecord proxies for solar activity on century and millennium time scales (*Abreu et al.*, 2012). So far there exists no solid evidence that these, and multidecadal, variations in solar activity have a strong influence on terrestrial climate, but the issue will probably be in the frontline of research on natural climate variability in the time to come. The work presented here cannot shed light on the physical cause of this oscillation, but it presents evidence that it is a phenomenon that stands out from the long-memory background of random temperature fluctuations. Its importance for our assessment of anthropogenic global warming is obvious from the observation that the oscillation seems to peak at the turn of the millennium and hence provides a possible explanation of the current hiatus in global temperature.

Acknowledgements. The authors are grateful to Ola Løvstetten for illuminating discussions and comments.

References

- Abreu, J. A., Beer, J., Ferriz-Mas, A., McCracken, K. G., and Steinhilber F. :, Is there a planetary influence on solar activity?, *Astronomy and Astrophys.*, A88, 548-557, doi:10.1051/0004-6361/201219997, 2012.
- Beran, J.: *Statistics for Long-memory Processes*, Monographs on statistics and applied probability, Chapman& Hall/CRC, Boca Raton, 1994.
- Blender, R. and Fraedrich, K.: Long time memory in global warming simulations, *Geophys. Res. Lett.*, 30, 1769, doi: 10.1029/2003GL017666, 2003.
- Bloomfield, P. and Nychka, K.: Climate spectra and detecting climate change, *Clim. Change*, 21, 275-287, doi: 10.1007/BF00139727, 1992.
- Cohn, T. A. and Lins, H. F.: Nature's style: Naturally trendy, *Geophys. Res. Lett.*, 32, L23402, doi:10.1029/2005GL024476, 2005.
- Efstathiou, M. N., Tzanis, C., Cracknell, A.P., and Varotsos, C. A.: New features of land and sea surface temperature anomalies, *Int. J. Remote Sensing*, 32, 3231-3238, doi:10.1080/01431161.2010.541504, 2011.
- Faticchi, S., Barbosa, S. M., Caporali, E., and Silva, M. E.: Deterministic versus stochastic trends: Detection and challenges, *J. Geophys. Res.*, 114, doi: 10.1029/2009JD011960, 2009.
- Flandrin, P.: Wavelet analysis and synthesis of fractional Brownian motion, *IEEE Trans. Inform. Theory*, 38, 910-917, doi:10.1109/18.119751, 1992.
- Foster, G. and Rahmstorf, S.: Global temperature evolution 1979–2010, *Environ. Res. Lett.*, 6, 044022, doi:10.1088/1748-9326/6/4/044022, 2011.
- Franzke, C.: Multi-scale analysis of teleconnection indices: Climate noise and nonlinear trends, *Nonlinear Processes Geophys.*, 16, 65-76, 2009.
- Franzke, C.: Long-range dependence and climate noise characteristics of Antarctic temperature data, *J. Clim.*, 23, 6074-6081, doi:10.1175/2010JCLI3654.1, 2010.
- Franzke, C. and T. Woollings, T.: On the persistence and predictability properties of North Atlantic climate variability, *J. Clim.*, 24, 466-472, 2011.
- Franzke, C.: Nonlinear trends, long-range dependence and climate noise properties of surface air temperature, *J. Clim.*, 25, 4172-4183, 2012a.
- Franzke, C.: On the statistical significance of surface air temperature trends in the Eurasian Arctic region, *Geophys. Res. Lett.*, 39, L23705, doi:10.1029/2012GL054244, 2012b.
- Franzke, C., Graves, T., Watkins, N.W., Gramacy, R. B., and Huges C.: Robustness of estimators of long-range dependence and self-similarity under non-Gaussianity, *Philos. Trans. R. Soc. A*, 370, 1250-1267, doi:10.1098/rsta.2011.0349, 2012.
- Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B.: *Bayesian Data Analysis*, Texts in Statistical Science Series, Chapman& Hall/CRC, Boca Raton, 2004.
- Gil-Alana, L. A.: Statistical modeling of the temperatures in the Northern Hemisphere using fractional integration techniques, *J. Clim.*, 18, 5357-5369, doi: http://dx.doi.org/10.1175/JCLI3543.1, 2005.
- Govindan, R. B., Vjushin, D., Brenner, S., Bunde, A., Havlin, S., and Schellnhuber H.-J.: Long-range correlations and trends in global climate models: Comparison with real data, *Physica A*, 294, 239-248, 2001.

- Halley, J., and Kugiumtzis, D.: Nonparametric testing of variability and trend in some climatic records, *Clim. Change*, 109, 549-568, doi:10.1007/s10584-011-0053-5, 2011.
- Hasselmann, K.: Optimal fingerprints for the detection of time dependent climate change, *J. Climate*, 6, 1957-1971, 1993.
- Hansen, J. and Sato, M. and Ruedy, R., et al.: Climate simulations for 1880-2003 with GISS modelE, *Climate Dynam.*, 29, doi: 10.1007/s00382-007-0255-8, 2007.
- Hansen, J., Sato, M., Kharecha, P., von Schuckmann, K.: Earth's energy imbalance and implications. *Atmos. Chem. Phys.*, 11, doi:10.5194/acp-11-13421-2011, 2011.
- Hegerl, G. C., von Storch, H., Hasselmann, K., Santer, B. D., Cubasch, U., Jones, P. D.: Detecting greenhouse gas induced climate change with an optimal fingerprint method, *J. Climate*, 9, 2281-2306, 1996.
- Hu, K., Ivanov, P. C., Chen, Z., Carpena, P., and Stanley H. E.: Effect of trends on detrended fluctuation analysis, *Phys. Rev. E*, 64, 011114, doi: 10.1103/PhysRevE.64.011114, 2001.
- Jones, P. D., Lister, D. H., Osborn, T. J., Harpham, C., Salmon, M., and Morice, C. P.: 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, 2012.
- Kantelhardt, J. W., Koscielny-Bunde, E., Rego, H. A., Havlin, S., and Bunde, A.: Detecting long-range correlations with detrended fluctuation analysis, *Physica A*, 295, 441-454, 2001.
- Kennedy, J. J., Rayner, N. A., Smith, R. O., Parker, D. E., and Saunby, M.: 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, 2011.
- Lean, J. L. and Rind, D. H.: How will Earth's surface temperature change in future decades? *Geophys. Res. Lett.*, 36, L15708, doi:10.1029/2009GL038932, 2009.
- Lennartz, S. and Bunde, A.: Trend evaluation in records with long-term memory. Application to global warming, *Geophys. Res. Lett.*, 36, L16706, doi:10.1029/2009GL039516, 2009.
- Lovejoy, S., and Schertzer, D.: *The Weather and Climate: Emergent Laws and Multifractal Cascades*, Cambridge University Press.
- Malamud, B. L., and Turcotte, D.: Self-Affine Time Series: I. Generation and Analyses, *Adv. Geophys.*, 40, 1-90, doi:10.1016/S0065-2687(08)60293-9, 1999.
- Markonis, Y. and Koutsoyiannis, D.: *Climatic Variability Over Time Scales Spanning Nine Orders of Magnitude: Connecting Milankovitch Cycles with Hurst-Kolmogorov Dynamics* (2013), *Surv. Geophys.*, 34, 181-207, doi:10.1007/s10712-012-9208-9, 2013.
- McLeod, A. I., Yu, H., and Krougly, Z. L.: Algorithms for linear time-series analysis, *J. of Statistical Software*, 23, 1-26, 2007.
- Pelletier, J. D., and Turcotte, D.: Self-Affine Time Series: II. Applications and Models, *Adv. Geophys.*, 40, 91-166, doi:10.1016/S0065-2687(08)60294-0, 1999.
- Rybski, D., Bunde, A., Havlin, S., and von Storch, H.: Long-term persistence in climate and the detection problem, *Geophysical Res. Lett.*, 33, L06718, doi: 10.1029/2005GL025591, 2006.
- Rybski, D., and Bunde, A.: On the detection of trends in long-term correlated records, *Physica A*, 388, 1687-1695, doi: 10.1016/j.physa.2008.12.026, 2009.
- Rypdal, M., and Rypdal, K.: Testing hypotheses about sun-climate complexity linking, *Phys. Rev. Lett.*, 104, 128501, doi:10.1103/PhysRevLett.104.128501, 2010.
- Rypdal, K. (2012) Global temperature response to radiative forcing: Solar cycle versus volcanic eruptions, *J. Geophys. Res.*, 117, D06115, doi:10.1029/2011JD017283, 2012.
- Rypdal, M., and Rypdal, K.: Is there long-range memory in solar activity on time scales shorter than the sunspot period?, *J. Geophys. Res.*, 117, A04103, doi:10.1029/2011JA017283, 2012.
- Rypdal, K., Østvand, L., and Rypdal, M.: Long-range memory in Earth's surface temperature on time scales from months to centuries, *J. Geophys. Res.*, 118, doi:10.1002/jgrd.50399, 2013.
- Rypdal, M., and Rypdal, K.: Long-memory effects in linear-response models of Earth's temperature and implications for future global warming, submitted to *J. Climate*, <http://arxiv.org/pdf/1305.5080v1.pdf>.
- Scafetta, N.: A shared frequency set between the historical mid-latitude aurora records and the global surface temperature, *Journal of Atmospheric and Solar-Terrestrial Physics*, 74, 145-163, doi:10.1016/j.jastp.2011.10.013, 2011.
- Scafetta, N.: Testing an astronomically based decadal-scale empirical harmonic climate model versus the IPCC (2007) general circulation models, *Journal of Atmospheric and Solar-Terrestrial Physics*, 80, 124-137, doi:10.1016/j.jastp.2011.12.005, 2012.
- Schlesinger, M. E. and Ramankutty, N.: An oscillation in the global climate system of period 65-70 years, *Nature*, 367, 723-726, 1994.
- Vallis, G. K.: *Climate and the Oceans*, Princeton Primers in Climate, Princeton University Press, Princeton, 2012.
- von Storch, H. and Zwiers, F. W.: *Statistical Analysis in Climate Research*, Cambridge University Press, 1999.
- Zorita, E., Stocker, T. F., and von Storch, H.: How unusual is the recent series of warm years, *Geophys. Res. Lett.*, 35, L24706, doi:10.1002/2008GL036228, 2008.
- Vyushin, D. I., Kushner, P. J., and Zwiers, F.: Modeling and understanding persistence of climate variability, *J. Geophys. Res.*, 117, D21106, doi:10.1029/2012JD018240, 2012.