

# Modeling and understanding persistence of climate variability

D. I. Vyushin,<sup>1</sup> P. J. Kushner,<sup>1</sup> and Francis Zwiers<sup>2</sup>

Received 11 June 2012; revised 3 September 2012; accepted 20 September 2012; published 6 November 2012.

[1] In this study, two parsimonious statistical representations of climate variability on interannual to multidecadal timescales are compared: the short-memory first order autoregressive representation (AR1) and the long-memory “power law” representation. Parameters for each statistical representation are fitted to observed surface air temperature at each spatial point. The parameter estimates from observations are found in general to be captured credibly in the Coupled Model Intercomparison Project 3 (CMIP3) simulations. The power law representation provides an upper bound and the AR1 representation provides a lower bound on persistence as measured by the lag-one autocorrelation. Both representations fit the data equally well according to goodness-of-fit-tests. Comparing simulations with and without external radiative forcings shows that anthropogenic forcing has little effect on the measures of persistence considered (for detrended data). Given that local interannual to multi decadal climate variability appears to be more persistent than an AR1 process and less persistent than a power law process, it is concluded that both representations are potentially useful for statistical applications. It is also concluded that current climate simulations can well represent interannual to multidecadal internal climate persistence in the absence of natural and anthropogenic radiative forcing, at least to within observational uncertainty.

**Citation:** Vyushin, D. I., P. J. Kushner, and F. Zwiers (2012), Modeling and understanding persistence of climate variability, *J. Geophys. Res.*, 117, D21106, doi:10.1029/2012JD018240.

## 1. Introduction

[2] Understanding climate persistence is one of the principal goals of statistical climatology. This study aims to better characterize climate persistence by comparing two parsimonious statistical representations of the temporal spectral density of climatic time series: the short-memory AR1 representation and the long-memory power law representation. (The term *statistical representation* is here used instead of *statistical model* to avoid confusion with numerical climate models.) This paper also places these two representations into a broader statistical context.

[3] The well known AR1 representation [e.g., Hasselmann, 1976] can be written as a stochastic process  $X_t = \phi X_{t-1} + \varepsilon_t$ , where  $\phi \in (0, 1)$  is the lag-one autocorrelation and  $\varepsilon_t$  represents white noise innovations. The AR1 process is *short-memory* with an autocovariance function that decays exponentially with lag  $\tau$ ,  $\gamma_{\text{AR1}}(\tau) \sim \phi^{|\tau|}$ , and a spectral density as a function of frequency  $f$

$$S_{\text{AR1}}(f) \sim (1 - 2\phi \cos(2\pi f) + \phi^2)^{-1}, \quad |f| \leq 1/2, \quad (1)$$

<sup>1</sup>Department of Physics, University of Toronto, Toronto, Ontario, Canada.

<sup>2</sup>Pacific Climate Impacts Consortium, University of Victoria, Victoria, British Columbia, Canada.

Corresponding author: D. I. Vyushin, Department of Physics, University of Toronto, 60 St. George St., Toronto, ON M5S 1A7, Canada. (vyushin@atmosphysics.utoronto.ca)

©2012. American Geophysical Union. All Rights Reserved.  
0148-0227/12/2012JD018240

that saturates to a constant near the origin (i.e. for  $|f| \ll 1/2$ ) (see Appendix A and, e.g., Brockwell and Davis [1998]). The *power law* representation [e.g., Beran, 1994] can be defined with reference to a fractionally integrated autoregressive process FAR(0,  $d$ ),  $(1 - B)^d X_t = \varepsilon_t$  where  $B$  is the backshift operator satisfying  $BM_t = M_{t-1}$  for random variable  $M_t$ , and  $d$  is the order of fractional integration [e.g., Beran, 1994]. It is conventional to write  $H = d + 1/2$ , where  $H$  is the so-called “Hurst exponent” [Hurst, 1951]; it is assumed that  $H \in (1/2, 1)$ . This process is *long-memory* with an algebraically decaying autocovariance function  $\gamma_{\text{PL}}(\tau) \sim |\tau|^{2H-2}$  and a spectral density that increases asymptotically with decreasing frequency according to a power law [e.g., Taqqu, 2002]

$$S_{\text{PL}}(f) \sim |f|^{1-2H}. \quad (2)$$

[4] In this article, we adopt the viewpoint that estimates of  $\phi$  and  $H$  provide related measures of persistence. For AR1, the fraction of predictable variance (FPV) of  $X_t$  given past history, i.e.  $X_{t-1}$ , is  $\phi^2$ . For FAR(0,  $d$ ) the FPV is  $1 - \Gamma^2(3/2 - H)/\Gamma(2 - 2H)$ , which is monotonically increasing in  $H \in (1/2, 1)$  [e.g., Beran, 1994]. Thus, greater values of  $\phi$  and  $H$  correspond to greater FPV in each representation. These two measures of persistence can be combined via the so-called FAR(1,  $d$ ) stochastic process  $(1 - \phi_1 B)(1 - B)^d X_t = \varepsilon_t$  that generalizes the AR1 and power law representations [e.g., Stephenson *et al.*, 2000]. The FAR(1,  $d$ ) reverts to AR1 with lag-one autocorrelation  $\phi = \phi_1$  when

$d \rightarrow 0$  and to the power law process  $\text{FAR}(0, d)$  with slope  $-2d = 1 - 2H$  when  $\phi_1 \rightarrow 0$ . The FPV for  $\text{FAR}(1, d)$  is

$$1 - \frac{1 - \phi_1^2}{\Gamma(1 - 2d)} \left\{ \frac{1}{\Gamma^2(1 - d)} + \frac{\sin(\pi d)}{\pi} \sum_{k=1}^{\infty} \phi_1^k \frac{\Gamma(k + d)}{\Gamma(k + 1 - d)} \right\}^{-1}, \quad (3)$$

which can be shown to revert to the FPV of the AR1 process when  $d \rightarrow 0$  (i.e. when  $H \rightarrow 1/2$ ) and to the FPV of the power law process  $\text{FAR}(0, d)$  when  $\phi_1 \rightarrow 0$ .

[5] The AR1 and power law stochastic processes provide distinct limiting cases of persistence and the behavior of the spectral density for low frequencies  $|f| \ll 1/2$ , which is controlled in practice by the length of a given time series. The AR1 representation captures the tendency for the variability of some climatic time series to saturate in the transition from weekly, to intraseasonal, annual, and decadal variability [e.g., *Frankignoul and Hasselmann, 1977; Feldstein, 2000*]. But the power law representation may be more appropriate for long instrumental and paleoclimate records, for which variability tends to build in the transition to decadal, centennial, and millennial timescales [e.g., *Bloomfield, 1992; Pelletier, 1997; Tsonis et al., 1999; Caballero et al., 2002; Eichner et al., 2003; Fraedrich and Blender, 2003; Vyushin et al., 2004; Huybers and Curry, 2006*]. This study explores the use of these two representations in the intermediate interannual-to-multidecadal range captured by the instrumental observational record and current climate simulations.

[6] While other higher order statistical representations [including  $\text{FAR}(1, d)$ ] are available, this study focuses on these two distinct, parsimonious, and commonly used representations. It evaluates their validity, and examines the ability of climate models to reproduce observed patterns of their parameters, and explain selected aspects of their behavior. In particular, some understanding of *internal* climate persistence, i.e. the persistence that occurs in the absence of anthropogenic and natural radiative forcing, can be obtained by comparing simulations with and without external radiative forcings. Only a few studies have sought to compare these two representations on an even footing [e.g., *Stephenson et al., 2000; Percival et al., 2001; Caballero et al., 2002; Vyushin et al., 2007; Vyushin and Kushner, 2009; Franzke, 2012*] and none have systematically compared them in climate simulations and observations or linked them in the way presented below.

[7] Observational data and simulations used and statistical methods employed are described in Section 2. Section 3 evaluates the ability of models to capture the observed spatial distribution of  $\phi$  and  $H$  and the dependence of these parameters on analysis timescale. Section 4 relates the two measures of persistence to each other and evaluates their validity. The study concludes with a discussion of implications of this analysis in Section 5. Appendix A outlines mathematical concepts in the study, and Appendix B provides details of a goodness-of-fit test employed in Section 4.

## 2. Data and Methods

[8] The observational products used are the NCEP/NCAR reanalysis [*Kalnay et al., 1996*], the ERA40 reanalysis [*Uppala et al., 2005*], and the NASA GISS surface air temperature (SAT) data set [*Hansen et al., 1999*] over the period September

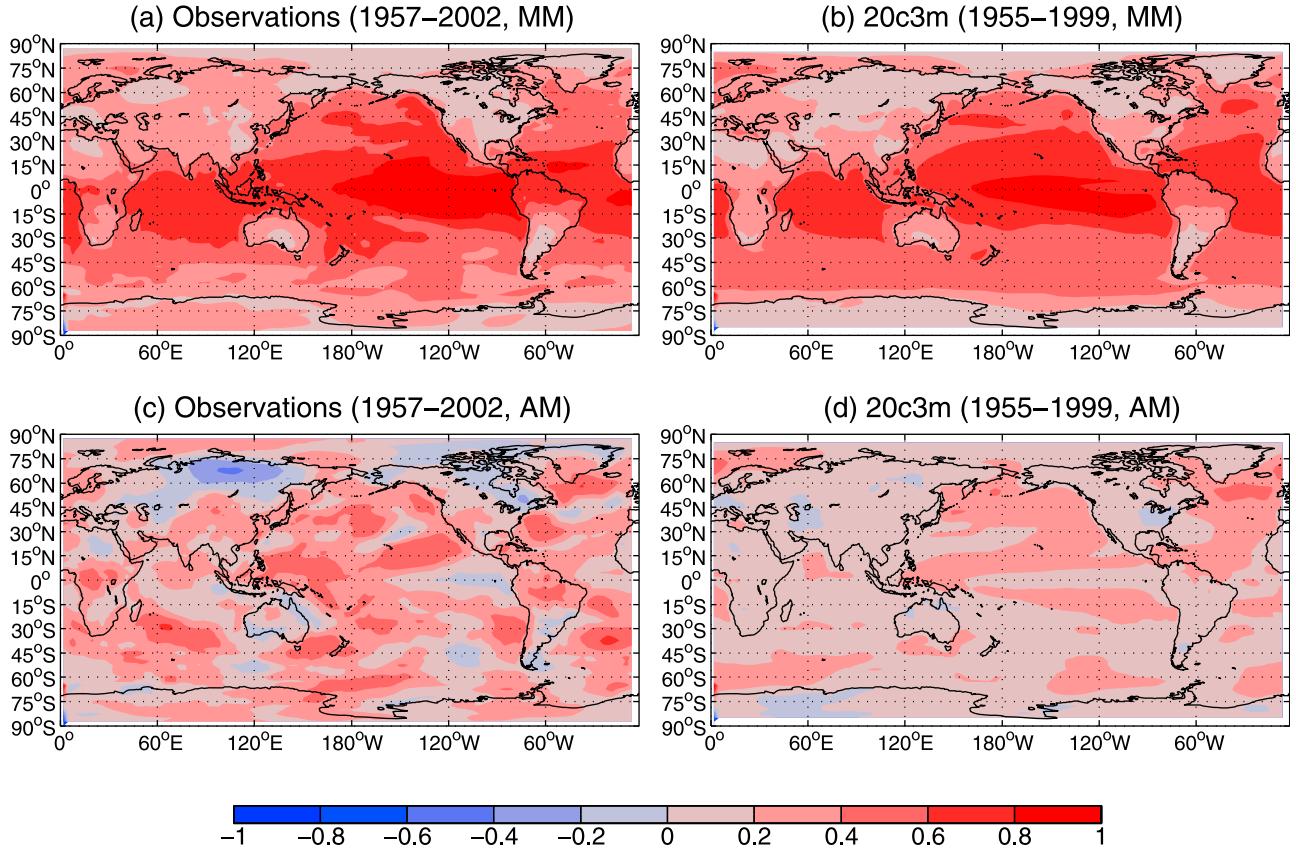
1957 to August 2002 covered by ERA40. The climate simulations used are the pre-industrial control (*picntrl*) and the 20th century (*20c3m*) simulations of 17 CMIP3 atmosphere-ocean coupled general circulation models: CGCM3.1(T47), CGCM3.1(T63), CSIRO-Mk3.0, CSIRO-Mk3.5, ECHAM5/MPI-OM, GFDL-CM2.0, GFDL-CM2.1, GISS-AOM, GISS-EH, GISS-ER, MIROC3.2(medres), MIROC3.2(hires), MRI-CGCM2.3.2, NCAR CCSM3.0, NCAR PCM, UKMO-HadCM3, UKMO-HadGEM1. A single realization is used from each model; results are insensitive to choice of realizations (not shown). Both 1955–1999 and 1900–1999 *20c3m* simulation segments are analyzed. Six 500 year *picntrl* simulations of six GCMs (CGCM3.1(T47), ECHAM5/MPI-OM, GFDL-CM2.0, GFDL-CM2.1, GISS-ER, MIROC3.2(medres)) are also analyzed. The seasonal cycle and its first three harmonics are filtered out from all time series.

[9] The lag-one autocorrelation coefficient  $\phi$  is estimated by the Yule-Walker method [*von Storch and Zwiers, 1999*] in the time domain and by maximum-likelihood fitting of the AR1 spectral density to the periodogram [*Beran, 1994*] in the spectral domain. Results from the two methods are consistent and only Yule-Walker results will be displayed, except in Appendix B and Figure 8. These calculations use linearly detrended data. Similarly, the Hurst exponent  $H$  is also estimated by time- and spectral-domain methods: detrended fluctuation analysis of the third order (DFA3) [*Kantelhardt et al., 2001*] (with modifications introduced by *Vyushin and Kushner [2009]*) in the time domain and the Gaussian Semiparametric Estimator (GSPE) [*Robinson, 1995*] in the spectral domain. Compared to the  $\phi$  estimates, GSPE and DFA3  $H$  estimates are relatively method dependent; consistency requires removal of periodic components and linear trends, and care to use equivalent frequency ranges [*Vyushin and Kushner, 2009*]. DFA3 is generally preferred in this study because it removes the effect of linear and quadratic trends without prior detrending. But GSPE is used in Figures 5 and 6 and Appendix B for reasons mentioned below. In the notation used below,  $\hat{\phi}$  and  $\hat{H}$  represent estimates of  $\phi$  and  $H$ .

[10] The estimators used in this article have been implemented in an open-source R package, *PowerSpectrum*, <http://tinyurl.com/4v985le> that we have developed. The package also provides estimators of trend confidence intervals based on different statistical representations for the residuals, various estimators of power and cross-spectrum with their confidence intervals, the spectral goodness-of-fit test, Monte-Carlo tests of the Hurst exponent estimators and the goodness-of-fittest, etc.

[11] In the results below,  $\phi$  and  $H$  are estimated at individual spatial points and the results averaged either across observational products/models or across selected regions defined as follows: North Atlantic ( $30^{\circ}\text{E}$ – $350^{\circ}\text{E}$ ,  $40^{\circ}\text{N}$ – $60^{\circ}\text{N}$ ), the North Pacific ( $149^{\circ}\text{E}$ – $230^{\circ}\text{E}$ ,  $20^{\circ}\text{N}$ – $57^{\circ}\text{N}$ ), the Southern Ocean ( $0^{\circ}\text{E}$ – $360^{\circ}\text{E}$ ,  $40^{\circ}\text{S}$ – $65^{\circ}\text{S}$ ), Main Development Region for Atlantic Hurricane formation (MDR:  $299^{\circ}\text{E}$ – $332^{\circ}\text{E}$ ,  $5^{\circ}\text{N}$ – $22^{\circ}\text{N}$ ), the Maritime Continent ( $98^{\circ}\text{E}$ – $158^{\circ}\text{E}$ ,  $5^{\circ}\text{S}$ – $5^{\circ}\text{N}$ ), Arctic (north of  $65^{\circ}\text{N}$ ), and Antarctic (south of  $65^{\circ}\text{S}$ ).

[12] The study also explores the impact of analysis timescale on climate persistence estimates. For estimates of  $\phi$ , calculations for monthly, annual, pentadal, decadal, and



**Figure 1.** (a) The spatial distribution of the lag-one autocorrelation estimate  $\hat{\phi}$  for the observed monthly mean 1957–2002 SAT. The map represents the average of  $\hat{\phi}$  of the available observational data sets at each spatial point. Note that the GISS SAT is only spatially complete northward of  $50^{\circ}\text{S}$  and so is not included in the average southward of this latitude. (b) As in Figure 1a, for the 1950–1999 20c3m simulations of CMIP3 (17 models); all models are included in the average at all points. (c) As in Figure 1a, for the annual mean (AM) SAT. (d) As in Figure 1b, for the AM SAT.

bidecadal means (denoted MM, AM, PM, DM, BDM) of SAT are carried out. For estimates of  $H$ , calculations for timescale ranges of 18 month–45 years (18 m–45 y), 18 m–100 y, and 5 y–45 y are carried out on MM data, and calculations for the timescale range of 20 y–500 y are carried out on AM data. The  $H$  analysis on the 18 m–45 y range was the focus of the work of *Vyushin and Kushner [2009]* and *Vyushin et al. [2009]*, which analyzed power law behavior in the recent observational record of free atmospheric air temperature.

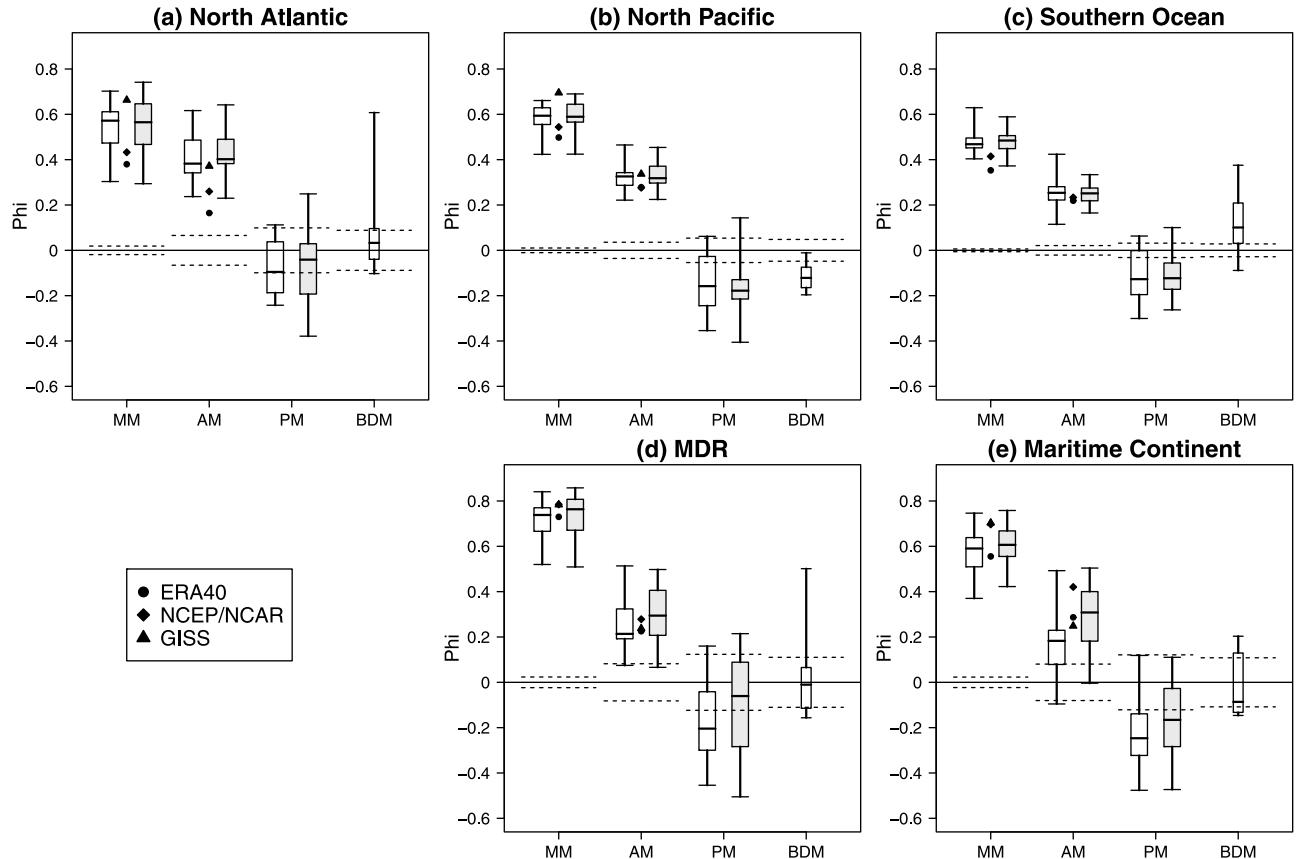
### 3. Estimates of Surface Air Temperature Persistence

#### 3.1. AR1 Representation

[13] The spatial distribution of the lag-one autocorrelation estimate  $\hat{\phi}$  for the observed monthly mean (MM) SAT (Figure 1a) is well captured by the models (Figure 1b). In particular, the models realistically simulate the contrast between large  $\hat{\phi}$  over the oceans and small  $\hat{\phi}$  over the inner continental areas that is expected from simple thermal inertia considerations [*Manabe and Stouffer, 1996*], as well as the observed relatively large  $\hat{\phi}$  in the tropics. Since there are more model data sets (17) than observational data sets

available, the average  $\hat{\phi}$  maps for the models tend to be relatively smooth.

[14] The “MM” portion of each panel in Figure 2 quantifies the observational and model uncertainty in these estimates in selected regions. For example, for the North Atlantic, observational estimates range from  $\hat{\phi} \approx 0.35$  for ERA40 to about  $\hat{\phi} \approx 0.65$  for GISS (points between the box and whisker plots in the MM portion of Figure 2a), and model estimates occupy a similar range with a median value of about  $\hat{\phi} \approx 0.55$ . The spread in model and observational estimates exceeds the sampling uncertainty represented by the confidence intervals plotted with the dashed lines in each panel. For the MM data, none of the regions show obvious model inaccuracies, i.e. the range of model estimates overlaps the range of observational estimates. Furthermore, comparing the *picntr* (unshaded box and whisker plot in the left column of MM portion of each panel) and 20c3m simulations (shaded box and whisker plot, right column), the model estimates are largely insensitive to external forcing (noting that data has been linearly detrended prior to the  $\hat{\phi}$  calculation). In sum, to within observational uncertainty, this measure of persistence appears to reflect basic coupled



**Figure 2.** (a) The North Atlantic spatial average of the lag-one autocorrelation estimates  $\hat{\phi}$  for monthly, annual, pentadal, and bidecadal mean SAT (MM, AM, PM, BDM). Individual points represent observations, unshaded box and whisker symbols represent *picntral* simulations, and shaded box and whisker symbols represent *20c3m* simulations. The time periods represented are 1957–2002 for observed MM and AM observations, 1955–1999 for *20c3m* MM and AM, arbitrary 45 y segments for *picntral* MM and AM, 1900–1999 for *20c3m* PM, arbitrary 100 y segments for *picntral* PM, and available 500 y segments for *picntral* BDM. CMIP3 models used are listed in Section 2. The horizontal dashed lines demonstrate approximate  $\pm 2\sigma$  confidence intervals for  $\hat{\phi} = 0$  [Vyushin, 2009]. The region boundaries are described in Section 2. (b–e) As in Figure 2a for the North Pacific, Southern Ocean, MDR, and Maritime Continent region. The averaged  $\hat{\phi}$  for the GISS SAT is not estimated for the Southern Ocean due to its poor coverage.

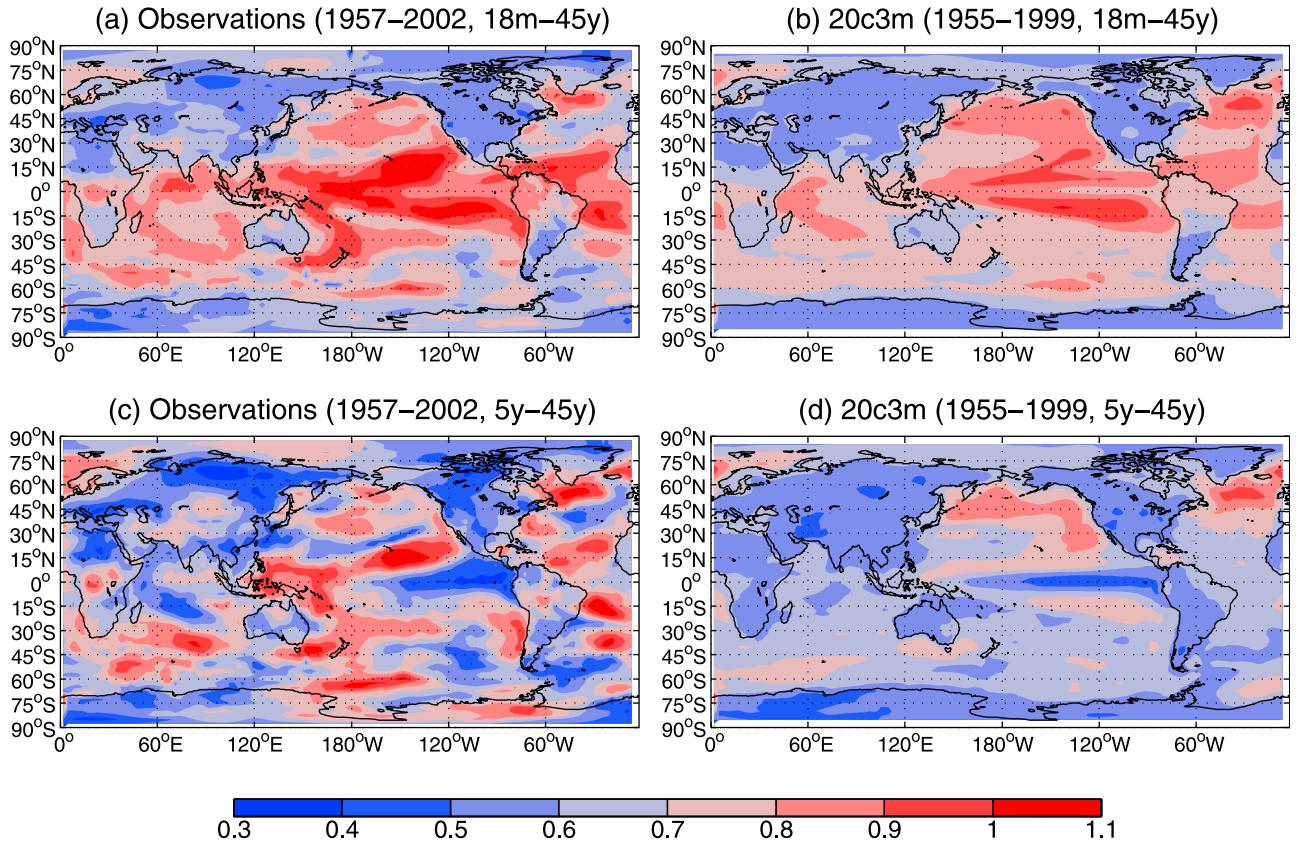
ocean-atmosphere dynamics that current generation climate models are capable of simulating realistically.

[15] This conclusion generally carries over to the observed  $\hat{\phi}$  for AM data (Figures 1c and 1d, “AM” portion of Figures 2a–2e). Not surprisingly, the  $\hat{\phi}$  for AM data is generally smaller than for MM data; this is captured in the simulations (Figures 1c and 1d). The impact of the choice of aggregation timescale (e.g. monthly mean versus annual mean) on measures of persistence like  $\phi$  will be discussed more fully in Section 4.2. In the tropical Pacific the observed and simulated  $\hat{\phi}$  displays west to east gradients of opposite sign depending on whether  $\phi$  is estimated from MM or AM data. The change in gradient with averaging timescale reflects, in part, the dominance of ENSO variability in eastern tropical Pacific in the 5 y–7 y band (see also Section 3.2).

[16] The observations include regions of relatively small and large  $\hat{\phi}$  compared to the models, but some of this difference reflects differences in the number of data sets going into each panel. As for  $\hat{\phi}$  for the MM data, the simulated

range for the AM data of the *20c3m* simulations for individual regions overlaps the observations, except perhaps in the North Atlantic  $\hat{\phi}$  (Figure 2a) where the simulations might be biased high compared to the observations. Another region of discrepancy between the models and observations is a region of anticorrelation from one year to the next over northern Siberia that is suggested in Figure 1c and that is present in all the observational products but either absent or has a somewhat different location in the models (not shown). This feature might be related to forcing of Eurasian temperature by the quasi-biennial oscillation (QBO) of the tropical stratosphere [Thompson et al., 2002], which is not captured in the CMIP3 models. While the extratropical AM  $\hat{\phi}$  is generally insensitive to external forcing, radiative forcing in the *20c3m* appears to boost persistence in the Maritime Continent relative to the *picntral*.

[17] Generally,  $\hat{\phi}$  drops for longer aggregation timescales (transition from MM to BDM in Figure 2). Except over the Southern Ocean, most of the models show insignificant or negative lag-one autocorrelations for PM or BDM,



**Figure 3.** The spatial distribution of the Hurst exponent estimate  $\hat{H}$  for the observed MM 1957–2002 SAT (a) and for the 1950–1999 20c3m simulations of the 17 CMIP3 models (b) calculated for the 18 m–45 y range. (c and d) As in Figures 3a and 3b but for the  $\hat{H}$  estimated using the 5 y–45 y range. As in Figure 1, the maps represent the average over the available observations or the models. Values of  $\hat{H} > 1$  are an artifact of inclusion of high-frequencies into the estimation range.

suggesting a saturation of coupled ocean-atmosphere variability at low frequencies in unforced control runs. The realism of this behavior cannot be assessed within the scope of this study; it is expected that cryospheric and biospheric processes not represented in the CMIP3 models might influence persistence on these longer timescales.

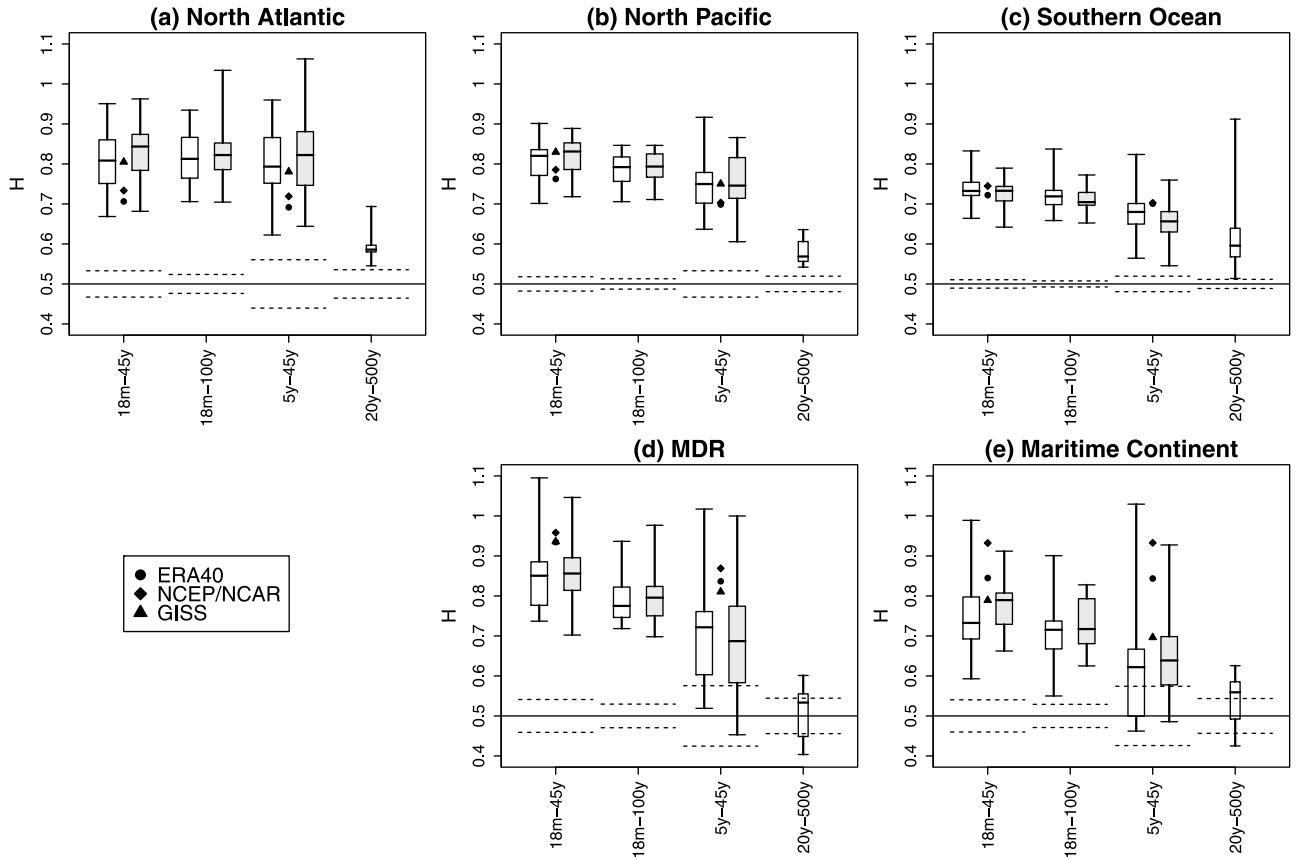
### 3.2. Power Law Representation

[18] Figure 3 displays estimates of the Hurst exponent,  $\hat{H}$ , for the same collections of data sets used to produce  $\phi$  in Figure 1. The CMIP3 models credibly represent the contrast in 18 m–45 y  $\hat{H}$  between land and ocean, including the observed peaks in the northern extratropical oceans and the decrease from west to east in the tropical Pacific and Atlantic (Figures 3a and 3b, “18 m–45 y” portion of Figures 4a–4c). The model 18 m–45 y  $\hat{H}$  is generally biased low over the tropical oceans and parts of the Southern Hemisphere (Figures 3a, 3b, 4d, and 4e). The  $\hat{H}$  values obtained from the climate model simulations are insensitive to increasing the low frequency cutoff of the timescale range from 45 y to 100 y (compare the 18 m–45 y to the 18 m–100 y portions of Figure 4). This makes sense because an increase in the low-frequency cutoff from 45 y to 100 y leads to the addition of only one DFA or periodogram data point at timescales longer than 45 y. This suggests that  $\hat{H}$  on the interannual to

multipdecadal scale may be robustly estimated with just a few decades of data.

[19] The  $\hat{H}$  values for the simulations shown here are largely consistent with previous modeling results [Fraedrich and Blender, 2003; Blender and Fraedrich, 2003; Blender et al., 2006; Rybski et al., 2008], to within model and analysis differences. The  $\hat{H}$  field is quite robust to the presence of external radiative forcing (in Figure 4 compare the *picntr* to the 20c3m results for the 18 m–45 y, 18 m–100 y, and 5 y–45 y ranges). DFA3 removes up to second order trends in the data and thus linear and quadratic trends do not directly affect  $\hat{H}$  in these figures. Consequently, external radiative forcing, whether anthropogenic or natural, does not appear to significantly affect this measure of climate persistence. The calculated  $\hat{H}$  of Huybers and Curry [2006] over the tropical oceans for the NCEP/NCAR reanalysis for the 2 month to 30 year timescale range is relatively large compared to those presented here, because of the relatively high high-frequency cutoff used in that work (comparison not shown).

[20] For the 20 y–500 y range (Figure 4), the CMIP3 models simulate values of  $\hat{H}$  in the extratropical oceans that are marginally significantly greater than 0.5, indicating that the models are capable of capturing some buildup of variance in the low frequency limit. Again, the realism of this behavior



**Figure 4.** (a) The North Atlantic spatial average of  $\hat{H}$  for 18 m–45 y, 18 m–100 y, 5 y–45 y, and 20 y–500 y timescale ranges. The horizontal dashed lines demonstrate the  $\pm 2\sigma$  confidence intervals for the  $\hat{H} = 1/2$  [Vyushin, 2009]; confidence intervals for  $\hat{H} > 1/2$  are similar. Data sources, region boundaries, and plotting conventions are as in Figure 2. (b–e) As in Figure 4a for the North Pacific, Southern Ocean, MDR, and Maritime Continent region.

at the multicentennial scale is difficult to assess within the scope of this study.

[21] As a whole, the CMIP3 models capture many of the important features of the two independently calculated measures of persistence,  $\hat{\phi}$  and  $\hat{H}$ , in the interannual to multidecadal range, particularly over the extratropical ocean basins. Furthermore, these features appear to be related to internal climate variability rather than external radiative forcing, because they are robust to the presence of external radiative forcing once the data has been detrended. Volcanic forcing increases  $H$  in tropical stratospheric air temperature [Vyushin et al., 2009], but the natural radiative forcings present in the CMIP3 models apparently do not strongly affect SAT.

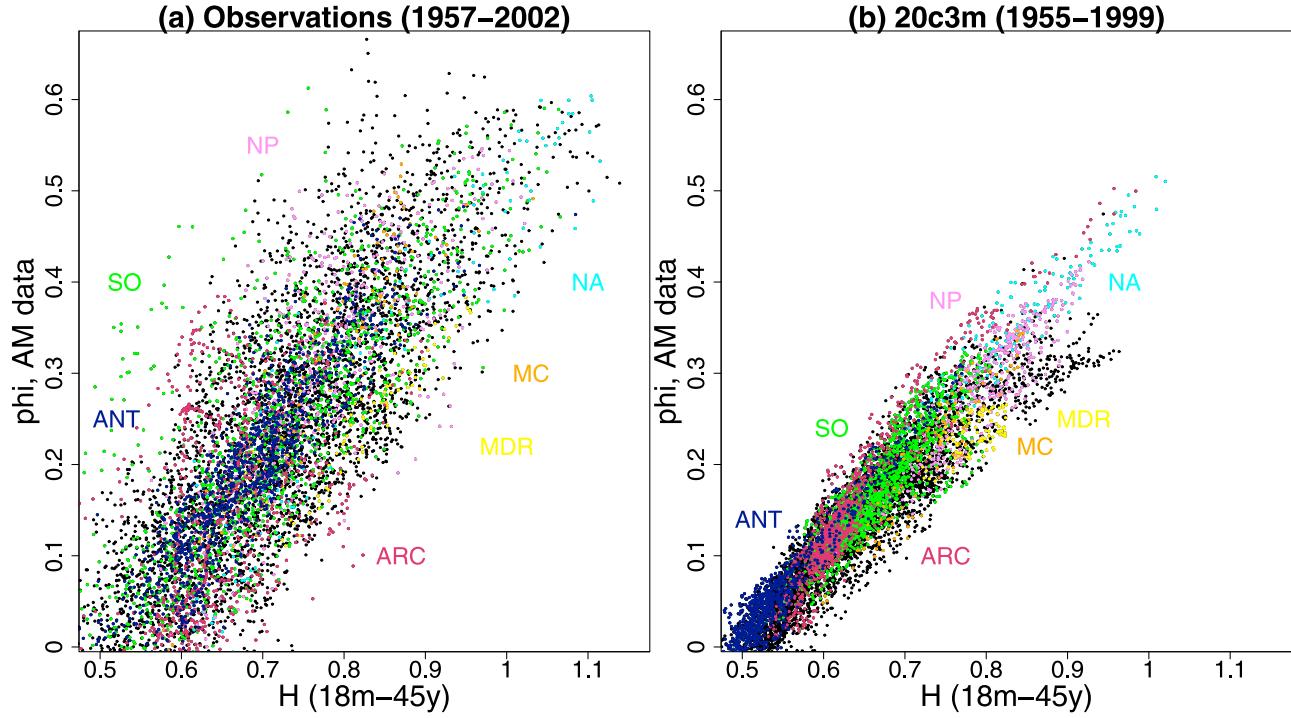
[22] For the  $\hat{H}$  values in the tropics, model spread is pronounced and the models tend to disagree with observations. The  $\hat{H}$  values are sensitive to analysis timescale in this region because of interannual ENSO-related variability, which in the CMIP3 models is simulated on timescales shorter than observed [see Randall et al., 2007, p. 624]. Thus, inconsistencies between the CMIP3 models and observations are expected here for power law fits that include interannual timescales. Tropics are also a region where, on decadal scales, observational products systematically disagree [e.g., Vyushin et al., 2009]. This suggests that less confidence should be

placed in observed estimates, and indeed our dynamical understanding, of tropical persistence on decadal timescales.

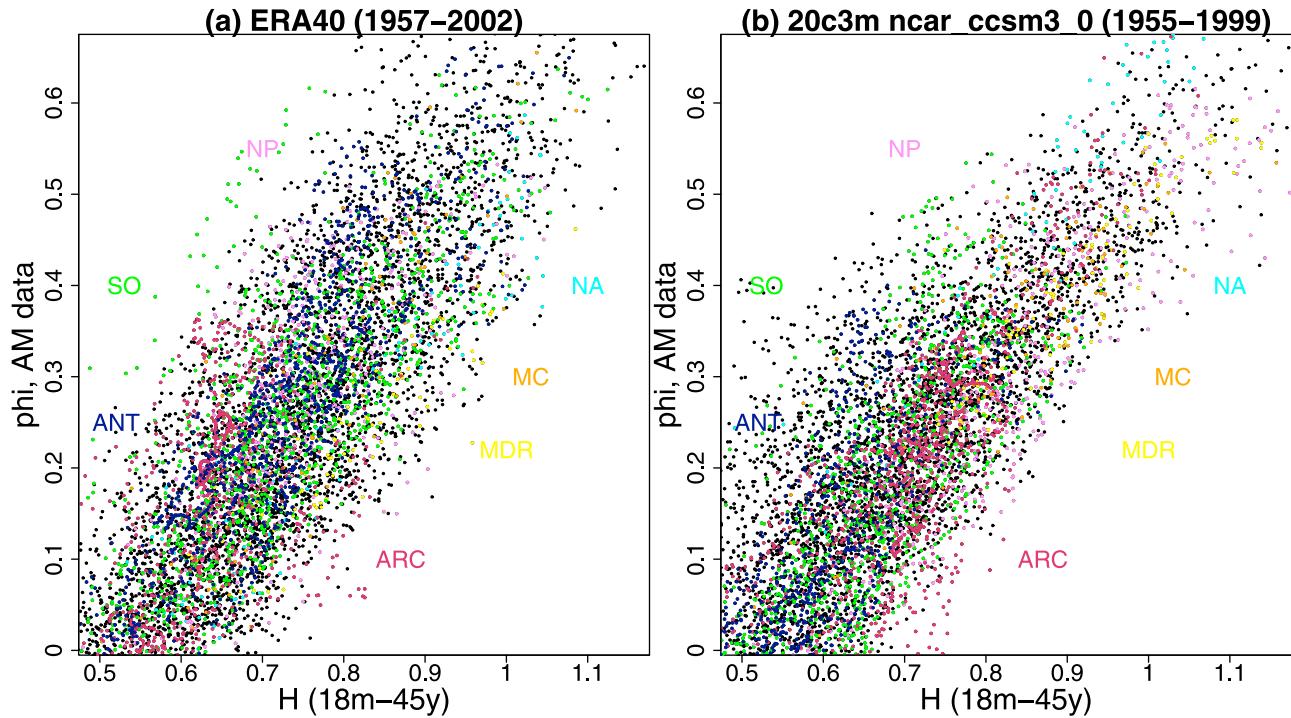
## 4. Comparison of the Two Representations

### 4.1. Relationship Between the Two Representations

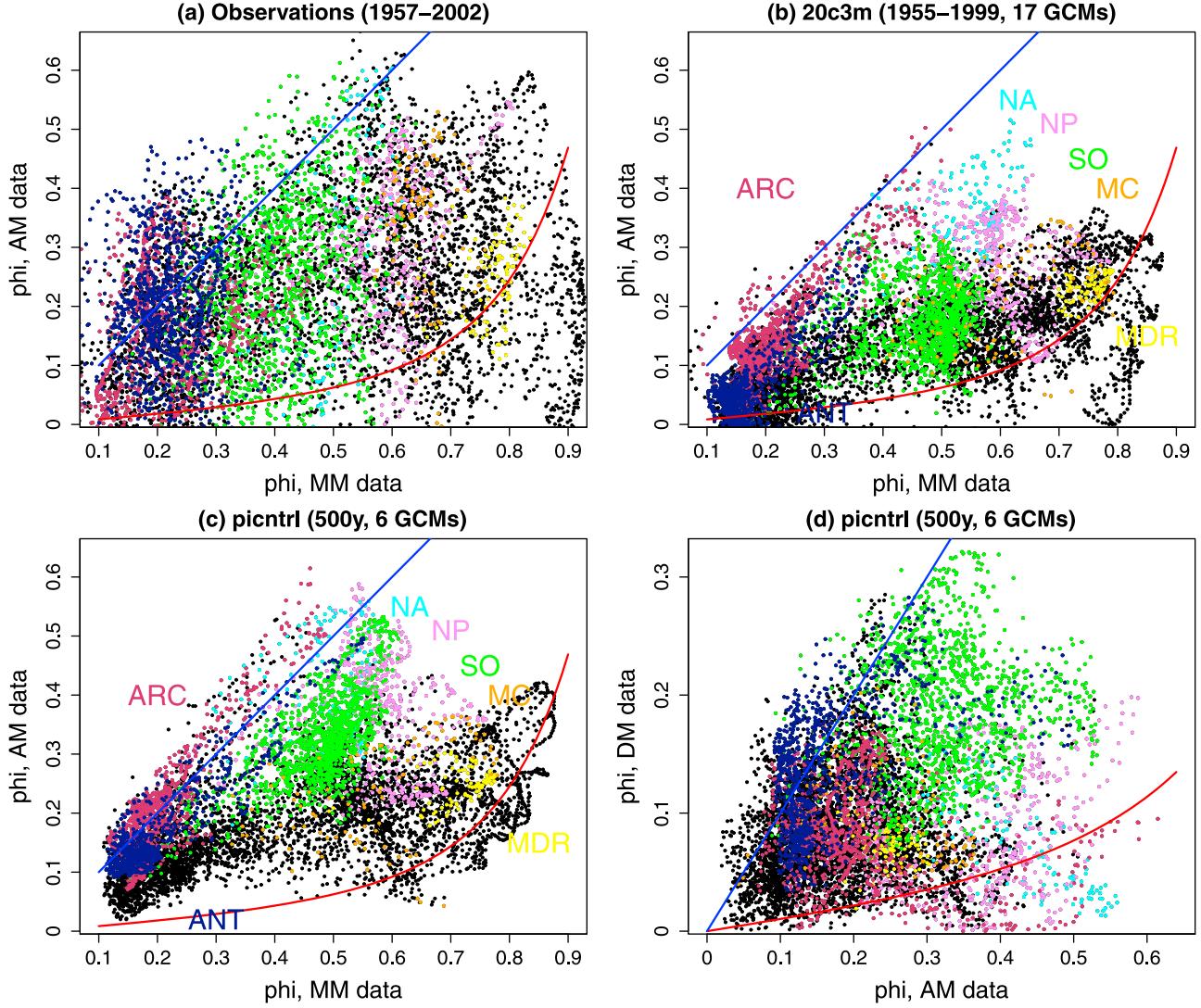
[23] The spatial distribution of the  $\hat{\phi}$  for AM data (Figure 1) and the 18 m–45 y  $\hat{H}$  (Figure 3) share some qualitative features, including larger values in the extratropical oceans, larger values over oceans than land, and relatively large values in the western tropical ocean basins. This similarity suggests that the two fields represent similar information on persistence on the interannual to multidecadal scale. Scattering  $\hat{\phi}$  for the AM data against 18 m–45 y  $\hat{H}$  (Figure 5) brings out a compact relationship between the two persistence measures, such that regions of high  $\hat{H}$  correspond to regions of high  $\hat{\phi}$ . Color coding highlights regions of interest identified in Section 2. The largest degree of persistence identified by both measures is found in the North Atlantic, followed by the North Pacific (see Section 2 for the definition of area boundaries). The Maritime Continent, the Main Development Region for Atlantic Hurricane formation, and the Southern Ocean have intermediate values. The points from the Arctic span a wide range, whereas most



**Figure 5.** (a) Scatterplot of  $\hat{\phi}$  for AM data against GSPE-estimated 18 m–45 y  $\hat{H}$  for SAT. The dots are color coded as follows: cyan: North Atlantic (NA); violet: North Pacific (NP); yellow: MDR; green: Southern Ocean (SO); orange: Maritime Continent (MC); maroon: Arctic (ARC); navy: Antarctica (ANT); black: the rest. The analogous figure with DFA3 estimates of  $H$  looks similar, but is more noisy (not shown). (b) As in Figure 5a, but for ensemble mean 1955–1999 20c3m CMIP3 simulations.



**Figure 6.** As Figure 5 but for (a) ERA40 (1957–2002) and (b) 20c3m 1955–1999 NCAR CCSM 3.0.



**Figure 7.** (a) Scatterplot of  $\hat{\phi}$  for AM data against  $\hat{\phi}$  for MM data for ensemble mean observed data. The blue line is equation (6) for the power law representation; the red curve is equation (5) for the AR1 representation. (b and c) As in Figure 7a for 17 20c3m simulations, for 6 picntrl 500 y simulations. (d) As in Figure 7c, for decadal mean (DM) and AM data (note different axis scales). The following percentage of the points is located between the red and blue curves: 66% (Figure 7a), 73% (Figure 7b), 85% (Figure 7c), and 83% (Figure 7d). Region color coding is as in Figure 5.

of the grid points from Antarctica demonstrate low values of  $\hat{\phi}$  and  $\hat{H}$  for the models and intermediate values [Franzke, 2010] for the observations. Similar compact relationships are found in the decadal-centennial band when decadal SAT  $\hat{\phi}$  is scattered against 20 y–500 y  $\hat{H}$  (not shown).

[24] The relationship between the different persistence measures is more compact for the model ensemble (Figure 5b) than for the observations (Figure 5a) at least in part because of an averaging effect. Scatterplots for an individual observational product and a representative simulation (Figure 6) illustrate a similar spread, because each relates to a single observed or simulated climate realizations.

#### 4.2. Relative Validity of the Two Representations

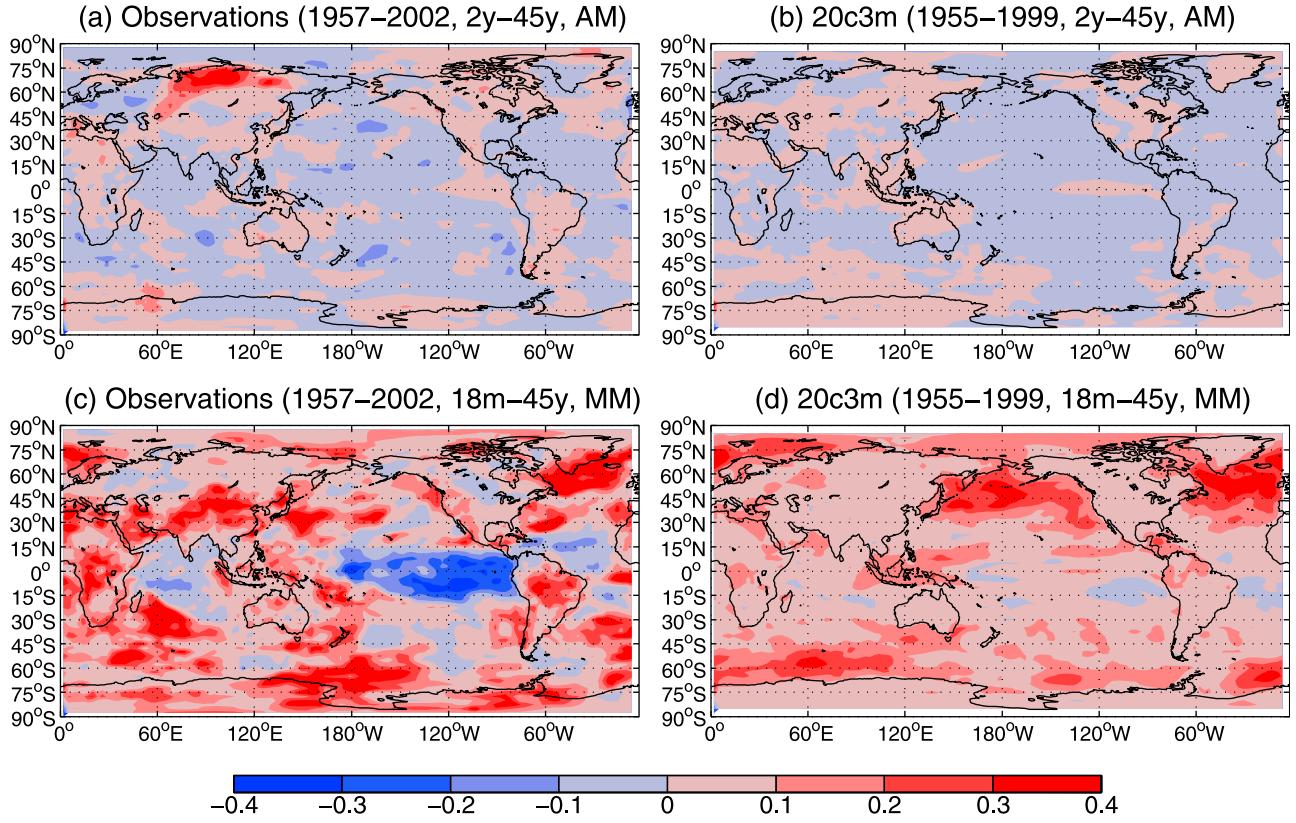
[25] Two tests are used to evaluate the relative validity of the AR1 and power law representations. The first test exploits

the distinctive behavior of the AR1 and power law statistical representations under *temporal aggregation* as is done when, for example, creating an annual mean (AM) time series based on January to December averages of a monthly mean (MM) time series. Define the temporally aggregated time series

$$X_j^{(T)} = \frac{1}{T} \sum_{t=1}^T X_{t+T(j-1)}, \quad j = 1, 2, \dots, N/T, \quad T \geq 1, \quad (4)$$

where  $X_t$ ,  $t = 1, 2, \dots, N$  is the original time series. In this notation,  $X_1^{(12)}$  would be the first value of an AM time series aggregated from the MM time series  $\{X_1, \dots, X_{12}, X_{13}, \dots\}$ . For AR1 processes the temporally aggregated process has lag-one autocorrelation

$$\text{AR1 : } \phi_{(T)} = \frac{\phi(1 - \phi^T)^2}{T(1 - \phi^2) - 2\phi(1 - \phi^T)}, \quad 0 \leq \phi \leq 1. \quad (5)$$



**Figure 8.** (a) The spectral goodness-of-fit test  $p$ -value for the power law fit minus the  $p$ -value for the AR1 fit for linearly detrended AM SAT in the 2 y–45 y range, for the observational products. Positive values indicate a better fit for the power law representation. (b) As in Figure 8a, but for the 20c3m CMIP3 simulations. (c) As in Figure 8a, but for MM SAT in the 18 m–45 y range. (d) As in Figure 8b, but for MM SAT in the 18 m–45 y range.

In (5),  $\phi_{(T)} = 0$  when  $\phi = 0$ ,  $\phi_{(T)} \rightarrow 1^-$  as  $\phi \rightarrow 1^-$ , and  $\phi_{(T)} < \phi$  for  $0 < \phi < 1$ . The shape of  $\phi_{(T)}$  as a function of  $\phi$  is shown by the red curves in Figure 7. A similar equation has been derived by Kushnir *et al.* [2006] for monthly versus daily persistence of the North Atlantic Oscillation (NAO).

[26] By contrast, temporal aggregation asymptotically has no impact on a power law stochastic process. More precisely, for a second order self-similar process, toward which a power law stochastic process, including FAR(1,  $d$ ) described in Section 1, converges in distribution on long timescales [Cox, 1984; Taqqu, 2002]

$$\text{Power law : } \phi_{(T)} = \phi, \quad 0 \leq \phi \leq 1, \quad (6)$$

where  $\phi$  is the lag-one autocorrelation of this process. The one-to-one blue lines in Figure 7 represent this relationship.

[27] When observed lag-one autocorrelation coefficients from MM data are scattered against those from AM data (Figure 7a), about two thirds of the points fall between the power law (6) and AR1 (5) curves. Thus, observed persistence is generally greater than would be expected from AR1 and less than would be expected from a pure power law process. Kushnir *et al.* [2006] obtained a similar result for an NAO index on subannual timescales, and Vyushin *et al.* [2010] found that the interannual variability in stratospheric ozone in many regions behaves similarly. The same general behavior is captured in the 45 y 20c3m simulation segments (Figure 7b), with a more compact distribution

arising from the ensemble averaging effect described. The scatterplot relationship is similar for 45 y *picntrl* segments (not shown), suggesting that external forcing plays only a minor role here. The results are also similar for the subset of 6 climate models with 500-year-long *picntrl* simulations (Figure 7c). Finally, the relationship between  $\hat{\phi}$  for DM data and  $\hat{\phi}$  for AM data (Figure 7d) is similar, although the overall magnitude of the correlations is reduced.

[28] In all four panels of Figure 7 most of the points lie below the blue line and above the red line (the exact percentages of the points located between the two curves are given in the figure caption). This, and the robustness of the relationships to the presence or absence of external radiative forcing, suggests that the AR1 representation provides a lower bound and the power law representation an upper bound for climate persistence on interannual to multidecadal timescales.

[29] Using a spectral goodness-of-fit test [Milhoj, 1981; Beran, 1992], it is found that neither representation provides a better fit to the observed and simulated spectral density (Appendix B and Figure 8). A key point made in the Appendix is that the goodness-of-fit test must be applied to consistent time ranges, or misleading results can follow.

[30] Thus, neither the temporal aggregation analysis (Figure 7) nor the spectral goodness-of-fit test (see Appendix B and Figure 8) provides objective evidence that favors either the AR1 or power law representation over the interannual to multidecadal timescale range. Instead, persistence of climate

variability, which is well characterized as internal climate variability, falls between the two statistical representations; neither representation provides a complete description.

## 5. Conclusions

[31] This article has set out to better characterize climate variability by comparing two representations of its temporal statistics that exhibit contrasting behavior at low frequencies. The analysis has focused on the interannual to multi-decadal scale that is expected to be reasonably well captured by current generation climate models. Two measures of persistence arising from these representations, the lag-one autocorrelation  $\phi$  and the spectral slope parameter, or the Hurst exponent,  $H$ , are generally credibly represented in climate models and are relatively insensitive to the presence of external radiative forcing for detrended data. Less confidence can be placed in observations and models for tropical data. Climate persistence appears to lie between short-memory (AR1) and long-memory (power law) processes, and the data does not suggest that one representation is superior.

[32] It might be argued that the two representations focused on here are too simple. We justify the exclusive focus on these two representations because (1) they are equivalently parsimonious, in the sense that each involves a single shape parameter and single parameter measuring overall power over the frequency band analyzed; and (2) they are extensively used in several applications in statistical climatology [e.g., Trenberth et al., 2007; World Meteorological Organization, 2007; Caballero et al., 2002; Vyushin et al., 2007, 2010; Franzke, 2012] and (3) they can be unified within the generalized FAR(1,  $d$ ) class of models that include both short and long memory behavior as demonstrated in the Introduction, and through the fraction of predictable variance at the one time increment lead. The fact that both provide related information does not imply, however, that they are equivalent in all respects.

[33] A key difference between a power law stochastic process and an autoregressive process of any finite order is that the spectral density of the former increases without bound as a power law near the origin (i.e. for  $|f| \ll 1/2$ ), whereas the spectral density of the latter saturates to a constant near the origin. Because the lowest frequency captured is controlled ultimately by the length of a given time series, for many applications, e.g. trend detection, it is necessary to make an assumption about spectral behavior near the origin. Autoregressive and power law stochastic processes provide two extreme cases of such assumption.

[34] Bloomfield and Nychka [1992] compared estimated linear trend confidence intervals based on white noise, AR1, AR2, AR8, and power law representations for globally averaged annual mean surface air temperature anomalies time series. They showed that the estimates based on the AR8 and power law representations are close to each other, about four times greater than the white noise based confidence interval, and approximately 70% greater than the AR1 and AR2 estimates. Vyushin et al. [2007] have found that the uncertainty of the long-term total ozone trend in Northern Hemisphere middle and high latitudes attributable to the anthropogenic chlorine estimated using a power law representation is about 50% greater than the corresponding AR1 estimate. Based on the results of this article we expect analogous differences in

temperature trend confidence interval to arise in persistent regions such as extratropical ocean basins.

## Appendix A: Spectral Representation Theorem

[35] This Appendix outlines some of the statistical concepts used in this analysis. For any real-valued discrete stationary process,  $X_t$ , with zero mean there exists an *orthogonal process*  $Z_X(f)$ , defined on the interval  $[-1/2, 1/2]$ , such that

$$X_t = \int_{-1/2}^{1/2} \exp(i2\pi ft) dZ_X(f). \quad (\text{A1})$$

The orthogonal process,  $Z_X(f)$ , has *orthogonal increments*  $dZ_X(f)$ . That is, if  $[f, f + df]$  and  $[f', f' + df']$  are nonintersecting subintervals of  $[-1/2, 1/2]$  then the increments satisfy

$$\text{cov}\{dZ_X(f), dZ_X(f')\} = E\{dZ_X^*(f') dZ_X(f)\} = 0, \quad (\text{A2})$$

where  $E\{Z\}$  is the expectation operator applied to the random variable  $Z$ , and

$$E\left\{|dZ_X(f)|^2\right\} = dS_X^{(I)}(f), \quad (\text{A3})$$

where  $S_X^{(I)}(f)$  is a bounded nondecreasing function called the *integrated spectrum*. In the case  $S_X^{(I)}(f)$  is differentiable everywhere on  $[-1/2, 1/2]$  and such that

$$dS_X^{(I)}(f) = S_X(f)df, \quad (\text{A4})$$

where  $S_X(f)$  is called the *spectral density function*, we have the following spectral representation of the *autocovariance function*  $\gamma_X(\tau)$  of  $X$ ,

$$\gamma_X(\tau) = E\{X_t X_{t+\tau}\} = \int_{-1/2}^{1/2} \exp(i2\pi f\tau) S_X(f) df. \quad (\text{A5})$$

This expression relates persistence captured by the autocovariance function to the statistics of variability across timescales captured by the spectral density. The variance of the discrete stationary process  $X_t$  can be represented as

$$\sigma_X^2 = \gamma_X(0) = \int_{-1/2}^{1/2} S_X(f) df. \quad (\text{A6})$$

Thus spectral density provides a spectral decomposition of the total variability of a discrete stationary process.

## Appendix B: Goodness of Fit Test

[36] A spectral goodness-of-fit test [Milhoj, 1981; Beran, 1992] is applied to the AR1 and power law fits of linearly detrended observed and simulated data. In this application, spectral domain fits (maximum likelihood estimation for AR1, GSPE for power law, as described in Section 2) are used instead of time domain fits. GSPE for the power law fit has the added advantage of using all available frequencies, while DFA3 has a short timescale cutoff of 18 time units [Kantelhardt et al., 2001; Vyushin and Kushner, 2009]. Figures 8a and 8b show the difference of  $p$  values for the 2 y–45 y fit for observed and simulated AM data, after an ensemble mean has been taking across all available data. For

observations and simulations this difference is small, except for observed temperature in northern Siberia, where the power law representation better fits the data. The latter feature is perhaps related to the anomalous behavior for the AR1 fit to AM data in this region (Section 3.1). Thus, for the most part, the two representations demonstrate similar performance. A similar conclusion is reached for other time-scale ranges and integrations, including the 10 y–100 y range for AM data in the 500 year *picntr* integrations (not shown).

[37] This conclusion might not hold if inconsistent timescales are used. For example, fitting AR1 to MM data in the 2 m–45 y range and power law in the 18 m–45 y range yields a better fit for the power law representation over most of the extratropical oceans and a worse fit for the power law representation in the tropics, especially for observations (Figures 8c and 8d). These differences in performance reflect the inconsistent high frequency cutoff in the two fits. This example is relevant because these are typical timescale ranges chosen for the representations in standard practice.

[38] **Acknowledgments.** This research has been supported by the Canadian Foundation for Climate and Atmospheric Sciences, the Natural Sciences and Engineering Research Council, and the Canadian Meteorological and Oceanographic Society. We thank the “R Foundation for Statistical Computing” for the *R environment* and David Pierce for the *ncdf* package. We acknowledge the modeling groups, the Program for Climate Model Diagnosis and Intercomparison (PCMDI) and the WCRP’s Working Group on Coupled Modeling (WGCM) for their roles in making available the WCRP CMIP3 multimodel data set. Support of this data set is provided by the Office of Science, U.S. Department of Energy. We are grateful to Alexander Korobov, David Stephenson, and Nicholas Watkins for fruitful discussions.

## References

- Beran, J. (1992), A goodness-of-fit test for time series with long range dependence, *J. R. Stat. Soc. B*, **54**, 749–760.
- Beran, J. (1994), *Statistics for Long-Memory Processes*, 315 pp., CRC, Boca Raton, Fla.
- Blender, R., and K. Fraedrich (2003), Long time memory in global warming simulations, *Geophys. Res. Lett.*, **30**(14), 1769, doi:10.1029/2003GL017666.
- Blender, R., K. Fraedrich, and B. Hunt (2006), Millennial climate variability: GCM-simulation and Greenland ice cores, *Geophys. Res. Lett.*, **33**, L04710, doi:10.1029/2005GL024919.
- Bloomfield, P. (1992), Trends in global temperature, *Clim. Change*, **21**, 1–16.
- Bloomfield, P., and D. Nychka (1992), Climate spectra and detecting climate change, *Clim. Change*, **21**, 275–287.
- Brockwell, P. J., and R. A. Davis (1998), *Time Series: Theory and Methods*, 2nd ed., 600 pp., Springer, New York.
- Caballero, R., S. Jewson, and A. Brix (2002), Long memory in surface air temperature: detection, modeling, and application to weather derivative valuation, *Clim. Res.*, **21**, 127–140.
- Cox, D. R. (1984), Long-range dependence: A review, in *Statistics: An Appraisal. Proceedings 50th Anniversary Conference Iowa State Statistical Laboratory*, edited by H. A. David and H. T. David, pp. 55–74, Iowa State Univ. Press, Ames.
- Eichner, J. F., E. Koscielny-Bunde, A. Bunde, S. Havlin, and H. J. Schellnhuber (2003), Power-law persistence and trends in the atmosphere: A detailed study of long temperature records, *Phys. Rev. E*, **68**, 046133, doi:10.1103/PhysRevE.68.046133.
- Feldstein, S. B. (2000), The timescale, power spectra, and climate noise properties of teleconnection patterns, *J. Clim.*, **13**, 4430–4440.
- 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.
- Frankignoul, C., and K. Hasselmann (1977), Stochastic climate models, Part II: Application to sea-surface temperature anomalies and thermocline variability, *Tellus*, **29**, 289–305.
- Franzke, C. (2010), Long-range dependence and climate noise characteristics of Antarctic temperature data, *J. Clim.*, **23**, 6074–6081.
- Franzke, C. (2012), Nonlinear trends, long-range dependence and climate noise properties of surface temperature, *J. Clim.*, **25**, 4172–4183.
- Hansen, J., R. Ruedy, J. Glascoe, and M. Sato (1999), GISS analysis of surface temperature change, *J. Geophys. Res.*, **104**, 30,997–31,022.
- Hasselmann, K. (1976), Stochastic climate models, part 1: Theory, *Tellus*, **28**, 473–485.
- Hurst, H. E. (1951), Long-term storage capacity of reservoirs, *Trans. Am. Soc. Civ. Eng.*, **116**, 770–799.
- Huybers, P., and W. Curry (2006), Links between annual, Milankovitch, and continuum temperature variability, *Nature*, **441**, 329–332, doi:10.1038/nature04745.
- Kalnay, E., et al. (1996), The NCEP/NCAR 40-year reanalysis project, *Bull. Am. Meteorol. Soc.*, **77**, 437–471.
- Kantelhardt, J. W., E. Koscielny-Bunde, H. H. A. Rego, S. Havlin, and A. Bunde (2001), Detecting long-range correlations with detrended fluctuation analysis, *Physica A*, **295**, 441–454.
- Kushnir, Y., W. A. Robinson, P. Chang, and A. W. Robertson (2006), The physical basis for predicting Atlantic sector seasonal-to-interannual climate variability, *J. Clim.*, **19**, 5949–5970, doi:10.1175/JCLI3943.1.
- Manabe, S., and R. J. Stouffer (1996), Low-frequency variability of surface air temperature in a 1000-year integration of a coupled atmosphere-ocean-land surface model, *J. Clim.*, **9**, 376–393.
- Milhoj, A. (1981), A test of fit in time series models, *Biometrika*, **68**, 177–188.
- Pelletier, J. D. (1997), Analysis and modeling of natural variability of climate, *J. Clim.*, **10**, 1331–1342.
- Percival, D. B., J. E. Overland, and H. O. Mofjeld (2001), Interpretation of North Pacific variability as a short- and long-memory process, *J. Clim.*, **14**, 4545–4559.
- Randall, D. A., et al. (2007), Climate models and their evaluation, in *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by S. Solomon et al., pp. 589–662, Cambridge Univ. Press, Cambridge, U. K.
- Robinson, P. M. (1995), Gaussian estimation of long range dependence, *Ann. Stat.*, **23**, 1630–1661.
- Rybksi, D., A. Bunde, and H. von Storch (2008), Long-term memory in 1000-year simulated temperature records, *J. Geophys. Res.*, **113**, D02106, doi:10.1029/2007JD008568.
- Stephenson, D. B., V. Pavan, and R. Bojariu (2000), Is the North Atlantic Oscillation a random walk?, *Int. J. Climatol.*, **20**, 1–18.
- Taqqu, M. S. (2002), Fractional Brownian motion and long-range dependence, in *Theory and Applications of Long-Range Dependence*, edited by M. T. P. Doukhan and G. Oppenheim, pp. 5–38, Birkhauser, Boston.
- Thompson, D. W. J., M. P. Baldwin, and J. M. Wallace (2002), Stratospheric connection to Northern Hemisphere wintertime weather: Implications for prediction, *J. Clim.*, **15**, 1421–1428.
- Trenberth, K. E., et al. (2007), Observations: Surface and atmospheric climate change, in *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by S. Solomon et al., pp. 235–335, Cambridge Univ. Press, Cambridge, U. K.
- Tsonis, A. A., P. J. Roebber, and J. B. Elsner (1999), Long-range correlations in the extratropical atmospheric circulation: origins and implications, *J. Clim.*, **12**, 1534–1541.
- Uppala, S. M., et al. (2005), The ERA-40 re-analysis, *Q. J. R. Meteorol. Soc.*, **131**, 2961–3012, doi:10.1256/qj.04.176.
- von Storch, H., and F. W. Zwiers (1999), *Statistical Analysis in Climate Research*, 494 pp., Cambridge Univ. Press, New York.
- Vyushin, D. I. (2009), Statistical approximation of natural climate variability, PhD thesis, Univ. of Toronto, Toronto, Ont., Canada.
- Vyushin, D. I., and P. J. Kushner (2009), Power-law and long-memory characteristics of the atmospheric general circulation, *J. Clim.*, **22**, 2890–2904, doi:10.1175/2008JCLI2528.1.
- Vyushin, D. I., 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.
- Vyushin, D. I., 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.
- Vyushin, D. I., P. J. Kushner, and J. Mayer (2009), On the origins of temporal power-law behavior in the global atmospheric circulation, *Geophys. Res. Lett.*, **36**, L14706, doi:10.1029/2009GL038771.
- Vyushin, D. I., T. G. Shepherd, and V. E. Fioletov (2010), On the statistical modeling of persistence in total ozone anomalies, *J. Geophys. Res.*, **115**, D16306, doi:10.1029/2009JD013105.
- World Meteorological Organization (2007), Scientific assessment of ozone depletion: 2006, Rep. 50, Global Ozone Res. Monit. Proj., Geneva, Switzerland.