Power-Law and Long-Memory Characteristics of the Atmospheric General Circulation

DMITRY I. VYUSHIN AND PAUL J. KUSHNER

Department of Physics, University of Toronto, Toronto, Ontario, Canada

ABSTRACT

The question of which statistical model best describes internal climate variability on interannual and longer time scales is essential to the ability to predict such variables and detect periodicities and trends in them. For over 30 yr the dominant model for background climate variability has been the autoregressive model of the first order (AR1). However, recent research has shown that some aspects of climate variability are best described by a ''long memory'' or ''power-law'' model. Such a model fits a temporal spectrum to a single power-law function, which thereby accumulates more power at lower frequencies than an AR1 fit. In this study, several power-law model estimators are applied to global temperature data from reanalysis products. The methods employed (the detrended fluctuation analysis, Geweke–Porter-Hudak estimator, Gaussian semiparametric estimator, and multitapered versions of the last two) agree well for pure power-law stochastic processes. However, for the observed temperature record, the power-law fits are sensitive to the choice of frequency range and the intrinsic filtering properties of the methods. The observational results converge once frequency ranges are made consistent and the lowest frequencies are included, and once several climate signals have been filtered. Two robust results emerge from the analysis: first, that the tropical circulation features relatively large power-law exponents that connect to the zonal-mean extratropical circulation; and second, that the subtropical lower stratosphere exhibits power-law behavior that is volcanically forced.

1. Introduction

In climate research, the principal statistical model for representing internal variability in the absence of external forcings is Hasselmann's (1976) model of a slow damped component that represents the ocean coupled to a fast, noisy component that represents the atmosphere. In time series analysis, Hasselmann's model is represented by the autoregressive model of the first order (AR1), which has a two-parameter power spectrum of the form $S(\lambda) = \sigma_\epsilon^2/[1 - 2\phi \cos(2\pi\lambda) + \phi^2]$, where σ_ϵ^2 and ϕ are constants and λ is the frequency, with $|\lambda| \leq 1/2$. In typical applications σ_ϵ^2 and ϕ are estimated from time series or from power spectra and are used to test for the presence of significant periodic or externally forced signals (e.g., Ghil et al. 2002). In other applications (e.g., Bretherton and Battisti 2000), the model is taken as a simplified physical model to analyze climate variability.

The AR1 model remains the default model for background climate variability, but over the past 15 yr evidence has accumulated that it is not the best model in all cases. For low frequencies the AR1 spectrum saturates to a value of $S(\lambda \to 0) = \sigma_\epsilon^2/(1 - \phi)^2$. However, for many climate variables there is an absence of power saturation in the low-frequency limit and evidence of temporal scaling behavior, suggesting that a better fit might be a two-parameter power spectrum model of the form $S(\lambda) = b|\lambda|^{1-2H}$, where b and H are positive constants. The constant H is called the Hurst exponent, and typically $1/2 \leq H < 1$, with $H = 1/2$ corresponding to a white-noise or short-memory spectrum and with $H \to 1$ corresponding to a ''1/f'' noise spectrum. In the time domain, the AR1 autocorrelation function $\gamma(t) \sim e^{-t/\tau}$ decays exponentially, while the power-law autocorrelation function $\gamma(t) \sim t^{2H-2}$ decays algebraically, which suggests a distinctive long-memory behavior.

Power-law behavior has been observed in globally and hemispherically averaged surface air temperature (Bloomfield 1992; Gil-Alana 2005), station surface air temperature (Pelletier 1997), geopotential height at 500 hPa (Tsonis et al. 1999), temperature paleoclimate proxies (Pelletier 1997; Huybers and Curry 2006), and

Corresponding author address: Dmitry I. Vyushin, 60 St. George St., Toronto, ON M5S 1A7, Canada.
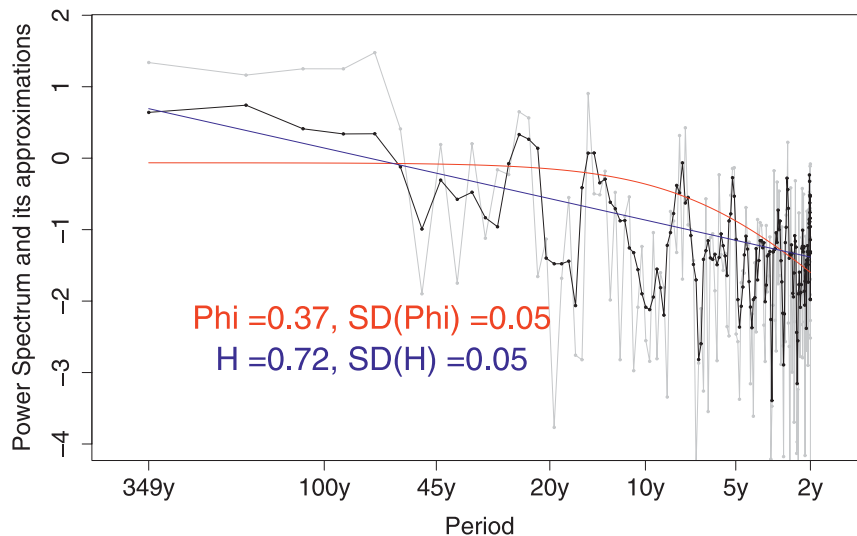E-mail: dmitry.vyushin@utoronto.ca

FIG. 1. CET (1659–2007) power spectrum estimators and their approximations. The raw periodogram power spectrum estimator (gray curve) and a three-sine-tapers multitaper power spectrum estimator (black curve, section 2a). A power spectrum of the AR(1) model fitted, using a maximum likelihood algorithm, to the CET time series (red solid curve). The power-law fit, estimated using the GPHE method (section 2a) to the multitaper is shown (blue line). Also included are numerical estimates for the AR1 parameter ϕ and its standard deviation sd(ϕ) (see Percival and Walden 1993, section 9.8) and the Hurst exponent estimate $\hat{H}$ and its standard deviation sd ($\hat{H}$) (see McCoy et al. 1998). The spectral and power-law estimators are described in section 2.

many other studies. However, only a few of these studies have performed quantitative tests to determine whether the power-law model is superior to the AR1 model. Those who have (Stephenson et al. 2000; Percival et al. 2001) find that both models demonstrate similar scores for the goodness-of-fit tests employed. Both Stephenson et al. (2000) and Percival et al. (2001) conclude that century-long time series are not long enough to clearly demonstrate the superiority of one model over the other.

To set the stage for our analysis of the modern temperature record, we apply the spectral density goodness-of-fit test (Beran 1992; Percival et al. 2001) to several of the temperature paleoclimate proxies used in Huybers and Curry (2006). The spectral density test evaluates the probability p to get a deviation from the null hypothesis shape of the spectrum at least as extreme as the one that was actually observed. We find that for the central U.K. temperature (CET) time series (1659–2007 annual means), spectrum $p = 0.67$ for the power-law model whereas $p = 0.2$ for the AR1 model. The large increase in p between the AR1 model and the power-law model indicates a closer fit for the latter.

In Fig. 1 we plot the CET periodogram (gray noisy curve) and the multitaper spectrum estimator (black curve) together with the AR1 and power-law fits to the spectrum. The methods are specified in the caption and

the power-law fit and multitaper methods will be described in section 2. Figure 1 demonstrates that the AR1 spectrum overestimates the CET power spectrum in high frequencies and underestimates it in low frequencies, whereas the power law, which also depends only on two parameters, similar to the AR1, does a much better job. When we repeat the analysis for the more recent record (corresponding to the reanalysis period of 09.1957–08.2002 analyzed in section 3), we find a similar slope that falls within the confidence intervals in Fig. 1, but the spectral density goodness-of-fit test no longer distinguishes between the power-law fit and the AR1 fit. Given this background, we now make the working assumption that the atmospheric general circulation might be well characterized by power-law behavior on interannual and longer time scales.

This study represents an attempt to systematically characterize power-law behavior in the atmospheric general circulation. However, first we must characterize power-law estimation methods, which, in this climate application, are surprisingly nonrobust. In section 2, we first describe and then use Monte Carlo benchmarking to compare a suite of power-law estimation methods. We then apply the methods to air temperature from reanalysis products for the last half-century, focusing on plots of zonal-mean cross sections of the Hurst exponent H (section 3). We attribute robust features of these

TABLE 1. The Hurst exponent estimation methods considered in this study. HF: high frequency and LF: low frequency (m denotes months, y denotes years).

Method	HF cutoff	LF cutoff	Remark
DFA(t)	$s_{short} = 18m$	$s_{long} = 11y$	Kantelhardt et al. (2001)
DFA(a)	$s_{short} = 18m$	$s_{long} = 45y$	Herein
GPHE(t)	$\lambda_{high} = 1/18m$	$\lambda_{low} = 1/15y$ ($l = 2$)	Robinson (1995b)
GPHE(a)	$\lambda_{high} = 1/18m$	$\lambda_{low} = 1/45y$ ($l = 0$)	Hurvich et al. (1998)
MTM GPHE(t)	$\lambda_{high} = 1/18m$	$\lambda_{low} = 1/15y$ ($l = 2$)	McCoy et al. (1998)
MTM GPHE(a)	$\lambda_{high} = 1/18m$	$\lambda_{low} = 1/45y$ ($l = 0$)	Herein
GSPE(t)	$\lambda_{high} = 1/18m$	$\lambda_{low} = 1/15y$ ($l = 2$)	Herein
GSPE(a)	$\lambda_{high} = 1/18m$	$\lambda_{low} = 1/45y$ ($l = 0$)	Robinson (1995a)
MTM GSPE(t)	$\lambda_{high} = 1/18m$	$\lambda_{low} = 1/15y$ ($l = 2$)	Herein
MTM GSPE(a)	$\lambda_{high} = 1/18m$	$\lambda_{low} = 1/45y$ ($l = 0$)	Herein

plots to specific physical climate processes and other less robust features to methodological artifacts. Although no single power-law fit technique should be employed in isolation, we identify a pair of techniques that characterize the range of results that might be typically expected. In section 3, we also discuss the connection between point statistics and zonal-mean statistics as the starting point for a more complete physical theory of power-law behavior in the general circulation.

2. Description and tests of power-law estimators

Many methods for estimating the Hurst exponent H are documented in the literature and a significant challenge in our analysis has been to reconcile the nonrobust aspects of these methods. In this section, we describe several of the documented methods, develop some variants of our own, and characterize them using Monte Carlo benchmarking. In the next section, we will apply the methods to temperature data from reanalysis products. The methods are summarized in Table 1. They include time domain methods, and periodogram and multitaper (spectral domain) methods. The Monte Carlo benchmarking will show that all of the methods agree reasonably well for simulated pure power-law stochastic processes. However, when we apply the methods to observed data in section 3, nonrobust results that are associated with frequency range choices and filtering properties of the methods will emerge.

a. Spectral methods

The spectral methods find H by estimating the spectral slope. These methods first calculate an estimate $\hat{S}(\lambda)$ from a finite-length time series of the true spectrum $S(\lambda)$, and then find the best power-law fit to $\hat{S}(\lambda)$. We consider two choices of spectral estimators $\hat{S}(\lambda)$: the periodogram estimator (corresponding to the raw discrete spectrum) and the multitaper estimator (Percival and Walden 1993). For a time series $X(t)$, $t = 1, \ldots, N$,

the periodogram estimator is simply the square amplitude of the discrete Fourier transform divided by the time series length

$$\hat{S}^{(p)}(\lambda_j) = \frac{1}{N}\left|\sum_{t=1}^{N} X(t)e^{-i2\pi t\lambda_j}\right|^2, \quad j = 1, \ldots, [N/2], \quad (1)$$

where $\lambda_j = j/N$ and the square brackets in this case denote rounding toward zero. The periodogram is an asymptotically unbiased but inconsistent spectrum estimator, because its variance is not a decreasing function of N; periodograms, as illustrated by the gray curve in Fig. 1, tend to appear noisy in spectral plots.

Multitaper spectral estimation (Thomson 1982) provides an estimated spectrum with relatively reduced variance compared to the periodogram. It employs a set of K orthogonal "tapers" $h_k(t)$, $k = 1, \ldots, K$ that is applied to the time series $X(t)$. The multitaper spectral estimate is given by

$$\hat{S}^{(mt)}(\lambda_j) = \frac{1}{K}\sum_{k=1}^{K} S_k^{(d)}(\lambda_j), \quad j = 1, \ldots, [N/2], \quad (2)$$

where

$$\hat{S}_k^{(d)}(\lambda_j) = \left|\sum_{t=1}^{N} h_k(t)X(t)e^{-i2\pi t\lambda_j}\right|^2, \quad j = 1, \ldots, [N/2], \quad (3)$$

is the kth direct spectral estimator. In this study we use sine tapers (Riedel and Sidorenko 1995)

$$h_k(t) = \sqrt{\frac{2}{N+1}}\sin\left(\frac{k\pi t}{N+1}\right), \quad t = 1, \ldots, N. \quad (4)$$

The number of tapers K usually used in geophysical applications ranges between 3 and 5 (e.g., Ghil et al. 2002; Huybers and Curry 2006). We choose $K = 3$ because of the large number of time series analyzed.

It can be shown that the variance of $\hat{S}^{(mt)}$ is a factor K smaller than the variance of $\hat{S}^{(p)}$ for large N. Thus, multitaper spectra appear smoother in spectral plots; the smoothing effect is evident in the multitaper spectral estimate shown by the black curve in Fig. 1.

Given the spectral density estimator $\hat{S}(\lambda)$, we find a power-law fit to $\hat{S}(\lambda)$ of the form $f(\lambda; b, H) = b|\lambda|^{1-2H}$ over a frequency range $\lambda_{low} \leq \lambda \leq \lambda_{high}$, where H is the Hurst exponent, b is a scaling factor, and λ_{low} and λ_{high} are low and high cutoff frequencies. For a recent review of these methods, known as spectral semiparametric estimation methods, see Moulines and Soulier (2002). We estimate b and H by minimizing

$$K(b, H) = \frac{1}{m - l} \sum_{j=l+1}^{m} k[\hat{S}(\lambda_j), f(\lambda_j; b, H)], \quad (5)$$

where $k(u, v)$ is the so-called contrast function, which can be thought of as a distance between functions u and v. In the summation, l and m are indices related to the low- and high-frequency cutoffs, $\lambda_{low} = \lambda_{l+1}$ and $\lambda_{high} = \lambda_m$.

The semiparametric power-law fits differ in their choice of contrast function $k(u, v)$. We here consider the Geweke–Porter-Hudak estimator (GPHE; Geweke and Porter-Hudak 1983), with $k(u, v) = [\log(u) - \log(v)]^2$, which corresponds to log-linear regression, and the Gaussian semiparametric estimator (GSPE; Fox and Taqqu 1986), with $k(u, v) = \log(u) + u/v$, which corresponds to a maximum likelihood estimator. GPHE is the best known and simplest of the two methods; the optimal b and H can be found in closed form along with confidence intervals. We use GPHE to obtain the CET power-law fit for the multitaper spectrum estimator in Fig. 1; the confidence intervals for GPHE with multitapering are found in McCoy et al. (1998).

GSPE is relatively more sophisticated and Robinson (1995a,b) has shown it to be superior to GPHE in various ways. Its optimization is not in closed form, but it reduces to a standard one-dimensional numerical optimization procedure. Robinson (1995b) and Hurvich et al. (1998) show that GSPE has a factor of $\pi^2/6 \approx 1.7$ smaller asymptotic variance than GPHE $\{1/[4(m - l)]$ versus $\pi^2/[24(m - l)]$ for large $N\}$. This property leads to practical advantages: in an analysis of power-law behavior in stratospheric ozone, Vyushin et al. (2007) found that GSPE yielded H estimates that were spatially smoother and more robust than GPHE.

b. Time domain method: Detrended fluctuation analysis

The detrended fluctuation analysis (DFA; Peng et al. 1993; Kantelhardt et al. 2001) time domain estimator of H is, along with GPHE, the best known power-law fitting technique and has been applied widely in the life sciences, the earth sciences, and physics. DFA works as follows. First, a cumulative sum time series is generated from the original time series $X(t)$. The cumulative sum time series is then split into segments of size s. Each of these segments is approximated by a least squares fit to a polynomial of order P, with P typically chosen to be between 1 and 5. The standard deviation of the residual of each least squares fit is then calculated for each segment and then averaged over all of the segments. This quantity is denoted $F(s)$ and is calculated for segment sizes s, where $P + 2 \leq s \leq s_{max}$. The standard method is to take $s_{max} = [N/4]$ (Kantelhardt et al. 2001), but we will test a variant with $s_{max} = N$. The so-called "fluctuation function" $F(s)$ characterizes the noise at each time scale s; if the spectral density $S(\lambda) \sim \lambda^{1-2H}$ for small λ, the fluctuation function $F(s) \sim s^H$ for large s (Taqqu et al. 1995; Heneghan and McDarby 2000).

Given this, we determine H by least squares linear regression of log F against log s in the range $s_{short} \leq s \leq s_{long}$ (Peng et al. 1993; Kantelhardt et al. 2001), where s_{short} and s_{long} are short and long time-scale cutoffs that correspond to the high- and low-frequency cutoffs for the spectral methods. For DFA3 (i.e., DFA with $P = 3$) we use a lower (high frequency) cutoff scale of $s_{short} = 18$ months, because it is only for longer time scales $F(s)$ for DFA3 might be well represented by a power-law function (Kantelhardt et al. 2001). The choice of s_{long} will be discussed in section 2c.

DFA is relatively straightforward to implement and can be used to infer information about the order of a trend of the time series. For example, a quadratic trend is effectively filtered out by DFA with $P > 2$. However, unlike the GPHE and GSPE spectral methods DFA is numerically based and thus lacks rigorous expressions for bias and confidence interval estimates.

c. Benchmark tests of the estimator methods

Before applying the power-law fit methods to reanalysis climate data we benchmark the methods using Monte Carlo tests of time series with known power-law behavior. These time series are generated from autoregressive fractionally integrated moving average (ARFIMA) models, which are linear models for power-law stochastic processes (Beran 1994; Taqqu 2002). By convention, an ARFIMA(0, d, 0) time series has a Hurst exponent $H = d + 0.5$ [i.e., the power spectrum $S(\lambda) \sim \lambda^{-2d}$]. To mimic our climate data analysis, we take the ARFIMA time series to represent monthly-mean records and estimate the Hurst exponent for frequencies lower than $\lambda_{high} = 1/(18$ months). We choose $\lambda_{high} = 18$ months for consistency with DFA3.
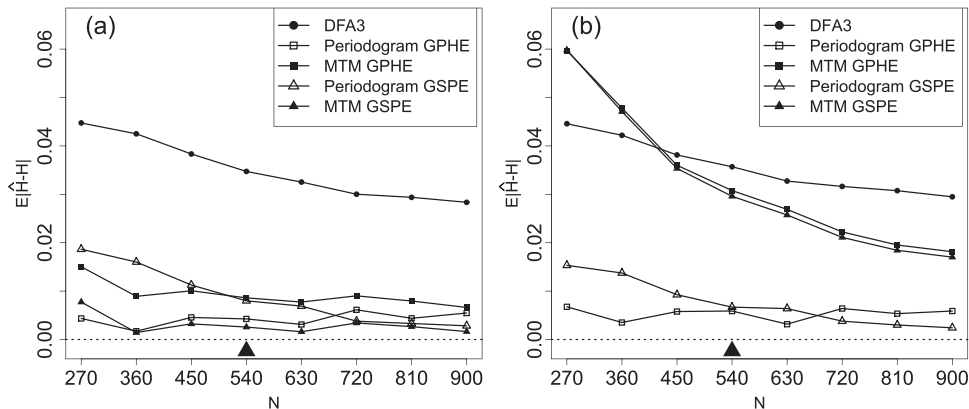
FIG. 2. Absolute bias of the Hurst exponent estimators as a function of time series length N. Synthetic time series were simulated by ARFIMA(0, 0.3, 0), that is, the true Hurst exponent was set to 0.8. Time series length $N = 540$ (solid triangle) corresponds to the length of the monthly-mean ERA-40 data. The results for (a) the trimmed "(t)" version of the methods, and (b) the all-frequency "(a)" version of the methods. See Table 1 and section 2 for a description of the methods.

As mentioned earlier, we have found that inconsistent choices of frequency range lead to nonrobust estimates of H in climate data. This is a practical issue encountered when dealing with time series that are not pure power-law stochastic processes with uniform behavior across all time scales. Many of the standard applications of power-law estimates build in these inconsistent ranges, for various reasons. For example, standard practice for DFA is to use $s_{\text{long}} = [N/4]$, for periodogram spectral methods to use $\lambda_{\text{low}} = 1/N$, and for the multi-tapered methods to use $\lambda_{\text{low}} \approx K/N$ (see Table 1 for references to each of these "conventional" methods). To test for the effect of these choices, we benchmark "all frequency" [denoted "(a)"] and "trimmed" [denoted "(t)"] versions of the methods. The all-frequency methods set the low-frequency (long time scale) cutoff as low as possible. The trimmed methods cut off some of the lowest frequencies. Table 1 lists two versions of the methods we use, and in connection with the table we note the following:

- The multitapered methods conventionally trim the lowest frequencies and the periodogram methods conventionally use all frequencies. We here test trimmed and all-frequency versions of all of the spectral methods.
- For DFA3, DFA3(t) with $s_{\text{max}} = s_{\text{long}} = [N/4]$ is the conventional method (Kantelhardt et al. 2001). We here test a version DFA3(a), whose time-scale range is consistent with the all-frequency spectral methods. DFA3(a) uses $s_{\text{max}} = s_{\text{long}} = N$.

We first test for the convergence of the magnitude of the estimators' bias $\langle |\hat{H} - H| \rangle$ as a function of time series length N, where $\hat{H}$ is the estimated value of the

ARFIMA time series. Figures 2a,b plot $\langle |\hat{H} - H| \rangle$ for the trimmed and all-frequency versions of the following five methods: DFA3, GPHE, multitapered (MTM) GPHE, GSPE, and MTM GSPE. Here the angled brackets represent the ensemble mean over 10 000 realizations of the ARFIMA model, for $H = 0.8$ and $270 \leq N \leq 900$. We see that the DFA3(t) estimator converges most slowly among the trimmed estimators, the periodogram spectral methods converge most quickly among the all-frequency estimators, and neither DFA3 nor the periodogram methods are sensitive to the trimming. The convergence rate of the multitaper spectral methods falls between the periodogram and DFA3 methods for the all-frequency estimators. The increase in bias from tapering for the all-frequency cases is expected from general statistical principles: heavier smoothing, that is, reduction of the variance from tapering, leads to an increase in the bias (von Storch and Zwiers 1999). The effect of including the additional low-frequency points degrades the convergence of the multitaper methods (Fig. 2b). This degradation is not surprising because the tapering impacts low frequencies most strongly. The impact of trimming on the spectral methods is consistent with theory (Hurvich et al. 1998; McCoy et al. 1998), but to our knowledge the robustness of DFA3 to changing from $s_{\text{long}} = [N/4]$ to $s_{\text{long}} = N$ has not been reported before.

Next, we test how the estimators' bias depend on H for $N = 540$, which corresponds to the length of the monthly reanalysis time series analyzed in section 3. In Figs. 3a,b we plot the bias $\langle \hat{H} - H \rangle$ over 10 000 realizations. All of the methods provide accurate estimates of H within the estimators' standard deviation (see Fig. 4). DFA3 exhibits the largest bias among the
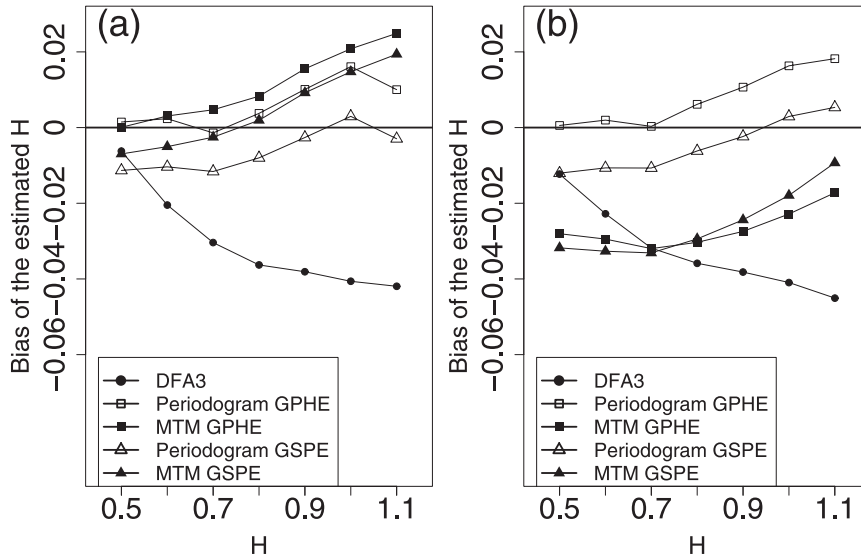
FIG. 3. Bias of the Hurst exponent estimators as a function of the true Hurst exponent. Time series length was fixed to 540, which corresponds to the length of the monthly-mean ERA-40 data. The rest of the description is similar to Fig. 2.

trimmed estimators. This bias increases in magnitude with increasing H and is robust to the trimming. The periodogram methods have the smallest bias, which is also robust to the trimming. The bias of the trimmed multitaper methods is comparable to that of the periodogram methods (Fig. 3a). However, the bias increases for the all-frequency multitaper methods, indicating again

that the multitaper H estimate degrades when all frequencies are included (Fig. 3b).

Figure 4 shows the relative robustness of the different methods as measured by the estimator standard deviation $\sqrt{\langle (\hat{H} - \langle \hat{H} \rangle)^2 \rangle}$ averaged across values of H from 0.5 to 1.1. In this figure we also include the large N asymptotic estimates for periodogram GPHE and GSPE;
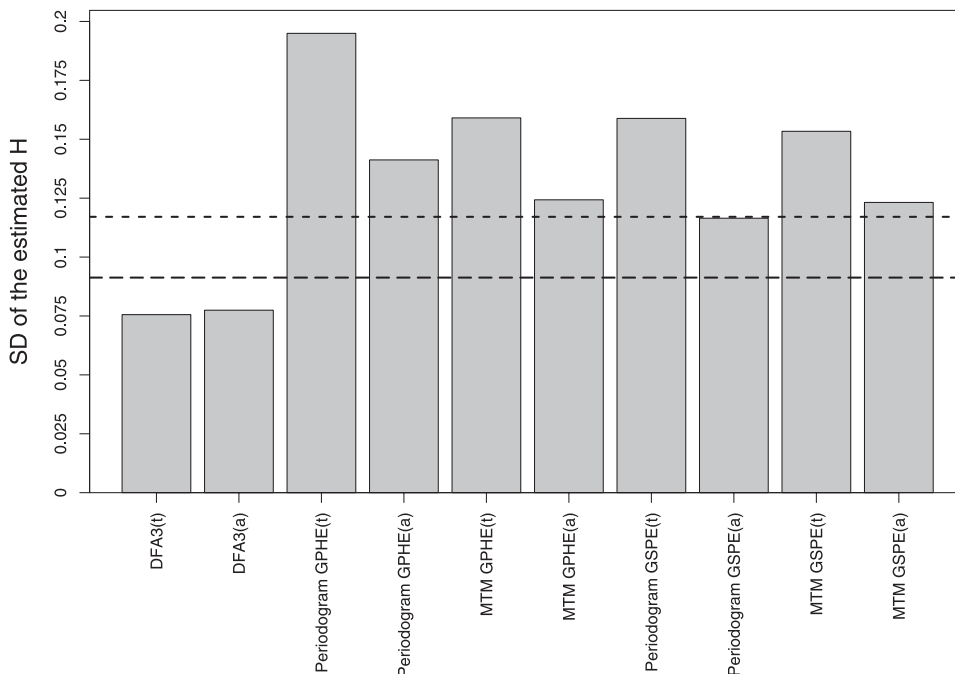

FIG. 4. Standard deviation of $\hat{H}$ averaged for each method in Table 1 over the H values shown in Fig. 3.
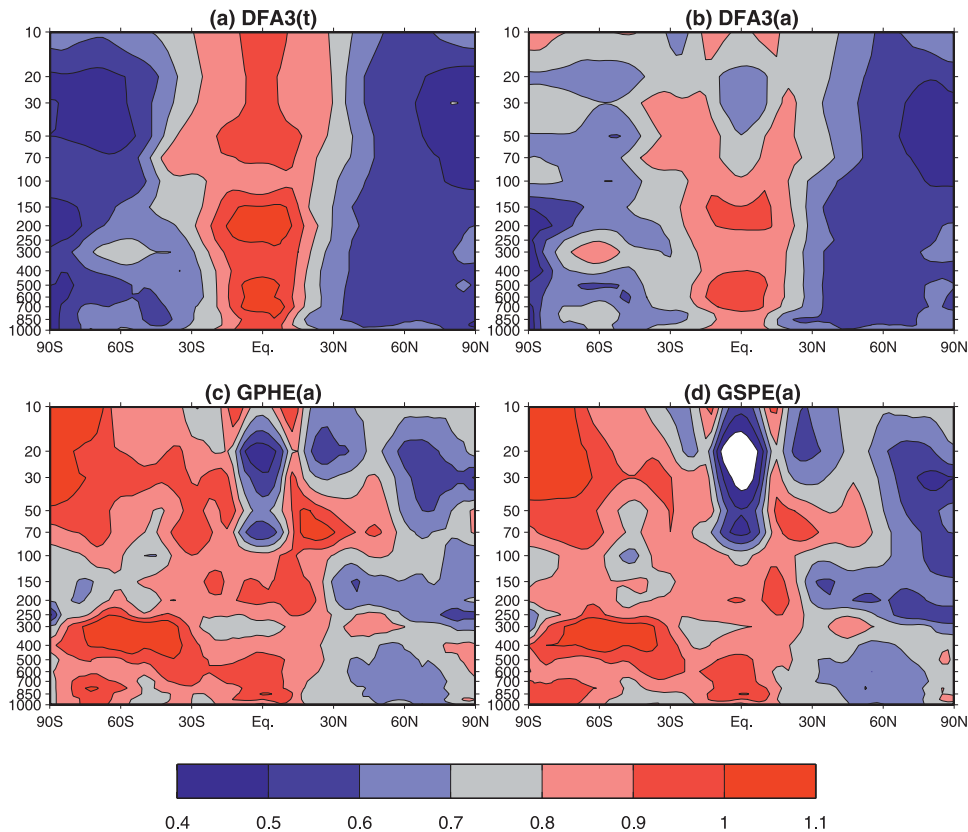
FIG. 5. Zonal-mean $\hat{H}$ of ERA-40 air temperature for (a) DFA3(t), (b) DFA3(a), (c) GPHE(a), and (d) GSPE(a). Values below 0.4 are shown in white.

these estimates are independent of H. DFA3 exhibits the least spread and is not sensitive to the trimming. However, all of the spectral methods have smaller variance when all frequencies are included. Among the spectral methods the periodogram GSPE(a) exhibits the least spread. The decrease of the $\hat{H}$ variance gained from multitapering is outweighed by the increased variance from the necessary trimming. The experimental standard deviations are consistently greater than the asymptotic ones for the time series length of 540, but the asymptotic results provide useful constraints in many applications.

Thus, all of the methods provide valid approaches to power-law fitting of pure power-law stochastic processes, but each method has the following distinct characteristics:

- DFA3 is robust (has the smallest variance), but has relatively large biases.
- GSPE is more complicated than GPHE, but produces more robust estimates than GPHE and less biased estimates than DFA3.
- The standard trimmed MTM [MTM GPHE(t), MTM GSPE(t)] are less robust than the correspond-

ing standard all-frequency periodogram methods [GPHE(a), GSPE(a)].

3. Power-law fits for reanalyzed air temperature data

a. Results for unfiltered data

Having benchmarked the H estimation methods we now apply them to the monthly-mean 40-yr European Centre for Medium-Range Weather Forecasts (ECMWF) Re-Analysis (ERA-40) air temperature for 09.1957–08. 2002. The annual cycle and three of its harmonics are removed from the temperature. The estimates of the Hurst exponent $\hat{H}$ are then carried out identically to the benchmark tests in section 2d. We calculate $\hat{H}$ at each longitude, latitude, and pressure, and take the zonal mean of the result to obtain a zonal cross section that characterizes the power-law behavior of the global atmosphere. We plot the resulting zonal-mean $\hat{H}$ in Figs. 5a–d for DFA3(t), DFA3(a), periodogram GPHE(a), and periodogram GSPE(a) (see Table 1 for parameter settings for these methods). By including these various methods
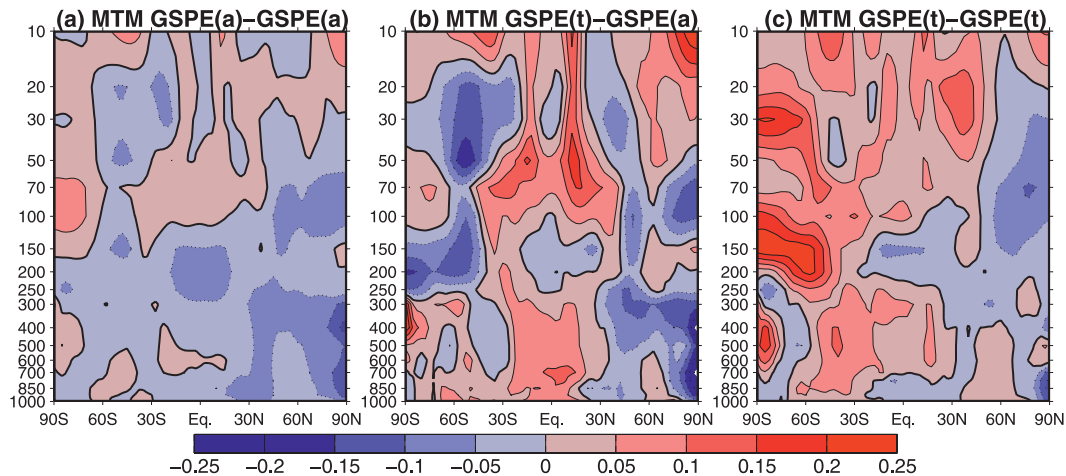
FIG. 6. (a) Zonal mean of $\hat{H}$ for MTM GSPE(a) minus that for GSPE(a). (b) Zonal mean of $\hat{H}$ for MTM GSPE(t) minus that for GSPE(a). (c) Zonal mean of $\hat{H}$ for MTM GSPE(t) minus that for GSPE(t). All $\hat{H}$ were estimated for linearly detrended (LTR filtered, in the notation of section 3c) ERA-40 air temperature.

and being clear about parameter settings we aim to reconcile existing results for the value of H for air temperature, because different studies use different methods and different frequency ranges (see, e.g., Pelletier 1997; Eichner et al. 2003; Fraedrich and Blender 2003; Huybers and Curry 2006).

Figure 5 shows that, apart from a common maximum H in the tropical troposphere, there are clear contrasts between DFA3(t) and DFA3(a), and also between DFA3 and the two spectral methods. The differences generally lie well outside the range of biases found in the benchmark tests, and one of our main aims is to pin down the source of these differences. DFA3(t) and DFA3(a) show a similar decrease of H from the tropics to the extratropics, but the values of H are generally lower in DFA3(a) than in DFA3(t). This is particularly true in the tropical lower stratosphere, where the methods disagree most strongly (section 3c). The spectral methods produce noisier plots, as might be expected by their generally larger variance (section 2c). They display pronounced maxima in the Southern Hemisphere that are not found in the DFA3 plots, and we will show that they are largely tied to linear trends in the data, some of which might arise from data inhomogeneities (section 3c). They also show strong minima and even blue-noise (positive spectral slope) behavior in the tropical stratosphere; these features are also discussed below.

b. Effect of multitapering and frequency range

We reported in section 2 that multitaper-based GPHE and GSPE have somewhat similar biases and standard deviations to the periodogram-based GPHE and GSPE. The difference between the MTM GSPE(a) and the periodogram GSPE(a) for linearly detrended data is shown in Fig. 6a. When all of the frequencies are included, the differences between the methods range from about -0.15 to about 0.1. The asymptotic limit of the GSPE Hurst exponent estimate standard deviation (one sigma) for this frequency range is $1/2\sqrt{30} \approx 0.1$, where 30 is the number of frequencies used. Therefore, the differences for the first case are statistically significant only for a few locations and the MTM GSPE plot of $\hat{H}$ is visually similar to Fig. 5d (not shown). However, we recall that MTM GSPE(a) has a larger bias than MTM GSPE(t) (section 2c). Thus, standard practice would suggest that we should compare MTM GSPE(t) and GSPE(a). Figure 6b plots the difference of $\hat{H}$ for MTM GSPE(t) and GSPE(a). The total area where the two methods disagree has increased relative to Fig. 6a. The difference between MTM GSPE(t) and GSPE(t) is even larger (see Fig. 6c), especially in the observation-sparse regions, such as the Southern Hemisphere stratosphere, which points to the different sensitivities to data inhomogeneities (section 3c) in periodogram- and MTM-based methods. Given this, it is clear that differences introduced by standard multitapering methods reflect not only the effect of multitapering itself but also (a) the selection of the frequency range for the power-law fit, (b) the effect of data inhomogeneities, and (c) the increased variance of H resulting from trimming.

At this point we are in a position to reduce the number of methods we consider. Together with the results of the Monte Carlo testing and additional testing with the GSPE method, we conclude that multitapering adds an unnecessary complication to the Hurst exponent spectral estimation procedure. Multitapering might produce graphically smoother spectral plots (as in Fig. 1), but it

does not provide obviously improved H estimates. We have also found that the all-frequency estimates work equally well for both DFA3 and the periodogram spectral methods. Finally, we find the GSPE method to be similar to, but moderately more robust than, the GPHE method. We thus proceed to focus mainly on the DFA3(a) and GSPE(a) methods, and we try to explain the robust and nonrobust aspects of their $\hat{H}$ portraits.

c. Effects of filtering and choice of reanalysis product

We now show that many of the differences between the spectral and time domain methods (e.g., between Figs. 5b and 5d) can be attributed to specific physical processes and methodological artifacts. We consider the effects of detrending, the quasi-biennial oscillation (QBO), ENSO, and volcanic aerosol forcing. The filters we use are as follows:

- LTR: We remove a simple linear trend from the data.
- QBO: We remove a QBO signal consisting of the equatorial zonally averaged zonal winds at 30 and 50 hPa (see information online at http://www.cpc.ncep.noaa.gov/data/indices/). We use winds at both 30 and 50 hPa, because they are about 90° out of phase, which allows for a better representation of the QBO signal.
- ENSO: We remove an ENSO signal consisting of the Niño-3.4 index lagged by 4 months. We choose 4-months lag because it maximizes correlations between the Niño-3.4 index and tropical troposphere air temperature (Yulaeva and Wallace 1994; Trenberth and Smith 2006).
- VOL: We remove the effect of volcanic aerosols by regressing air temperature on meridionally and time-dependent historical reconstructions of volcanic aerosols optical depth (Ammann et al. 2003).

All of the signals described above are modulated by the seasonal cycle in our filtering procedure (e.g., Vyushin et al. 2007). We also have carried out additional calculations involving solar and Atlantic multidecadal variability signals, but these did not show significant effects on H estimates. The impact of removing each signal on $\hat{H}$ is plotted in Fig. 7. For all rows in Fig. 7 the left column corresponds to DFA3(a) and the right column to GSPE(a).

The first row of Fig. 7 (Figs. 7a,b) shows the effect of the linear detrending (filtering LTR) on $\hat{H}$. Specifically, $\hat{H}$ with LTR filtering is subtracted from $\hat{H}$ with no filtering. As expected, detrending has little effect on the DFA3-based estimate, because this method effectively filters out polynomial trends up to the second order. However, the effect is significant for GSPE, especially in the Northern Hemisphere lower stratosphere and

Southern Hemisphere troposphere. The presence of the linear trend increases $\hat{H}$ by 0.1–0.25 for the spectral methods, because a linear trend increases power at low frequencies and therefore steepens the spectral slope. It is well known that climate trends in reanalysis products often reflect data inhomogeneities (Dell'Aquila et al. 2007; Bromwich and Fogt 2004; Marshall 2002; Randel and Wu 1999; Randel et al. 2000). We do not aim to evaluate the realism of these trends; instead, we want to point out the relative sensitivities of the H estimation methods to detrending.

The second row of Fig. 7 (Figs. 7c,d) shows the impact of removing the QBO. Specifically, $\hat{H}$ with LTR + QBO filtering was subtracted from $\hat{H}$ with LTR filtering. For DFA3(a) and GSPE(a), removing the QBO reduces $\hat{H}$ in the tropical and subtropical lower stratosphere, but the impact is much greater for GSPE(a) than for DFA3(a). Thus, the presence of a quasi-periodic signal appears to significantly impact the spectral method. In contrast to the linear trend case, the QBO boosts frequencies near the high-frequency cutoff of 18 months and shallows the spectral slope. Thus, the presence of the QBO reduces $\hat{H}$.

In another analysis, we have found that for the trimmed DFA, DFA3(t), which is the standard method in the literature, the effect of filtering the QBO was to significantly *increase* $\hat{H}$ in the lower stratosphere (not shown), opposite to what is seen in Figs. 7c,d. This effect can be attributed to the spreading of the QBO signal by the DFA smoothing (e.g., Jánosi and Müller 2005; Marković and Koch 2005). The effect is seen in the difference between the DFA3(t) and DFA3(a) plots in Figs. 5a,b. This again illustrates how sensitive the H estimation methods are to the frequency range choice.

Figures 7e,f show the impact of removing the ENSO signal. The location of the difference is in the tropical troposphere and the sense of the impact is similar to the QBO case. ENSO represents a high- (interannual) frequency signal that is significantly correlated with tropical temperatures, and so the ENSO and QBO effects on $\hat{H}$ are analogous. Again, the impact on $\hat{H}$ for DFA3(a) is minimal, but it is more significant for the standard-practice DFA3(t) (not shown).

Unlike for the other filtered signals, the impact of the volcanic signal on $\hat{H}$ is similar for both DFA3 and GSPE (Figs. 7g,h). Volcanic forcing appears to increase $\hat{H}$ in the tropical and subtropical lower stratosphere. In climate simulations with and without volcanic forcings, we have been able to reproduce this volcanic signature in $\hat{H}$ (Vyushin et al. 2009, manuscript submitted to *Geophys. Res. Lett.*), and Vyushin et al. (2004) have reported a similar boost of surface temperature $\hat{H}$ from volcanic forcing in climate of the twentieth-century simulations.
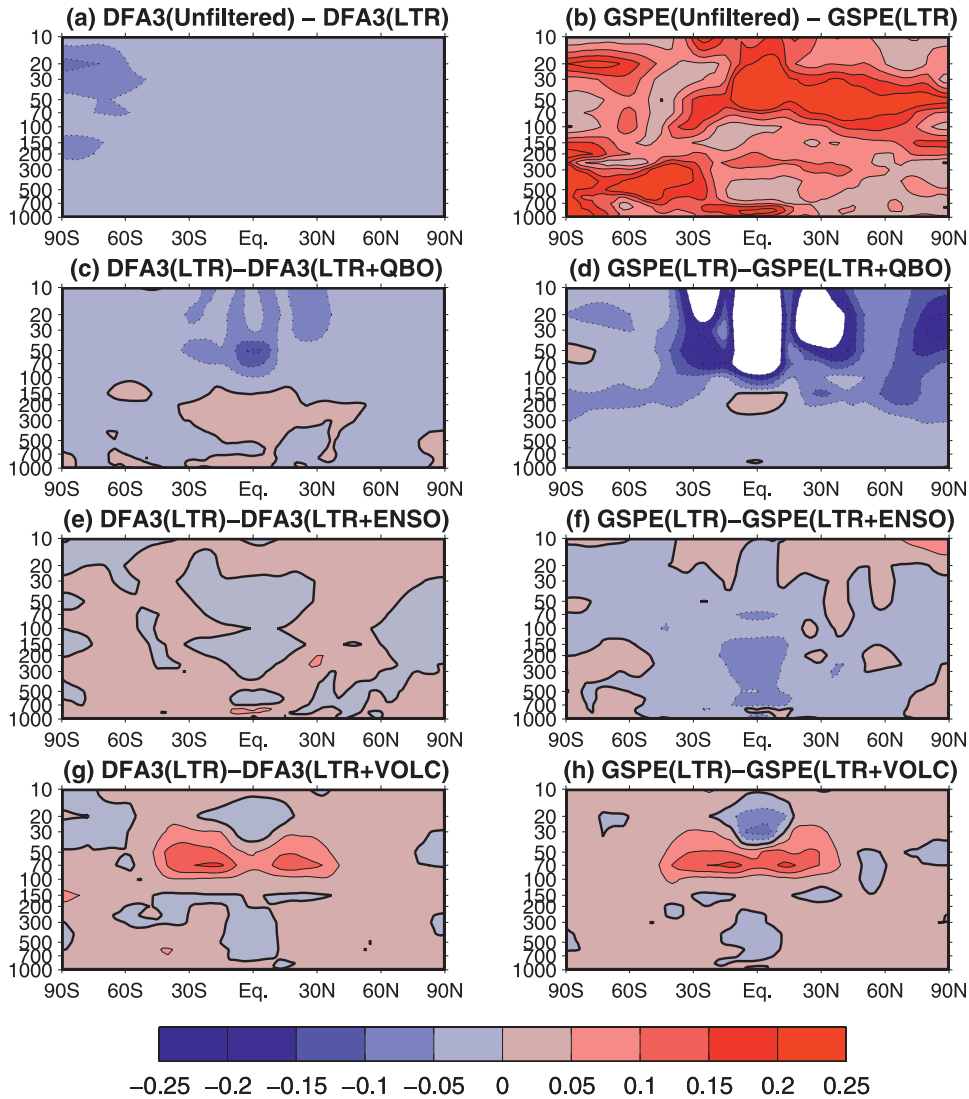
FIG. 7. Impact on zonal-mean $\hat{H}$ of filtering different climate signals. (left) Difference plots for DFA3(a) and (right) difference plots for GSPE(a). (a),(b) Zonal-mean $\hat{H}$ for the unfiltered time series minus that for $\hat{H}$ with LTR filtering. (c),(d) Zonal-mean $\hat{H}$ for LTR filtering minus that for $\hat{H}$ with LTR + QBO filtering. (e),(f) Same as in (c) and (d), but for ENSO instead of QBO filtering. (g),(h) Same as in (c) and (d), but for VOL instead of QBO filtering.

The fact that volcanic forcing leads to power-law behavior points to an ambiguity in how to interpret power-law spectra as indicators of long-memory processes. In this case the long-memory process is the geophysical one of volcanism, which leads to intermittent pulses of shortwave forcing, rather than a process internal to the atmospheric general circulation.

We have accounted for the various effects of trends, QBO, ENSO, and volcanoes, and we can now compare the spectral and time domain methods again. Figures 8a,b show $\hat{H}$ for DFA3(a) and GSPE(a) with all of the LTR, QBO, ENSO, and VOL signals removed. Outside of the

Southern Hemisphere stratosphere, the plots are similar, with the GSPE providing somewhat larger $\hat{H}$. An overall larger value of $\hat{H}$ is expected for GSPE based on the Monte Carlo testing (Fig. 3b). Compared to the corresponding plots in Figs. 5b,d, the methods have converged considerably. Both methods show relatively large $\hat{H}$ in the tropical troposphere, the subtropical lower stratosphere, the tropical stratosphere above 20 hPa, and the extratropical Southern Hemisphere.

The Southern Hemisphere stratosphere, where GSPE(a) and DFA3(a) continue to disagree in Figs. 8a,b, is a highly problematic area for this kind of analysis because
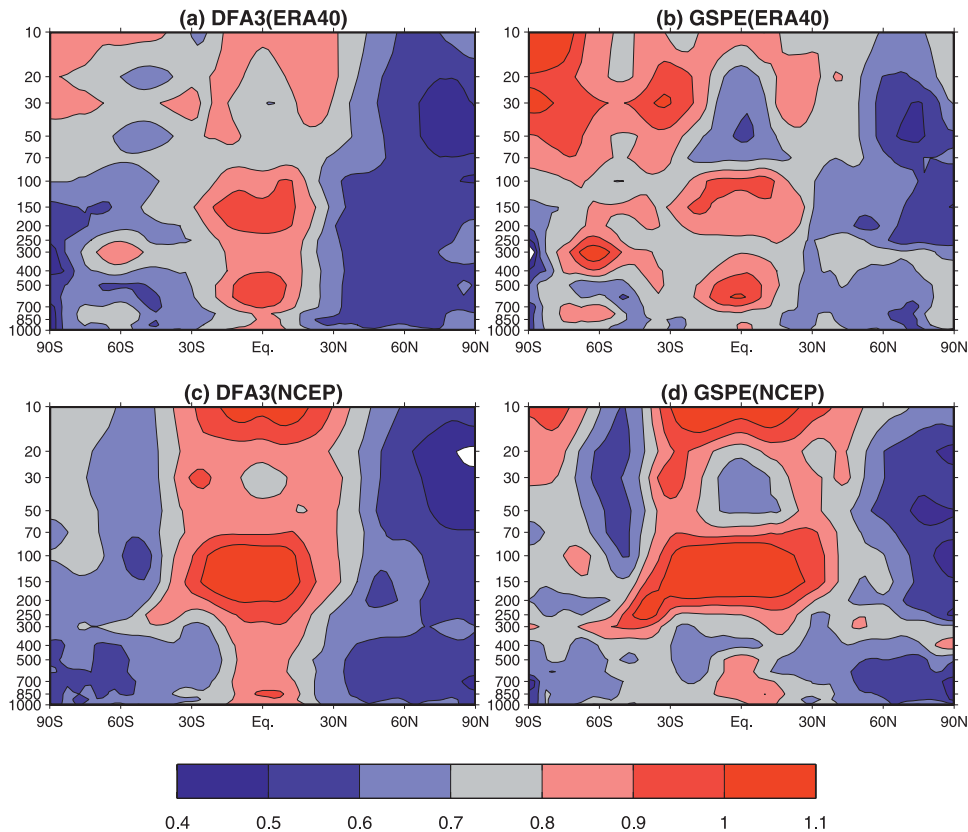
FIG. 8. Zonal-mean $\hat{H}$ with LTR + QBO + ENSO + VOL filtering for (a) DFA3(a) and ERA-40 data, (b) GSPE(a) and ERA-40 data, (c) DFA3(a) and NCEP–NCAR data, and (d) GSPE(a) and NCEP–NCAR data.

of inhomogeneities in the reanalyzed data (Randel and Wu 1999; Marshall 2002). For example, a red spot in $\hat{H}$ structure at 60°S, 300 hPa is caused by an obvious jump in temperature related to the assimilation of the Vertical Temperature Profile Radiometer data (e.g., Bromwich and Fogt 2004; Dell'Aquila et al. 2007). To test the robustness of the H estimates for ERA-40, we calculate $\hat{H}$ using the National Centers for Environmental Prediction–National Center for Atmospheric Research (NCEP–NCAR) reanalysis air temperature with the same filtering applied. Figures 8c,d show that the main features of the $\hat{H}$ portraits found in the ERA-40 data remain robust for the NCEP–NCAR data. However, in the data-poor Southern Hemisphere polar stratosphere, the four panels disagree significantly. It is known that the Southern Hemisphere stratosphere record has nonlinear temperature trends (Vyushin et al. 2007) related to photochemical ozone loss, and that these cannot have been filtered out by LTR in GSPE; however, this does not explain why GSPE(a) gives different results for between NCEP–NCAR and ERA-40. We thus do not expect to find a robust estimate of H in this region from reanalyzed data.

d. $\hat{H}$ of zonal-mean temperature

Figures 5–8 represent the zonal average of $\hat{H}$ values calculated at each point. However, energy and momentum conservation constraints, along with the theory of eddy mean flow interactions in the atmospheric general circulation (e.g., Lorenz 1967; Schneider 2006), suggest that $\hat{H}$ values of the zonal-mean circulation might also be dynamically interesting. With this very general motivation, we show $\hat{H}$ for the zonally averaged ERA-40 air temperature in Fig. 9, which, analogously to Fig. 8, has had all of the (LTR, QBO, ENSO, and VOL) signals removed. We include both DFA3(a) and GSPE(a) estimates in Figs. 9a,b, and plot the difference fields ($\hat{H}$ for the zonal-mean temperature minus the zonal mean of $\hat{H}$ for the temperature at each point) in Figs. 9c,d.

We see in Fig. 9 that the zonal-mean temperature statistics exhibit considerably more power-law behavior in the extratropics than the point temperature statistics. For example, the regions with $\hat{H} > 0.8$ are confined between 15°S and 10°N in Fig. 8a, but are between 40°S and 20°N in Fig. 9a. For GSPE this difference is even
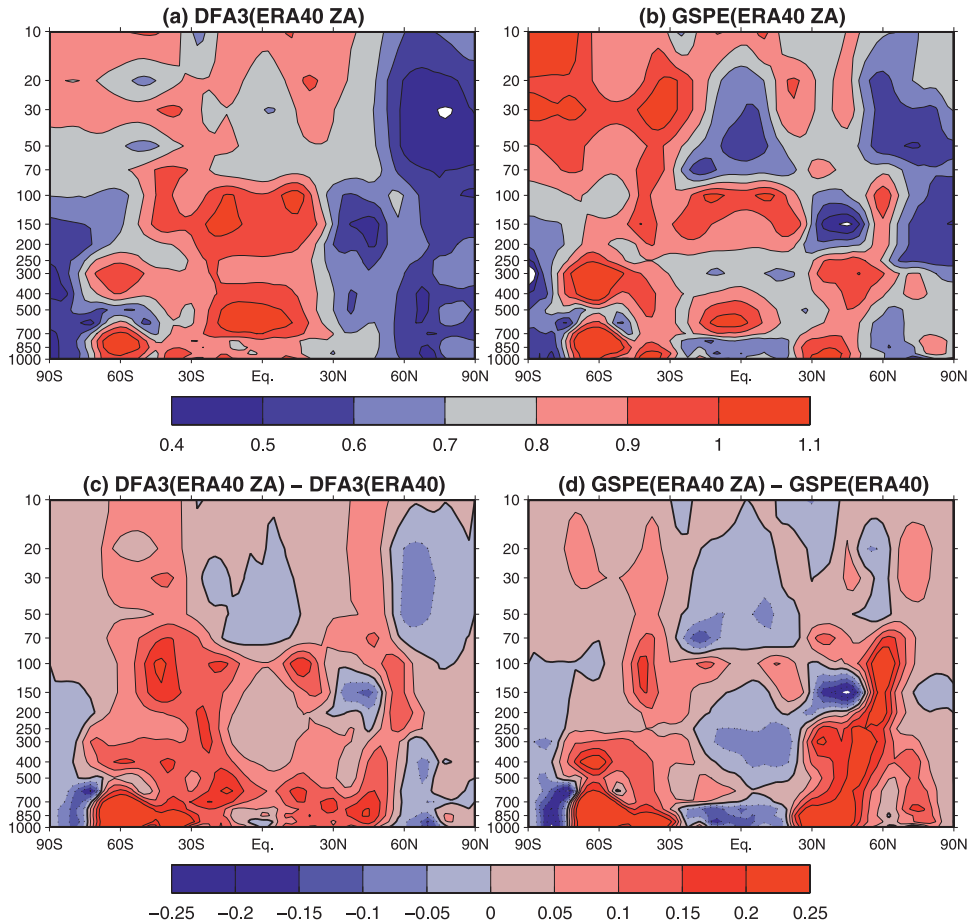
FIG. 9. $\hat{H}$ of the zonal-mean temperature for ERA-40 data for (a) DFA3(a) and (b) GSPE(a). H of the zonal-mean temperature minus zonal mean of $\hat{H}$ for point temperatures, for (c) DFA3(a) and (d) GSPE(a).

more pronounced. The boost provided by taking the zonal mean first, which is shown in Figs. 9c,d, is remarkably robust between the two H estimate methods.

There is a possibility that some of the boost seen in Figs. 9c,d comes about because of an aggregation effect that arises when independent power-law time series are averaged. In particular, when independent power-law time series are averaged, the H of the mean is greater than the mean H of the individual time series (Granger 1980). Although the temperature time series are spatially correlated, this aggregation effect might still operate on sufficiently large scales. We test for the aggregation effect by the following Monte Carlo test: we create a set of independent synthetic temperature time series with values of H equal to the estimated H at each spatial point in the ERA-40 grid represented in Fig. 8b. Therefore, we simulate $144 \times 73 \times 18$ mutually uncorrelated time series using the ARFIMA(0, d, 0) model. The zonal-mean $\hat{H}$ of this dataset is, by construction, the same as that seen in Fig. 8b. We note that we do not

include spatial correlations in order to focus on the aggregation effect. We then estimate $\hat{H}$ of the zonal averages of these synthetic time series. The obtained spatial patterns (not shown) are noisier than Figs. 8a,b but are numerically close to it. Thus, for mutually uncorrelated time series the statistical aggregation effect is negligible. This suggests that the boost in H from using the zonal-mean temperature is of dynamical origin and stems from systematic zonal correlations of the eddy fields.

The boost in Figs. 9c,d is relatively small in the tropics, consistent with the idea of Sobel et al. (2002) that point temperatures in the tropics are well correlated with the zonal-mean tropical temperature field. The enhanced values of H in midlatitudes, at the surface, and in the lower stratosphere suggest that the long-memory behavior in the tropics is coupled to the midlatitudes via the eddy-driven zonal-mean overturning circulation (Held and Schneider 1999). In particular, it might be that interdecadal variability in the tropics drives long-memory behavior in the extratropics via

extratropical teleconnection dynamics. These ideas clearly warrant further study.

4. Conclusions

Under the working assumption that the atmospheric general circulation exhibits power-law behavior, we have estimated the Hurst exponent H for the global atmosphere using several statistical methods. Monte Carlo benchmarking with pure power-law time series reveals no obvious discrepancies between the methods, but when we apply these methods to observed climate data we find a striking degree of inconsistency among the results. We summarize our current understanding of the methods as follows:

- DFA3 results are insensitive to trends and can be made insensitive to high-frequency periodicities, provided trimming is not applied and all time scales are used, that is, provided s_{long} is set to N.
- Multitapered and periodogram spectral methods can be made consistent with one another provided that consistent frequency ranges are used and the lowest frequencies are included. Because the two methods yield consistent H estimates (Fig. 6a), there is no obvious advantage to using multitapering in H estimation, at least in this application.
- The DFA3 and the spectral methods $\hat{H}$ results are quite inconsistent unless filtering is applied, consistent frequency ranges are chosen, and the lowest frequencies are included. The spectral methods are sensitive to periodicities and trends, and DFA3 appears to be more robust in this regard.

Given our current understanding, we recommend the use of either DFA3(a) and GSPE(a), or alternatively DFA3(a) and GPHE(a), and tests to filter well-known climate signals, such as the QBO, ENSO, and external climate forcings, to provide a representative picture of power-law behavior in climate time series. Another issue that has arisen is the different sensitivities the methods exhibit to data inhomogeneities (e.g., temperature jumps induced by instrumentation changes). A proper examination of this issue is beyond the scope of our study, but previous work suggests that DFA3 is also more robust in the presence of such inhomogeneities (Berton 2004; Chen et al. 2002; Hu et al. 2001).

Using DFA3(a) and GSPE(a), we have found several robust high H regions in the atmospheric general circulation. In particular, we have found that point temperature statistics exhibit robust power-law behavior in the tropics that decreases with latitude. A connection between the tropics and extratropics becomes evident when $\hat{H}$ is calculated for the zonal-mean temperature.

We are just beginning to analyze this behavior, which is connected to interannual-to-decadal coupled ocean–atmosphere variability generated in the tropics once ENSO effects are filtered out.

These results may have practical implications for analysis of tropical tropospheric temperature trends (e.g., Santer et al. 2005; Randel and Wu 2006; Thorne et al. 2007). These trends are highly nonrobust: satellite measurements, radiosondes, and climate model simulations all provide different values for these trends. However, if tropical temperatures exhibit power-law behavior, confidence intervals on these trends would very likely be underestimated using AR1-based noise models. The calculation of confidence intervals according to a power-law model for the residuals as carried out in Smith (1993) and Vyushin et al. (2007) would lead to a significant increase of the trends confidence intervals. Thus, at least some of the apparent discrepancies could be accounted for by properly representing long-range temporal correlations in the tropical atmosphere.

Another robust result, found both for DFA3(a) and GSPE(a), is that volcanic forcing increases $\hat{H}$ in the lower tropical and subtropical stratosphere. Volcanic forcing has also been found to have an effect on H at the surface (Vyushin et al. 2004), and it still remains to reconcile the surface and stratospheric H signatures. Furthermore, because the volcanic forcing record can be imprinted in the deep oceanic circulation (Delworth et al. 2005; Gleckler et al. 2006), this result suggests that some of the long-memory behavior seen in the coupled ocean–atmosphere system might be attributed to a volcanic forcing effect.

REFERENCES

Ammann, C. M., G. A. Meehl, W. M. Washington, and C. S. Zender, 2003: A monthly and latitudinally varying volcanic forcing dataset in simulations of 20th century climate. *Geophys. Res. Lett.,* **30,** 1657, doi:10.1029/2003GL016875.

Beran, J., 1992: A goodness-of-fit test for time series with long range dependence. *J. Roy. Stat. Soc.,* **54B,** 749–760.

——, 1994: *Statistics for Long-Memory Processes.* Chapman and Hall, 315 pp.

Berton, R. P. H., 2004: Influence of a discontinuity on the spectral and fractal analysis of one-dimensional data. *Nonlinear Processes Geophys.,* **11,** 659–682.

Bloomfield, P., 1992: Trends in global temperature. *Climatic Change,* **21,** 1–16.

Bretherton, C., and D. Battisti, 2000: An interpretation of the results from atmospheric general circulation models forced by the time history of the observed sea surface temperature distribution. *Geophys. Res. Lett.,* **27,** 767–770.

Bromwich, D. H., and R. L. Fogt, 2004: Strong trends in the skill of the ERA-40 and NCEP/NCAR reanalyses in the high and middle latitudes of the Southern Hemisphere, 1958–2001. *J. Climate,* **17,** 4603–4619.

Chen, Z., P. C. Ivanov, K. Hu, and H. E. Stanley, 2002: Effect of nonstationarities on detrended fluctuation analysis. *Phys. Rev.,* **65E,** 041107, doi:10.1103/PhysRevE.65.041107.

Dell'Aquila, A., P. M. Ruti, S. Calmanti, and V. Lucarini, 2007: Southern Hemisphere midlatitude atmospheric variability of the NCEP-NCAR and ECMWF reanalyses. *J. Geophys. Res.,* **112,** D08106, doi:10.1029/2006JD007376.

Delworth, T. L., V. Ramaswamy, and G. L. Stenchikov, 2005: The impact of aerosols on simulated ocean temperature and heat content in the 20th century. *Geophys. Res. Lett.,* **32,** L24709, doi:10.1029/2005GL024457.

Eichner, J., E. Koscielny-Bunde, A. Bunde, S. Havlin, and H. Schellnhuber, 2003: Power-law persistence and trends in the atmosphere: A detailed study of long temperature records. *Phys. Rev.,* **68E,** 046133, doi:10.1103/PhysRevE.68.046133.

Fox, R., and M. Taqqu, 1986: Large sample properties of parameter estimates for strongly dependent stationary Gaussian time series. *Ann. Stat.,* **17,** 517–532.

Fraedrich, K., and R. Blender, 2003: Scaling of atmosphere and ocean temperature correlations in observations and climate models. *Phys. Rev. Lett.,* **90,** 108501, doi:10.1103/PhysRevLett.90.108501.

Geweke, J., and S. Porter-Hudak, 1983: The estimation and application of long memory time series models. *J. Time Ser. Anal.,* **4,** 221–238.

Ghil, M., 2002: Advanced spectral methods for climatic time series. *Rev. Geophys.,* **40,** 1003, doi:10.1029/2000RG000092.

Gil-Alana, L., 2005: Statistical modeling of the temperatures in the Northern Hemisphere using fractional integration techniques. *J. Climate,* **18,** 5357–5369.

Gleckler, P. J., K. AchutaRao, J. M. Gregory, B. D. Santer, K. E. Taylor, and T. M. L. Wigley, 2006: Krakatoa lives: The effect of volcanic eruptions on ocean heat content and thermal expansion. *Geophys. Res. Lett.,* **33,** L17702, doi:10.1029/2006GL026771.

Granger, C., 1980: Long memory relationships and the aggregation of dynamic models. *J. Econometrics,* **14,** 227–238.

Hasselmann, K., 1976: Stochastic climate models. Part 1: Theory. *Tellus,* **28,** 473–485.

Held, I. M., and T. Schneider, 1999: The surface branch of the zonally averaged mass transport circulation in the troposphere. *J. Atmos. Sci.,* **56,** 1688–1697.

Heneghan, C., and G. McDarby, 2000: Establishing the relation between detrended fluctuation analysis and power spectral density analysis for stochastic processes. *Phys. Rev.,* **62E,** 6103–6110.

Hu, K., Z. Chen, P. C. Ivanov, P. Carpena, and H. E. Stanley, 2001: Effect of trends on detrended fluctuation analysis. *Phys. Rev.,* **64E,** 011114, doi:10.1103/PhysRevE.64.011114.

Hurvich, C., R. Deo, and J. Brodsky, 1998: The mean squared error of Geweke and Porter-Hudak's estimator of the memory parameter of a long-memory time series. *J. Time Ser. Anal.,* **19,** 19–46.

Huybers, P., and W. Curry, 2006: Links between annual, Milankovitch, and continuum temperature variability. *Nature,* **441,** 329–332, doi:10.1038/nature04745.

Jánosi, I., and R. Müller, 2005: Empirical mode decomposition and correlation properties of long daily ozone records. *Phys. Rev.,* **71E,** 056126, doi:10.1103/PhysRevE.71.056126.

Kantelhardt, J., E. Koscielny-Bunde, H. Rego, S. Havlin, and A. Bunde, 2001: Detecting long-range correlations with detrended fluctuation analysis. *Physica A,* **295,** 441–454.

Lorenz, E., 1967: *The Nature and Theory of the General Circulation of the Atmosphere.* WMO, 161 pp.

Marković, D., and M. Koch, 2005: Sensitivity of Hurst parameter estimation to periodic signals in time series and filtering approaches. *Geophys. Res. Lett.,* **32,** L17401, doi:10.1029/2005GL024069.

Marshall, G. J., 2002: Trends in Antarctic geopotential height and temperature: A comparison between radiosonde and NCEP–NCAR reanalysis data. *J. Climate,* **15,** 659–674.

McCoy, E., A. Walden, and D. Percival, 1998: Multitaper spectral estimation of power law processes. *IEEE Trans. Signal Process.,* **46,** 655–668.

Moulines, E., and P. Soulier, 2002: Semiparametric spectral estimation for fractional processes. *Theory and Applications of Long-Range Dependence,* P. Doukhan et al., Eds., Birkhauser, 251–301.

Pelletier, J., 1997: Analysis and modeling of the natural variability of climate. *J. Climate,* **10,** 1331–1342.

Peng, C., S. Buldyrev, A. Goldberger, S. Havlin, M. Simons, and H. Stanley, 1993: Finite-size effects on long-range correlations: Implications for analyzing DNA sequences. *Phys. Rev.,* **47E,** 3730–3733.

Percival, D., and A. Walden, 1993: *Spectral Analysis for Physical Applications.* Cambridge University Press, 611 pp.

——, J. Overland, and H. Mofjeld, 2001: Interpretation of North Pacific variability as a short- and long-memory process. *J. Climate,* **14,** 4545–4559.

Randel, W., and F. Wu, 1999: Cooling of the Arctic and Antarctic polar stratospheres due to ozone depletion. *J. Climate,* **12,** 1467–1479.

——, and ——, 2006: Biases in stratospheric and tropospheric temperature trends derived from historical radiosonde data. *J. Climate,* **19,** 2094–2104.

——, ——, and D. Gaffen, 2000: Interannual variability of the tropical tropopause derived from radiosonde data and NCEP reanalyses. *J. Geophys. Res.,* **105,** 15 509–15 524.

Riedel, K. S., and A. Sidorenko, 1995: Minimum bias multiple taper spectral estimation. *IEEE Trans. Signal Process.,* **43,** 188–195.

Robinson, P., 1995a: Gaussian semiparametric estimation of long range dependence. *Ann. Stat.,* **23,** 1630–1661.

——, 1995b: Log-periodogram regression of time series with long range dependence. *Ann. Stat.,* **23,** 1048–1072.

Santer, B., and Coauthors, 2005: Amplification of surface temperature trends and variability in the tropical atmosphere. *Science,* **309,** 1551–1556.

Schneider, T., 2006: The general circulation of the atmosphere. *Annu. Rev. Earth Planet. Sci.,* **34,** 655–688.

Smith, R., 1993: Long-range dependence and global warming. *Statistics for the Environment,* V. Barnett and F. Turkman, Eds., John Wiley, 141–161.

Sobel, A., I. Held, and C. Bretherton, 2002: The ENSO signal in tropical tropospheric temperature. *J. Climate,* **15,** 2702–2706.

Stephenson, D., V. Pavan, and R. Bojariu, 2000: Is the North Atlantic Oscillation a random walk? *Int. J. Climatol.,* **20,** 1–18.

Taqqu, M., 2002: Fractional Brownian motion and long-range dependence. *Theory and Applications of Long-Range Dependence,* P. Doukhan et al., Eds., Birkhauser, 5–38.

——, V. Teverovsky, and W. Willinger, 1995: Estimators for long-range dependence: An empirical study. *Fractals,* **3,** 785–798.

Thomson, D. J., 1982: Spectrum estimation and harmonic analysis. *Proc. IEEE,* **70,** 1055–1096.

Thorne, P., and Coauthors, 2007: Tropical vertical temperature trends: A real discrepancy? *Geophys. Res. Lett.,* **34,** L16702, doi:10.1029/2007GL029875.

Trenberth, K., and L. Smith, 2006: The vertical structure of temperature in the tropics: Different flavors of El Niño. *J. Climate,* **19,** 4956–4973.

Tsonis, A., P. Roebber, and J. Elsner, 1999: Long-range correlations in the extratropical atmospheric circulation: Origins and implications. *J. Climate,* **12,** 1534–1541.

von Storch, H., and F. Zwiers, 1999: *Statistical Analysis in Climate Research.* Cambridge University Press, 494 pp.

Vyushin, D., I. Zhidkov, S. Havlin, A. Bunde, and S. Brenner, 2004: Volcanic forcing improves atmosphere–ocean coupled general circulation model scaling performance. *Geophys. Res. Lett.,* **31,** L10206, doi:10.1029/2004GL019499.

——, V. E. Fioletov, and T. G. Shepherd, 2007: Impact of long-range correlations on trend detection in total ozone. *J. Geophys. Res.,* **112,** D14307, doi:10.1029/2006JD008168.

Yulaeva, E., and J. Wallace, 1994: The signature of ENSO in global temperature and precipitation fields derived from the microwave sounding unit. *J. Climate,* **7,** 1719–1736.