Testing the Annular Mode Autocorrelation Time Scale in Simple Atmospheric General Circulation Models

EDWIN P. GERBER, SERGEY VORONIN, AND LORENZO M. POLVANI

Department of Applied Physics and Applied Mathematics, Columbia University, New York, New York

ABSTRACT

A new diagnostic for measuring the ability of atmospheric models to reproduce realistic low-frequency variability is introduced in the context of Held and Suarez's 1994 proposal for comparing the dynamics of different general circulation models. A simple procedure to compute τ, the *e*-folding time scale of the annular mode autocorrelation function, is presented. This quantity concisely quantifies the strength of low-frequency variability in a model and is easy to compute in practice. The sensitivity of τ to model numerics is then studied for two dry primitive equation models driven with the Held–Suarez forcings: one pseudospectral and the other finite volume. For both models, τ is found to be unrealistically large when the horizontal resolutions are low, such as those that are often used in studies in which long integrations are needed to analyze model variability on low frequencies. More surprising is that it is found that, for the pseudospectral model, τ is particularly sensitive to vertical resolution, especially with a triangular truncation at wavenumber 42 (a very common resolution choice). At sufficiently high resolution, the annular mode autocorrelation time scale τ in both models appears to converge around values of 20–25 days, suggesting the existence of an intrinsic time scale at which the extratropical jet vacillates in the Held and Suarez system. The importance of τ for computing the correct response of a model to climate change is explicitly demonstrated by perturbing the pseudospectral model with simple torques. The amplitude of the model's response to external forcing increases as τ increases, as suggested by the fluctuation–dissipation theorem.

1. Introduction

Held and Suarez (1994, hereinafter HS94), established a benchmark for comparing the numerical schemes of different dynamical cores, general circulation models (GCMs) that integrate the primitive equations with idealized physics. They proposed a simple set of forcings that produce a realistic climate without complex parameterizations, allowing a comparison of the dynamical fidelity of GCMs independently of differences in their radiation, convection, and boundary layer schemes. HS94 suggested that two key diagnostics be computed: the time and zonal mean zonal winds, and the time and zonal mean temperature variance. Together, these diagnostics ensure that a core's dynamics produce a realistic 1) mean climate and 2) synoptic eddies. In this paper, we propose a new diagnostic for the

HS94 benchmark, one that characterizes the variability of a model on lower frequencies. This new diagnostic establishes the fitness of a GCM for the very long integrations that are typically needed for studies of low-frequency variability.

The computational efficiency of GCMs with the HS94 forcing, or other similar idealized forcings, has made them an attractive tool for dynamical studies of atmospheric variability on intraseasonal and interannual time scales. They have been used to explore stratospheric–tropospheric interactions (Taguchi et al. 2001; Polvani and Kushner 2002; Taguchi and Yoden 2002a,b; Taguchi 2003a,b; Song and Robinson 2004; Kushner and Polvani 2004, 2005, 2006; Eichelberger and Hartmann 2005), and to understand the behavior of the North Atlantic Oscillation, annular modes, and zonal index (e.g., Robinson 1996; Franzke et al. 2004; Son et al. 2008). In addition, a well-defined forcing with less reliance on subgrid-scale parameterizations enables greater control over the climatology. Theories can be tested over a wide range of parameter settings (e.g., Schneider 2004; Walker and Schneider 2006), a strategy

Corresponding author address: Edwin P. Gerber, Dept. of Applied Physics and Applied Math, Columbia University, 200 S. W. Mudd Building, MC 4701, 500 W. 120th St., New York, NY 10027.
E-mail: epg2108@columbia.edu

that cannot be implemented with more comprehensive GCMs without retuning parameterizations.

In a recent study with a model nearly identical to the pseudospectral dynamical core used by HS94, Gerber and Vallis (2007) noted a sensitivity of the low-frequency variability to the model's resolution. The model, however, consistently produced a realistic climate by the standards of the HS94 diagnostic. This sensitivity poses a problem for the interpretation of studies with idealized GCMs, particularly if the processes in question depend on the nature of the low-frequency variability. Paradoxically, Gerber and Vallis (2007) observed a *worsening* of their model's ability to capture realistic low-frequency variability with an *increase* in the vertical resolution, also raising questions concerning numerical convergence. Boer and Denis (1997) proposed a test for the numerical convergence of a dynamical core in terms of the mean flow and synoptic variability, but the issue of convergence has not been investigated in the context of low-frequency variability.

To address these concerns, in this paper we first propose and then apply a simple test designed to assess a model's ability to produce realistic low-frequency variability. As shown in section 2, the test involves computation of a single number τ, the autocorrelation time scale of the annular mode. The procedure for computing τ is described in section 2a. In section 3, we investigate the dependence of τ on model numerics using two dynamical cores, one pseudospectral and the other finite volume, forced with the HS94 scheme. From a theoretical perspective, we ask the following question: What is the actual autocorrelation time scale for the physical system described by the dry primitive equations driven by the HS94 forcing, and how does this time scale compare to the one observed in the atmosphere? From a practical standpoint, we wish to determine what resolution is sufficient for these models to be used to study low-frequency variability, and to provide a benchmark for comparison with other dynamical cores. In section 4, we illustrate a potential consequence of poorly simulated low-frequency variability: erroneous response to climate change. Our conclusions are found in section 5.

2. The autocorrelation time scale of the annular mode

We seek a simple diagnostic to determine whether a model exhibits realistic low-frequency variability. By low-frequency, we mean intraseasonal, that is, time scales greater than 10 days, and so longer than the periods associated with baroclinic instability, but generally shorter than those associated with the seasonal cycle and ocean processes, which are often not represented in simple GCMs. On these time scales, the leading patterns of variability in the extratropics are the well-known annular modes (Thompson and Wallace 2000). The northern and southern annular modes are defined as the first empirical orthogonal functions (EOFs) of sea level pressure or geopotential height in the Northern and Southern Hemispheres, respectively. The observed annular mode patterns reveal a robust hemispheric-scale dipole in pressure, or mass, between the high and midlatitudes that, by geostrophy, characterize vacillations in the position of the extratropical jets. In this respect the annular modes are related to the zonal index, a measure of the zonal circulation first investigated by Rossby (1939), and refined by Namias (1950). Given near-geostrophic balance on large spatial scales, there is a very high degree of correlation between the time series associated with the annular mode patterns and the zonal index, especially in idealized models. While other, more regional patterns of variability, such as the North Atlantic Oscillation, are also prominent in the atmosphere, Feldstein (2000b) shows that the time scale tends to increase with the zonal scale of the pattern: the broader the pattern in longitude, the slower the time scale. Hence a characterization of the annular mode time scale establishes an upper bound on the low-frequency variability in a model.

For our test, we define the annular mode to be the first EOF of the zonally averaged surface[1] pressure $\overline{p_s}$, corrected for latitude bias as in North et al. (1982). The advantage of this variable is that, in the absence of topography, it requires no vertical interpolation from the actual model levels, as would be the case if, for instance, the 250-hPa zonal winds were used. Thompson and Wallace (2000) show that patterns identified from zonally averaged variables are equivalent to those identified from the two-dimensional latitude–longitude fields, so that little information with respect to the annular mode is lost by first averaging p_s around the latitude circle. The principal component time series associated with the annular mode EOF pattern is denoted the "annular mode index." We quantify the behavior of the annular mode by computing τ, the *e*-folding time scale of the annular mode index's autocorrelation function. As seen in Fig. 1, such autocorrelation functions are not perfectly exponential, but the *e*-folding time scale provides a fair estimate for comparison.

It is worth emphasizing that τ is not an artificial statistical construction: its value has physical significance

---

[1] If the model includes topography, zonally averaged sea level pressure should be used instead of $\overline{p_s}$.
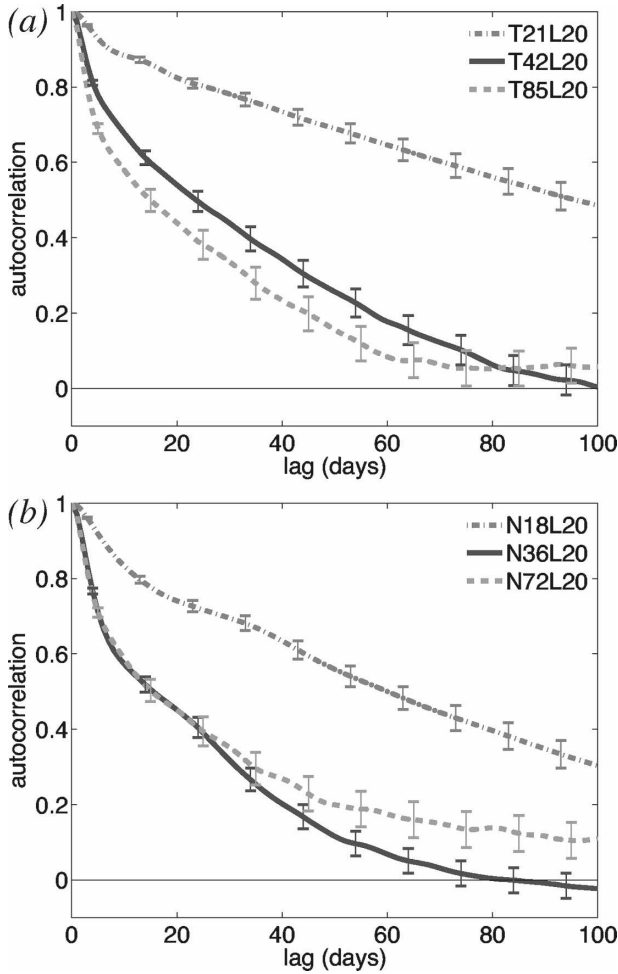
FIG. 1. The autocorrelation functions for the (a) pseudospectral and (b) finite-volume dynamical cores as a function of lag for different horizontal resolutions. Error bars provide one-std-dev bounds on the autocorrelation function. The vertical resolution is the same in all integrations: 20 evenly spaced σ levels.

both for medium-range weather forecasting and the interpretation of interannual variability. The observed value of τ varies roughly between 10 and 20 days, depending on the season (Baldwin et al. 2003). This reflects the fact that large-scale perturbations of the jet stream tend to persist longer than typical synoptic events. As articulated by Feldstein (2000a), the likelihood of extreme deviations of the climate on interannual time scales due simply to "climate noise" on intraseasonal time scales increases significantly with the value of τ. In the HS94 system, τ is one of two important time scales generated spontaneously[2] by the sys-

tem's internal dynamics, the other being the eddy turnover time scale, which describes the synoptic variability associated with baroclinic eddies. Therefore, as part of model validation, it is important to determine whether the HS94 system of equations, solved with a specific numerical scheme running at a particular resolution, generates a value of τ is comparable to the one obtained from observations.

a. Procedure for computing τ

For the sake of clarity and reproducibility, we start by spelling out a simple procedure for computing the annular mode autocorrelation time scale. Given a finite integration of length N days, we can only estimate the "true" autocorrelation time scale τ by and empirical estimate, which we denote τ_N. Clearly, as N increases, the empirical estimate τ_N will converge to τ. We discuss the rate of convergence, and how to estimate the uncertainty as a function of N in the next subsection. The procedure is as follows:

1) Calculate the daily zonal mean surface pressure $\overline{p}_s(t, \phi)$ (or sea level pressure, if the model has topography) from equator to pole for each hemisphere, separately.
2) Weight the pressure field to account for the latitude bias, $\overline{p}_s^w(t, \phi) = \sqrt{\cos\phi}\, \overline{p}_s(t, \phi)$.
3) Compute[3] the first principal component time series, $\alpha_1(t)$, of the weighted surface pressure field $\overline{p}_s^w(t, \phi)$. This is the annular mode index of the model.
4) Calculate the autocorrelation function $r_N(t)$ of the principal component time series $\alpha_1(t)$. If the forcing is hemispherically symmetric, as in the standard HS94 scheme, average the functions $r_N(t)$ from both hemispheres to reduce the uncertainty.
5) Compute the estimate τ_N of the autocorrelation time scale τ by fitting $r_N(t)$ to a simple exponential. Specifically, τ_N is chosen so as to minimize the RMS distance between $r_N(t)$ and e^{-t/τ_N} for all t such that $r_N(t) > e^{-1}$. This focuses the fit on the intraseasonal time scales, and not ultralow frequencies characterized by the tail of the autocorrelation function.

b. Estimating the uncertainty in τ_N

An important practical question needs be addressed at this point: How long must a model be integrated to obtain an accurate estimate of τ? To answer this question, even approximately, we must assume that the annular mode index can be modeled as an autoregressive

---

[2] This is in contrast to the thermal and momentum damping time scales of the model, which are prescribed as part of the HS94 forcing.

[3] See von Storch and Zwiers (1999, chapter 13) for a complete discussion of EOF analysis.

(AR) process of first order (AR-1) or, in the continuous limit, an Ornstein–Uhlenbeck process.[4] Error estimates for the autocorrelation function of simple statistical processes were obtained by Bartlett (1946). In particular, the persistence of an AR-1 process is fully determined by the single parameter τ, which represents the decay time scale of the process' autocorrelation function $r(t) = \exp(-t/\tau)$. For simplicity, we take the time t to increment at unit intervals, specifically, days. Given a time series of length N generated by a process with unknown autocorrelation time scale τ, we estimate the true autocorrelation function $r(t)$ by computing the empirical autocorrelation function, $r_N(t)$. Bartlett (1946) shows that the variance of $r_N(t)$ about $r(t)$ is a function of N, τ, and the lag t,

$$\text{var}[r_N(t)] = \frac{1}{N}\left\{\frac{[1 + \exp(-2/\tau)][1 - \exp(-2t/\tau)]}{1 - \exp(-2/\tau)}\right.$$
$$\left. - 2t\exp(-2t/\tau)\right\}. \qquad (2.1)$$

At time $t = 0$ the uncertainty is zero, as the autocorrelation function always begins at unity, while for large lags, $t \gg \tau$, the variance approaches τ/N. For our purposes, we need to know the uncertainty of τ_N, the best-fit e-folding time scale of $r_N(t)$. As shown in the appendix, (2.1) can be used to estimate the magnitude of the uncertainty using the standard deviation

$$\text{std}(\tau_N) \approx k\tau^{3/2}N^{-1/2}, \qquad (2.2)$$

where $k \approx e(1 - 3e^{-2})^{(1/2)} \approx 2$. Since the distribution of both $r_N(t)$ and τ_N are non-Gaussian, we use the standard deviation as a rough measure of the uncertainty.

The fact that the uncertainty of τ_N depends on the value of τ itself makes it difficult to set an appropriate integration length a priori. To achieve a given level of accuracy, we must first execute a short integration to estimate τ and then use (2.2) to predict the necessary run time N. If we define accuracy as the relative size of the standard deviation to the value of τ itself, $\sigma_N = \text{std}(\tau_N)/\tau$, the appropriate integration time is roughly

$$N \approx 4\tau\sigma_N^{-2}. \qquad (2.3)$$

For example, if τ is 10 days, roughly 4000 days of integration are needed to reduce the standard deviation of τ down to 1 day, while if $\tau = 25$, roughly 10 000 days of

---

[4] It is not possible, in general, to make an error estimate without some knowledge of the process in question. Bartlett provides a more general formula for an AR-n process, but from our experience the annular mode autocorrelation functions are reasonably well described by an AR-1 process for the purposes of uncertainty estimates.

integration are needed to get the same accuracy, corresponding a standard deviation of 2.5 days. The dependence of the uncertainty on τ/N is related to the effective degrees of freedom, discussed in the climate literature by Oort and Yienger (1996). If an event decorrelates on a time scale of τ days, one needs $n\tau$ days of integration to observe n independent events.

Equation (2.2) provides a very rough estimate of the uncertainty in τ_N, and was derived primarily to estimate the dependence of the uncertainty on τ and N. For the error estimates presented in section 3, we use (2.1) to compute $\text{var}[r_N(t)]$. Upper and lower bounds on τ_N are then found by fitting exponentials to $r_N(t) \pm \text{std}[r_N(t)]$, respectively. The uncertainty can be reduced further if the model's two hemispheres are statistically equivalent. Provided the run time $N \gg \tau$, the autocorrelation functions from the two hemispheres can be averaged together, and the value $2N$ used when computing the uncertainty.

Based on results in section 3, the autocorrelation time scale of the annular mode in the HS94 system is approximately 25 days. As a rule of thumb, 2000–3000 days of integration is sufficient to estimate τ in this range to $\pm 20\%$ accuracy, especially if both hemispheres are statistically the same. If the value of τ in the model is larger, a longer integration would, of course, be needed to accurately compute it. An integration of this length, however, is sufficient to determine whether the τ of model is near 20–30 days, or whether it is not.

3. Case study: A comparison between pseudospectral and finite-volume dynamical cores

a. Experimental setup

To illustrate the importance of testing the low-frequency variability in simple GCMs, we compare two idealized general circulation models developed by the Geophysical Fluid Dynamics Laboratory (GFDL). These models use very different numerical schemes, one pseudospectral and the other finite volume, to solve the dry primitive equations on evenly spaced $\sigma = p/p_s$ levels. Both are forced as specified in HS94, and differ only in their numerics. In brief, the HS94 forcing is a simple scheme designed to produce a reasonable climate without radiative, convective, or boundary layer parameterizations. Temperature is linearly relaxed to specified profile, a state close to radiative–convective equilibrium. The relaxation rate is 40 days everywhere but the lower-level tropics, where it is increased to 4 days to create a more realistic Hadley circulation. Momentum is removed by a linear drag with a

TABLE 1. The integration length of the resolution sweep experiments, for the pseudospectral and finite-volume cores. The models are initialized from a state of rest, and given 500 days to spin up before data are sampled.

		Horizontal (pseudospectral)			Horizontal (finite volume)		
		T21 64 × 32	T42 128 × 64	T85 256 × 128	N18 60 × 36	N36 120 × 72	N72 240 × 144
Resolution							
Vertical	10	20 000	10 000	5000	20 000	10 000	5000
	20	20 000	10 000	5000	20 000	10 000	5000
	40	20 000	10 000	5000	20 000	10 000	5000
	80	20 000	10 000	5000	20 000	10 000	5000

1-day damping rate at the surface that decreases linearly in pressure to zero at 700 hPa.

The pseudospectral model is essentially the same as the one used in HS94. The truncation is triangular, and integrations are performed at three horizontal resolutions: T21, T42, and T85. These correspond to grids with 64 × 32, 128 × 64, and 256 × 128 points in longitude and latitude, respectively. For each of these horizontal resolutions, four integrations are completed with increasing vertical resolution, specifically with 10, 20, 40, and 80 evenly spaced σ levels. HS94 consider dissipation, necessary to remove enstrophy at smaller scales, as part of the numerics. Unless otherwise specified, we apply the same ∇^8 hyperdiffusion as used in their original test, which is adjusted with resolution so that the highest resolved wavenumber is damped with time scale 1/10 day.

The finite-volume model is described by Lin (2004). Since the finite-volume scheme is implicitly diffusive, the model requires no explicit dissipation. The code is most efficient when the ratio of the number of longitudinal to latitudinal grid points is approximately 1.6 (S. Lin 2006, personal communication). To achieve a meaningful comparison with the pseudospectral model, then, we choose grids that contain roughly equivalent net resolutions. Thus, we set

$$n_{\lambda,\text{fv}} \times n_{\phi,\text{fv}} \approx n_{\lambda,\text{ps}} \times n_{\phi,\text{ps}} \quad \text{and} \quad (3.1)$$

$$n_{\lambda,\text{fv}}/n_{\phi,\text{fv}} \approx 1.6, \quad (3.2)$$

where n_λ and n_ϕ refer to the number of longitudinal and latitudinal grid points and fv and ps refer to the finite-volume and pseudospectral codes, respectively. For illustration, consider the pseudospectral model with T21 horizontal resolution, which has a 64 × 32 grid. We compare this with the finite-volume core with a 60 × 36 grid (denoted N18 resolution)[5] to satisfy both con-

straints (3.1) and (3.2). The N36, with a 120 × 72 grid, and N72, with a 240 × 144 grid, finite-volume resolutions are comparable to the pseudospectral model at T42 and T85 resolution, respectively. Again, four vertical resolutions were tested at each horizontal resolution, with 10, 20, 40, or 80 evenly spaced σ levels.

A summary of the resolution sweep experiments is shown in Table 1. Two suites of 12 experiments, one set with each model, allow us to independently vary the vertical and horizontal resolution by a factor of 8 and 4, respectively. Integration times are set to minimize the uncertainty in the time-scale estimates, subject to available computational resources. In all cases, the model is allowed to spin up for 500 days before data are collected. Since τ happens to be quite large in both models at low horizontal resolution, extremely long integrations (20 000 days) are needed in those cases to accurately determine its value.

b. Results

A key result of this paper is that the annular mode autocorrelation time scale of the HS94 system is very sensitive to resolution for both the finite-volume and pseudospectral models. Figure 1 illustrates the autocorrelation function of the annular mode index for varying horizontal resolutions, but a fixed vertical grid of 20 equally spaced σ levels. A similar picture (not shown) is found with 10, 40, or 80 levels. At low horizontal resolution, the annular mode is extremely persistent, decorrelating on a time scale τ of approximately 140 days in the pseudospectral model and 80 in the finite-volume model. With higher resolution, the annular mode decorrelates on a time scale of 15–30 days in both models—much closer to the observed autocorrelation time scale of the annular modes.

The values of τ obtained from the resolution sweep experiments are summarized in Table 2. The associated uncertainties provide a rough estimate of a one-standard-deviation bound on the annular mode time scale. As marked by boldface in the table, we find a sharp distinction between integrations with unrealisti-

---

[5] We use a convention that quantifies the resolution of the finite-volume core using the number of grid points from the equator to the pole.

TABLE 2. The autocorrelation time scale τ of the annular mode, for the resolution sweep experiments with the pseudospectral and finite-volume models. The uncertainties provide one-std-dev bounds about the estimates. Boldface shows unrealistic values.

Resolution		Horizontal (pseudospectral)			Horizontal (finite volume)		
		T21	T42	T85	N18	N36	N72
Vertical	10	**73 ± 7**	19.5 ± 1.4	14.0 ± 1.5	**46.9 ± 3.3**	27.7 ± 2.5	18.5 ± 2.1
	20	**136 ± 15**	33.1 ± 3.2	22.0 ± 3.1	**84 ± 7**	22.5 ± 1.9	23.2 ± 3.4
	40	**562 ± 170**	**96 ± 15**	20.6 ± 2.9	**101 ± 9**	26.3 ± 2.3	27.0 ± 3.6
	80	**238 ± 50**	**100 ± 18**	19.0 ± 2.8	**106 ± 9**	27.0 ± 2.5	22.6 ± 3.0

cally long autocorrelation and those clustering about values from 20 to 30 days. Taking both models together, we conclude that at sufficient resolution, the HS94 system of PDEs exhibits an internal time scale of approximately 25 days, a time scale characterizing vacillations of the extratropical jet. The τ values associated with the L10 integrations in both models tend to cluster at somewhat lower values. Analysis of the standard HS94 diagnostics, the time and zonal average jet, and the temperature variance (not shown) suggest that the other elements of the climate may be degraded in these low vertical resolution integrations, particularly in the upper atmosphere.

We first consider the problems at low horizontal resolution. At T21 and N18, the models produce extremely long time scales for all vertical resolutions. Given that the annular mode depends critically on the eddy momentum fluxes (e.g., Vallis et al. 2004) and that the fluxes are not sufficiently resolved at these low resolutions (Held and Phillipps 1993), it is perhaps not so surprising that the integrations provide unrealistic values of τ. Once the horizontal resolution becomes sufficient to adequately resolve eddy processes, the time scales generally cluster about the more realistic value of 25 days. Held and Phillipps (1993) find that meridional resolution is most important in controlling the eddy fluxes. The asymmetric grid in the finite-volume core provides better meridional resolution than the comparable pseudospectral model. Consistent with this, the time scales and jet structure (not shown) are slightly better at N18 resolution than T21.

The potential for a problem with the annular model autocorrelation time scale at these resolutions could perhaps be detected with the standard HS94 diagnostics, as illustrated in Fig. 2 where the mean climatology and temperature variance between the three pseudospectral resolutions with 20 σ levels in the vertical are shown. The mean jet is not well formed at T21 resolution, with westerlies extending to the equator producing super rotation at height. The eddy temperature variance is quite low, suggesting a lack of synoptic eddy activity. T42 or better resolution is required to get

a realistic extratropical circulation. Similar differences are found between the N18, N36, and N72 finite-volume integrations (not shown).

There is a notable exception to the improved quality of the integrations as the resolution is increased: the T42 model with 40 or 80 vertical levels. As seen in Fig. 3, the pseudospectral model is very sensitive to the vertical resolution. The value of τ jumps from 30 to 100 days as the model levels are increased from 20 to 40 levels. Hence, *higher vertical resolution degrades the quality of the model's annular mode time scale.* The finite-volume core does not have this problem, exhibiting little sensitivity to vertical resolution at all three horizontal resolutions. The pseudospectral model is sensitive to vertical resolution at T21 resolution as well; the annular modes in the T21L40 and L80 integration persist on the order of years! The problem seems to disappear at T85 resolution, however, where the autocorrelation time scale of the annular mode is statistically indistinguishable for most vertical resolutions.

Unlike the problems at low horizontal resolution, this failure in the pseudospectral model could not easily be detected in the HS94 test. To illustrate this subtlety, we focus on four integrations, three with realistic annular mode autocorrelation time scales, T42L20, N36L20, and N36L40, and one with unrealistic variability, T42L40. The HS94 diagnostics for these integrations are plotted in Fig. 4. All the jets are of the same magnitude and have the same overall structure. The only difference lies in the position of the extratropical jet: in the T42L40 integration, the extratropical jet has shifted a few degrees toward the equator, merging with the subtropical jet. The eddy statistics, as quantified by the variance of the temperature, are also nearly identical in all integrations, except for the slight equatorward shift in T42L40. Given no a priori reason to suspect that a small equatorward shift of the extratropical jet will alter the low-frequency variability, this defect can only be detected by checking the annular mode time scale. This demonstrates the practical need for computing τ to validate a model's ability to realistically capture atmospheric low-frequency variability.
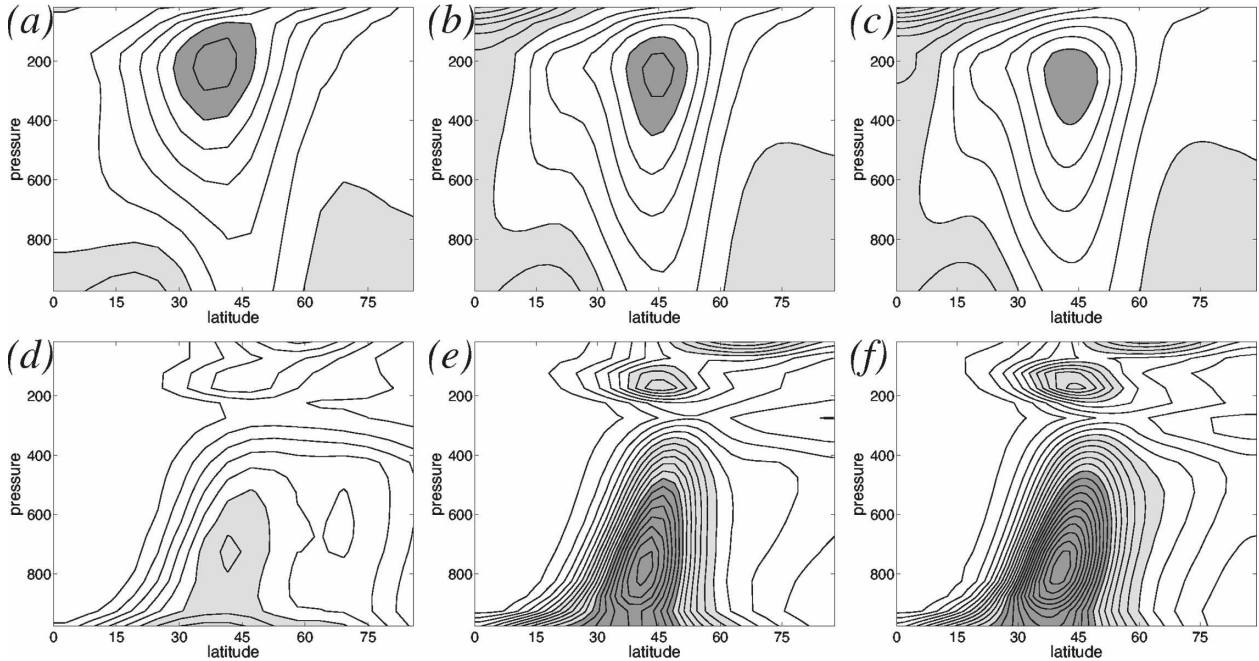
FIG. 2. The HS94 diagnostics for the T21L20, T42L20, and T85L20 integrations. (top) The time- and zonal-average zonal wind and (bottom) the time- and zonal-average of the eddy temperature variance are shown for (a), (d) T21L20; (b), (e) T42L20; and (c), (f) T85L20. In (a)–(c), the contour interval is 5 m s^{-1}; regions where the velocity is negative are lightly shaded, and regions where the velocity exceeds 25 m s^{-1} are darkly shaded. In (d)–(f) the contour interval is 2 K^2, regions above 12 K^2 are lightly shaded, and regions above 24 K^2 are darkly shaded.

c. Discussion

We consider whether the high sensitivity to vertical resolution in the pseudospectral model is related to the vertical advection scheme. To test this hypothesis, four additional T42 integrations of the model were completed, the first four listed in Table 3: two runs each using fourth-order finite-difference and finite-volume parabolic vertical advection; this is in contrast with the standard second-order scheme used in the calculations discussed above. As can be seen from the last column in Table 3, changing to the vertical advection scheme does not correct the problem. In both cases, the annular mode autocorrelation time scale increases substantially when the vertical resolution is raised. Relative to the second-order finite-difference integrations, fourth-order finite-difference advection increases τ at 20 vertical levels (33 → 43 days) and decreases it at 40 levels (96 → 91 days). The finite-volume scheme reduces the time scale at both resolutions, but there is still more than a doubling of the time scale with a doubling of the vertical resolution, from 21 to 46 days. These experiments indicate that τ is not well constrained at T42 horizontal resolution, even if the vertical resolution is fixed. With 20 levels, τ ranges from 21 to 43 days, depending on the choice of vertical advection, while with 40 levels it varies from 47 to 96 days. The dynamics

controlling the autocorrelation time scale are extremely sensitive to the choice of vertical advection at T42 horizontal resolution.

Lindzen and Fox-Rabinovitz (1989) suggest that the aspect ratio of the model grid, $\Delta z/\Delta\phi$, may be more important than the absolute vertical or horizontal resolution. For example, if the horizontal grid is fine in comparison to the vertical spacing, underresolved motions in the vertical may eventually corrupt the resolved motions in the horizontal, nullifying the impact of the added resolution. Indeed, both the T21 and N18 models perform better with lower vertical resolution, 10 levels in particular. Evidence that the aspect ratio matters in the pseudospectral model is also found in a short integration at T85 horizontal resolution with 160 evenly spaced vertical levels, the last described in Table 3. Utilizing the data from both hemispheres, we estimate τ to be 43 ± 15 days. Therefore we can say with some certainty that the time scale is greater than 22 days, the upper limit for the T85 simulations with fewer vertical levels; as at T42 resolution, increased vertical resolution has degraded the model's fidelity. It is worth noting that Lindzen and Fox-Rabinovitz (1989) found that most models were underresolved in the vertical. Our T85 simulations suggest that underresolving the vertical is not as problematic as overresolving it, at least in terms
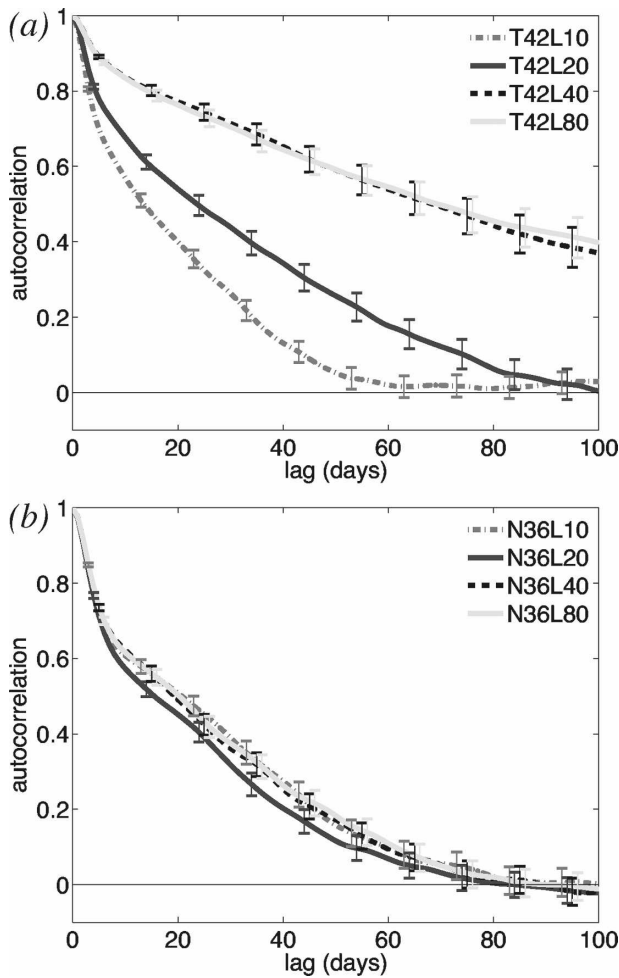
FIG. 3. Autocorrelation functions for the (a) pseudospectral and (b) finite-volume cores as a function of lag and vertical resolution at fixed horizontal resolution, T42 and N36.

of the low-frequency variability. This suggests that it may ultimately be a problem of diffusion. Increased horizontal resolution permits smaller-scale motions that might smooth out problems in the vertical.

The numerics of the finite-volume core are, by nature, more diffusive than those of the pseudospectral core, lending further support that a lack of diffusion may be the root of the problem in the spectral core. To test this hypothesis, we experiment with a broader hyperdiffusion in the pseudospectral model, ∇^4, again tuning the coefficient so that the diffusion time on the highest resolved wavenumber is 1/10 day. It should be noted that increased diffusion could also push the model in the wrong direction; greater diffusion decreases the effective horizontal resolution, and we know that the model performs extremely poorly at T21. The ∇^4 integrations, however, produce a realistic climatology and eddy statistics by the HS94 benchmark (not shown), and exhibit improved annular mode behavior, as seen in Fig. 5. While the annular mode autocorrelation time scale varies considerably with vertical resolution, from 27 to 39 days, it is not the threefold increase observed with ∇^8 diffusion. A more appropriate change to the model may be to add a vertical diffusion, and is the subject of present work.

Gerber and Vallis (2007) and Son et al. (2008) independently noted that, as the HS94 forcing is modified in various ways, an increase in τ is accompanied by an equatorward shift of the jet (and the associated synoptic eddies). Gerber (2005) found that these properties were also correlated when just the resolution alone is altered. For the runs discussed above, the relationship between τ and the jet location is explored in Fig. 6,
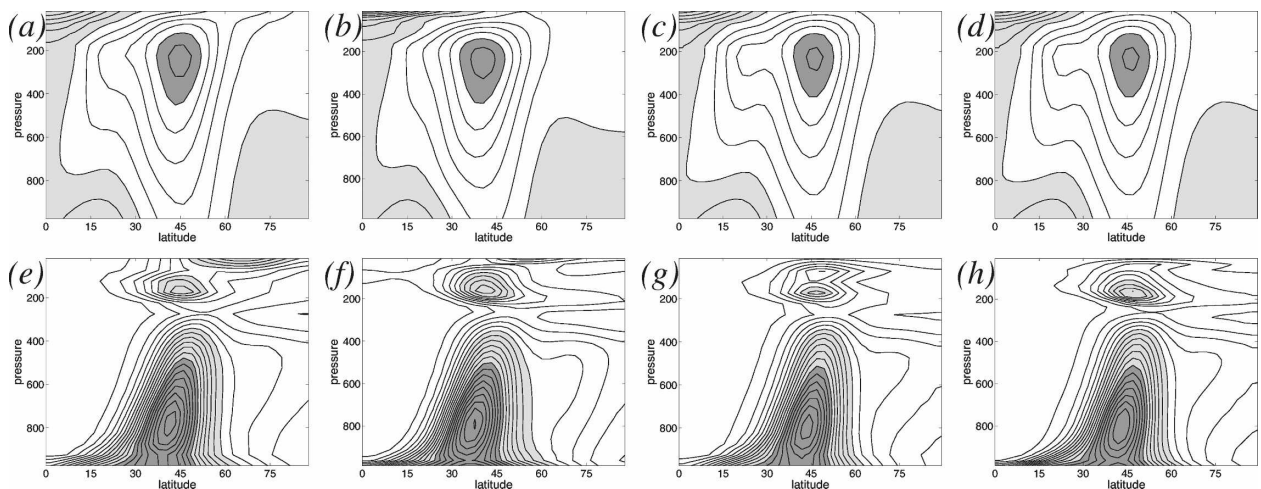

FIG. 4. As in Fig. 2, but for integrations (a), (e) T42L20; (b), (f) T42L40; (c), (g) N36L20; and (d), (h) N36L40. All integrations exhibit an annular mode with time scale τ of approximately 20–30 days, except T42L40, where the jet vacillates on a time scale of approximately 100 days.

TABLE 3. Integration details and annular mode autocorrelation time scale τ for additional pseudospectral dynamical core experiments. Length refers to length of the integration, after a 500-day spinup period, with the exception of the T85L160 run, where we only allowed 300 days of spinup. Estimates of τ are less constrained on the upper bound than the lower bound, particularly when the uncertainty is large. Hence, for T85L160 we provide more detailed error bounds.

Integration	Horizontal resolution	Vertical resolution	Vertical advection	Diffusion order	Length (days)	τ (days)
Fourth-order-20	T42	20	Fourth-order finite difference	∇^8	10 000	43.0 ± 4.7
Fourth-order-40	T42	40	Fourth-order finite difference	∇^8	10 000	91 ± 13
fvp-20	T42	20	Finite-volume parabolic	∇^8	10 000	20.6 ± 1.7
fvp-40	T42	40	Finite-volume parabolic	∇^8	10 000	46.4 ± 5.6
∇^4-20	T42	20	Second-order finite difference	∇^4	10 000	27.1 ± 2.2
∇^4-40	T42	40	Second-order finite difference	∇^4	10 000	37.2 ± 4.4
T85L160	T85	160	Second-order finite difference	∇^8	1300	43 ± 15

where the time scale of the annular mode is plotted against the latitude of extratropical jet. Each connected cluster of points mark a set of simulations with identical horizontal resolutions (and different model levels). Note how, remarkably, all three of the sets corresponding to T42 pseudospectral integrations (see the circles in Fig. 6) show a similar slope of approximately −12 days degree^{-1}, that is, a 1° poleward shift in the jet is associated with a 12-day decrease in τ. This relationship, however, does not hold for *any* other set of simulations, suggesting that is an artifact of T42 resolution, which has many deficiencies, as we have already demonstrated.

It is noteworthy that the latitude of the jet, a first-order quantity of much interest for climate modeling, shows no sign of convergence between the models, or even as a function of resolution for either model. For example, in the finite-volume core the jet is found near 46.6° for all three vertical resolutions at N36 horizontal

resolution. When the resolution is doubled to N72, the jet clusters near 44.8°. Such a shift can have dramatic implications for climate, as observed with the GFDL coupled climate models 2.0 and 2.1 (Delworth et al. 2006). We also note that the magnitude of the jet shift with resolution (roughly 1°–2°) in these simple models is of the same order as the response of the jet to twenty-first-century greenhouse gas forcing, as found in Intergovernmental Panel on Climate Change (IPCC) climate simulations (Yin 2005). While this by no means discounts the validity of the IPCC conclusions, it suggests that there may still be uncertainty in the dynamics, where there is great confidence in the governing equations but the numerics are not fully resolved.
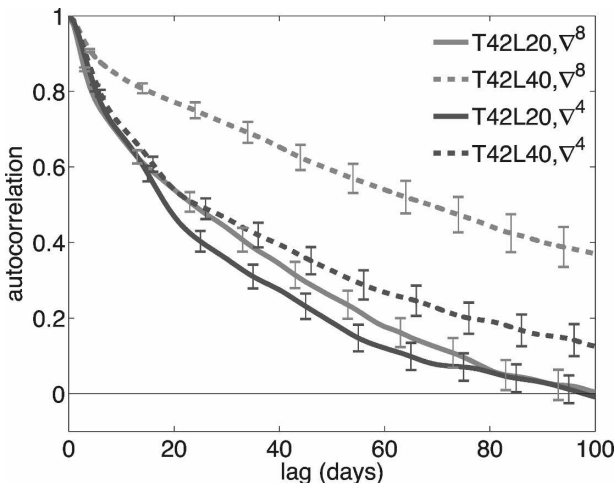

FIG. 5. Autocorrelation functions of the pseudospectral model as a function of lag with 20 and 40 levels in the vertical with different diffusion, ∇^8 and ∇^4.
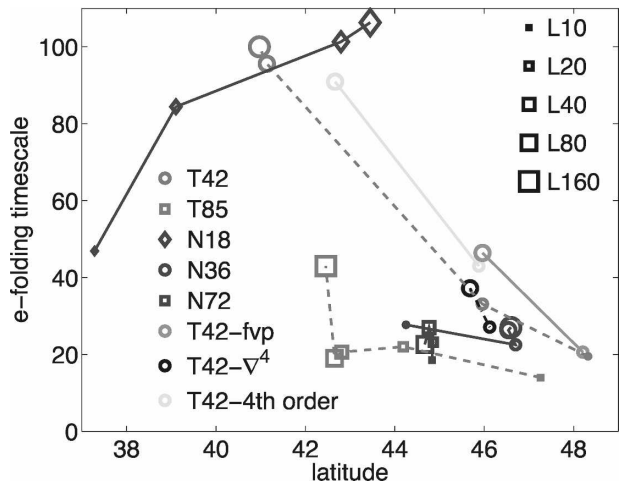

FIG. 6. The time scale of the annular mode as a function of the latitude of the surface wind maximum. Integrations differing only in vertical resolution have been connected together; the size of the marker indicates the number of σ levels. Here "fvp" refers to integrations with finite-volume parabolic vertical advection, "∇^4" to integrations with fourth-order horizontal diffusion, and "4th order" to integrations with fourth-order vertical advection. The jet latitude is found by parabolic interpolation. The T21 integrations are too persistent to be plotted without losing detail for the better-resolved integrations.
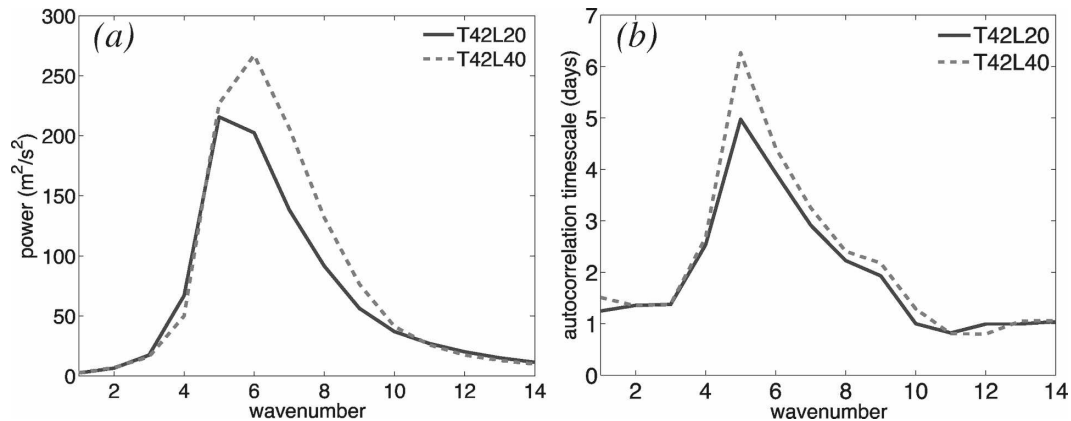
FIG. 7. Characterization of nonannular variability in the T42L20 and T42L40 models: (a) the power and (b) the autocorrelation time scale of the energy in the meridional wind in the jet core as a function of wavenumber. The plots were constructed by decomposing $v(t, \lambda)$ at 40°N and 275 (262) hPa in the 20- (40)-layer models into wavenumber space, $\tilde{v}(t, n)$, and computing the time average and e-folding autocorrelation time scales of $|\tilde{v}(t, n)|^2$ for each wavenumber n separately.

Last, we explore how specific the anomalous variability in the pseudospectral model is to the Held–Suarez forcing. We consider the annular mode decorrelation time scale in the same dynamical core, but with the forcing schemed developed by Schneider and Walker (2006). The Schneider–Walker forcing differs chiefly in the prescribed equilibrium temperature profile, which is unstable to convection, and the addition of a simple convection parameterization. We use the same parameters as in Son and Polvani (2007), and run the model at T42 horizontal resolution with 10, 20, 40, and 80 evenly spaced σ levels in the vertical. The annular mode autocorrelation time scale in this model is much less sensitive to vertical resolution: 33.0 ± 3.5, 24.8 ± 2.1, 25.5 ± 2.2, and 23.4 ± 2.4 days for the 10-, 20-, 40-, and 80-level configurations, respectively. The longest time scale is found in the coarsest integration, opposite the trend observed with the HS94 forcing. While the difference may stem from differences in the temperature profile, convection may aid in diffusing the flow in the vertical, and so remediate the numeric problems encountered with the Held–Suarez forcing. It is interesting to note that the time scales here are about the same as those in the properly resolved HS94 model, about 25 days, despite substantial differences in the forcing and climatology (not shown).

The exact cause of the anomalous persistence in the high vertical resolution pseudospectral integrations forced with the Held–Suarez forcing is difficult to attribute. The problem is particularly pernicious in that the model's climate is otherwise quite reasonable. As seen in Fig. 7, for example, the energy and time scales associated with planetary and synoptic eddies change little between the T42L20 and T42L40 integrations, de-

spite the factor-of-3 difference in the annular mode time scales. (This is not the case in the low-horizontal-resolution T21 and N18 integrations, where synoptic variability is clearly anomalous; in some cases, eddy wave packets persist on monthly time scales.) The large τ values in the T42L40 and L80 models are rather associated with a strong eddy–zonal wind feedback where the eddies and jet vacillate in concert, as documented by Gerber and Vallis (2007). Because of the sensitivity both to the details of the numerics and the model forcing, it may be of limited value to understand in great detail how the numerical truncation error leads to the unnatural feedback. We present this case study chiefly to illustrate the need to test a model, as anomalous low-frequency variability can be difficult to detect without directly testing for it, and since the conditions which lead to it—for example, changes in forcing or numerics—can be difficult to predict.

4. Implications for climate change

Fluctuation–dissipation theory relates the response of a system to external forcings to the characteristics of its natural, unforced variability. The fluctuation–dissipation relationship was developed in the context of thermodynamic equilibrium, where the statistics of the system are Gaussian, and knowledge of the covariance structure provides sufficient information to predict the response to external forcing. Leith (1975) proposed that the theory may have utility in modeling the atmosphere and ocean system, despite the fact that statistics are not necessarily Gaussian. More recently, Majda et al. (2005) has extended the fluctuation–dissipation relationship for application in climate sciences and non-

Gaussian systems. To test the applicability of the theorem to the Held–Suarez system, we consider a suite of 10 additional experiments based on the work of Ring and Plumb (2007). The theory predicts that the steady-state response of a particular mode to external perturbation is related to the integrated weight of the autocorrelation function associated with the mode; in the case of exponential decay, this implies that the response should be proportional to τ, the autocorrelation time scale. Thus we consider the response of two spectral model configurations, T42L20 and T42L40, where τ varies by a factor of three, and ask the following: Is the T42L40 model 3 times more sensitive to external forcing?

Following Ring and Plumb (2007) we consider the impact of Gaussian-shaped torques T_u on the zonal mean wind. Specifically, we add on the right-hand side of the zonal momentum equation a forcing term of the form

$$T_u(\phi, p) = \alpha_T \sum_{j=1}^{2} (-1)^j \exp\left[\frac{(p - p_T)^2}{w_P^2}\right]$$

$$\times \exp\left\{\frac{[\phi - (-1)^j \phi_T]^2}{w_\phi^2}\right\}. \qquad (4.1)$$

This forcing is westerly in the Northern Hemisphere and easterly in the Southern Hemisphere, so that there is no net torque on the atmosphere. The torques are placed in the lower troposphere by setting $p_T = 750$ hPa, and centered on the poleward flanks of the climatological jets by choosing $\phi_T = 55°$ in T42L20 and $50°$ in T42L40, with half widths $w_p = 200$ hPa and $w_\phi = 8°$ in height and latitude, respectively. We perform integrations with $\alpha_T = 0.125, 0.25, 0.375, 0.5$, and 0.75 m s^{-1} day^{-1} with the T42L20 model and $\alpha_T = 0.05, 0.1, 0.125, 0.25$, and 0.375 m s^{-1} day^{-1} with the T42L40 model. Weaker torque is required for the L40 integrations to keep the response linear. In all cases, the model is integrated for 5250 days and the first 250 days discarded.

We use the first EOF of the zonally averaged zonal winds as a function of latitude and height, $E_u(\phi, p)$, to quantify the forcing and response of the models to remain as close as possible to Ring and Plumb (2007). The EOF based on zonal winds captures the same variability as the EOF based on the surface pressure discussed above; the principal component time series associated with the zonal wind and surface pressure EOFs are correlated at 0.86 and 0.94 in the 20- and 40-layer models, respectively. The e-folding autocorrelation time scales associated with the E_u modes are slightly larger than those based on surface pressure, 37 days in the L20 model and 111 days in L40, but the ratio of time scales
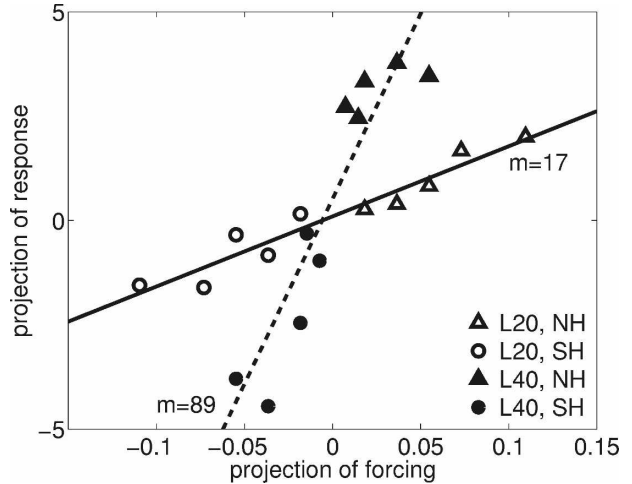

FIG. 8. The response as a function of the forcing for the first EOF of zonal mean zonal wind in the T42L20 and T42L40 models. The forcing and response are quantified by projecting the torque and the resulting change in the zonal mean zonal wind onto the first EOF of zonal wind, respectively. The magnitude of the response in the L40 model is on average $89/17 \approx 5$ times that in the L20 model, as determined by the slopes of the best-fit lines. In other words, the L40 configuration of the model is roughly *5 times* as sensitive as the L20 model to the same external perturbation.

is the same. The forcing of the zonal wind EOF mode is determined by projecting the torque T_u onto E_u. The response of the mode is determined by projecting the difference between the time and zonal average zonal winds of the perturbed integration, $[\overline{u}]_{\text{pert}}$, and that of the unperturbed integration, $[\overline{u}]_{\text{control}}$, onto E_u. We treat each hemisphere separately, allowing us to analyze the impact of both positive and negative torques.

The key result is shown in Fig. 8, where the projection of the response ($[\overline{u}]_{\text{pert}} - [\overline{u}]_{\text{control}}$) onto E_u is plotted against the projection of the external torque T_u onto E_u. The more persistent T42L40 model is *much more* sensitive to external perturbation than the T42L20 model, as predicted by the fluctuation–dissipation relationship.

Quantitatively, however, the result is not entirely consistent with the theory. The ratio of the slopes, $89/17 \approx 5$, indicates that the L40 model is on average 5 times as sensitive to changes in the forcing, while the ratio of autocorrelation time scales would suggest a threefold ($\approx 111/37$) increase in sensitivity. This discrepancy may be due to a number of reasons. First, deviations about the best-fit lines suggest that more integration time is required for convergence, particularly in the extremely persistent T42L40 model, and that some of our forced integrations may lie outside the linear regime. Second, we have restricted our analysis

to the zonal wind, hence neglecting other dynamic variables. Last, as noted by Leith (1975), the impact of other modes in the system must be accounted for if they are exhibit significant lag correlations with the annular mode. This may be particularly important in the more realistic T42L20 model, where there is poleward propagation of zonal wind anomalies (Son et al. 2008).

Despite the lack of quantitative consistency, however, this result indicates that correct simulation of internal (unforced) variability τ is necessary in climate change experiments, even if the model is primarily being used to determine the response in the time mean. As suggested by the fluctuation–dissipation relationship, *a model with poor internal variability cannot be expected to exhibit the correct response to perturbed climate forcing.*

5. Conclusions

We have proposed that a new diagnostic be computed for validating simple atmospheric general circulation models within the HS94 framework. The HS94 test was originally designed to compare different numeric schemes, but in recent years primitive equation models with HS94-like forcings have been used in climate dynamics studies, especially those requiring very long integrations. We argue that, in addition to the evaluation of the mean circulation and eddy statistics of a model, as in the original proposal, the low-frequency variability should be assessed by determining the annular mode autocorrelation time scale τ. The value of τ quantifies the time scale at which the latitudinal location of the extratropical jet varies. HS94 specifically welcome the introduction of new diagnostics, provided they meet two criteria: first, that the diagnostics isolate the dynamics away from physical parameterizations, and second, that they be based on long-term statistical properties of a realistic three-dimensional global circulation. Our proposed assessment of the low-frequency variability satisfies both criteria. We offer a relatively simple procedure for estimating the new diagnostic τ and a means of computing the uncertainty of the estimate. Based on the results of two resolution sweep experiments, we find that the HS94 forcing of the primitive equations exhibits an intrinsic time scale of jet vacillation, with two distinct numerical models converging around values of 20–25 days. These values establish a benchmark for comparison and testing the convergence of future numerical schemes.

The need for the third component of the test is demonstrated in a case study of two simple GCMs developed by the Geophysical Fluid Dynamics Laboratory. The two models are driven by the same HS94 forcing but utilize different numerics: pseudospectral and finite volume. At moderate horizontal resolution, T42 and N36, both models produce realistic climatologies and eddy statistics over a range of vertical resolutions, but the low-frequency variability in the pseudospectral model is extremely sensitive to model details. Changes to the number of vertical levels, the vertical advection scheme, or the horizontal diffusion have dramatic and unpredictable impacts on the value of τ. The fact that T42 pseudospectral resolution is generally considered sufficient for dynamical studies (e.g., Held and Phillipps 1993; Boer and Denis 1997) highlights the need for the new diagnostic proposed here. At higher resolution T85 and N72, the time scale of the annular mode appears to have largely converged.

Tests with an alternative forcing scheme developed by Schneider and Walker (2006) indicate that the problems with vertical resolution in the T42 pseudospectral model may be very specific to the HS94 forcing. The values of τ in the Schneider–Walker forced integrations of the T42 model with varying vertical resolution exhibit little sensitivity to the number of layers. The time scales are quite similar to those found in the HS94 model, around 20–25 days, despite substantial differences in the forcing and climatology. Hence τ for both forcing schemes is not too far from that of the observed annular mode, which varies between 10 and 20 days over the seasons.

This diagnostic is particularly important for studies in which a simple model, because of its computational efficiency, is used to study low-frequency variability on intraseasonal to interannual time scales or climate change. For studies of internal variability, if the persistence of the extratropical circulation is unrealistically long, the model must be run much longer to achieve the same statistical certainty provided by a model with more realistic variability, as a new state is sampled on the order of ever τ days (e.g., Kushner and Polvani 2004). Potentially more worrisome, the unnatural separation in time scales between the low-frequency variability and synoptic variability may lead to unphysical coupling between eddies and the mean flow, or between the stratosphere and troposphere. In this respect, particular caution should be taken when using a simple model to analyze stratospheric–tropospheric coupling. The finer vertical resolution necessary to resolve processes in the upper atmosphere may have a negative impact on the tropospheric circulation; higher vertical resolution does not always entail better simulation of the dynamics.

The fluctuation–dissipation relationship suggests that proper representation of a model's unforced (or internal) variability is important for studies of climate

change—even if the internal variability is not the focus of the investigation. We find that the extratropical jet is much more sensitive to applied torque in models where the annular mode is more persistent. This suggests that the response of a model to more general perturbations, such as those associated with increased greenhouse gas forcing or ozone depletion and recovery, may be abnormal if the low-frequency variability is poorly simulated.

APPENDIX

The Uncertainty of τ_N

We here derive (2.2), the uncertainty in the empirical estimate of the decorrelation time scale τ_N. To obtain our rough estimate, we use the fact that decay time scale of an exponential can be estimated by the determining at what time the function crosses the e^{-1} line. From (2.1), the variance of the empirical autocorrelation function $r_N(t)$ about the actual AR-1 autocorrelation function $r(t) = e^{-t/\tau}$, at time τ, is given by

$$\mathrm{var}[r_N(\tau)] = \frac{1}{N}\left[\frac{(1 + e^{-2/\tau})(1 - e^{-2})}{1 - e^{-2/\tau}} - 2\tau e^{-2}\right].$$

$$(A.1)$$

For large τ (i.e., $\tau \gg 2$), we take the first terms of the Taylor expansion to approximate $\exp(-2/\tau) \approx 1 - 2/\tau$. Hence

$$\mathrm{var}[r_N(\tau)] \approx \frac{1}{N}\left\{\frac{[1 + (1 - 2/\tau)](1 - e^{-2})}{1 - (1 - 2/\tau)} - 2\tau e^{-2}\right\},$$

$$(A.2)$$

$$\approx \frac{1}{N}\left[\frac{(2 - 2/\tau)(1 - e^{-2})}{2/\tau} - 2\tau e^{-2}\right], \quad (A.3)$$

$$\approx \frac{1}{N}[(1 - e^{-2})\tau - 2e^{-2}\tau], \quad (A.4)$$
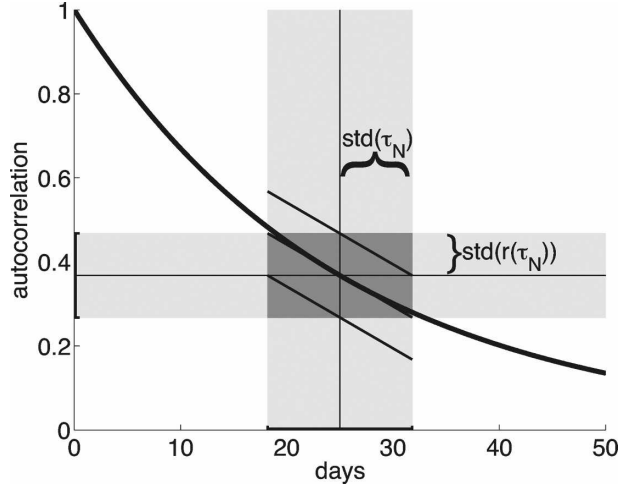
$$\approx \frac{\tau}{N}(1 - 3e^{-2}), \quad (A.5)$$

FIG. A1. A cartoon depicting the strategy for estimating the uncertainty in the empirical decorrelation time scale τ_N, given the uncertainty in the autocorrelation function $r_N(t)$. We linearize $r_N(t)$ around the point (τ, e^{-1}), estimating the derivative as that of a pure exponential.

where we assume $1/\tau \ll 1$ in (A.3). Hence the standard error in the height of $r_N(t)$ at time τ is given by $(1 - 3e^{-2})^{1/2}\tau^{1/2}N^{-1/2}$.

To determine the uncertainty in τ_N, we estimate the variability in the time when $r_N(t)$ reaches e^{-1}. We do so linearizing $r_N(t)$ about the point (τ, e^{-1}), assuming that $r_N(t)$ has an exponential form. As shown in Fig. A1, we use the slope $(dr/dt)|_{t=\tau} = -e^{-1}/\tau$ to estimate the standard error in τ_N. Thus

$$-e^{-1}/\tau \approx \frac{\mathrm{rise}}{\mathrm{run}}, \quad (A.6)$$

$$|-e^{-1}/\tau| \approx \left|\frac{\mathrm{std}[r_N(\tau_N)]}{\mathrm{std}(\tau_N)}\right|, \quad (A.7)$$

$$e^{-1}/\tau \approx \frac{(1 - 3e^{-2})^{1/2}\tau^{1/2}N^{-1/2}}{\mathrm{std}(\tau_N)}, \quad (A.8)$$

$$\mathrm{std}(\tau_N) \approx k\tau^{3/2}N^{-1/2}, \quad (A.9)$$

where $k = e(1 - 3e^{-2})^{1/2} \approx 2$. The annular mode index is generally not entirely an AR-1 process, and so $r_N(t)$ is not perfectly exponential. These arguments are only intended to provide a rough estimate of the uncertainty, and how it scales with τ and N.

REFERENCES

Baldwin, M. P., D. B. Stephenson, D. W. J. Thompson, T. J. Dunkerton, A. J. Charlton, and A. O'Neill, 2003: Stratospheric memory and skill of extended-range weather forecasts. *Science,* **301,** 636–640.
Bartlett, M. S., 1946: On the theoretical specification and sam-

pling properties of autocorrelated time-series. *Suppl. J. Roy. Stat. Soc.,* **8,** 27–41.

Boer, G. J., and B. Denis, 1997: Numerical convergence of the dynamics of a GCM. *Climate Dyn.,* **13,** 359–374.

Delworth, T. L., and Coauthors, 2006: GFDL's CM2 global coupled climate models. Part I: Formulations and simulation characteristics. *J. Climate,* **19,** 643–674.

Eichelberger, S. J., and D. L. Hartmann, 2005: Changes in the strength of the Brewer–Dobson circulation in a simple AGCM. *Geophys. Res. Lett.,* **32,** L15807, doi:10.1029/2005GL022924.

Feldstein, S. B., 2000a: Is interannual zonal mean flow variability simply climate noise? *J. Climate,* **13,** 2356–2362.

——, 2000b: The timescale, power spectra, and climate noise properties of teleconnection patterns. *J. Climate,* **13,** 4430–4440.

Franzke, C., S. Lee, and S. B. Feldstein, 2004: Is the North Atlantic Oscillation a breaking wave? *J. Atmos. Sci.,* **61,** 145–160.

Gerber, E. P., 2005: A dynamical and statistical understanding of the North Atlantic Oscillation and annular modes. Ph.D. thesis, Princeton University, 252 pp.

——, and G. K. Vallis, 2007: Eddy–zonal flow interactions and the persistence of the zonal index. *J. Atmos. Sci.,* **64,** 3296–3311.

Held, I. M., and P. J. Phillipps, 1993: Sensitivity of the eddy momentum flux to meridional resolution in atmospheric GCMs. *J. Climate,* **6,** 499–507.

——, and M. J. Suarez, 1994: A proposal for the intercomparison of the dynamical cores of atmospheric general circulation models. *Bull. Amer. Meteor. Soc.,* **75,** 1825–1830.

Kushner, P. J., and L. M. Polvani, 2004: Stratosphere–troposphere coupling in a relatively simple AGCM: The role of eddies. *J. Climate,* **17,** 629–639.

——, and ——, 2005: A very large, spontaneous stratospheric sudden warming in a simple AGCM: A prototype for the Southern Hemisphere warming of 2002? *J. Atmos. Sci.,* **62,** 890–897.

——, and ——, 2006: Stratosphere–troposphere coupling in a relatively simple AGCM: Impact of the seasonal cycle. *J. Climate,* **19,** 5721–5727.

Leith, C. E., 1975: Climate response and fluctuation dissipation. *J. Atmos. Sci.,* **32,** 2022–2026.

Lin, S. J., 2004: A "vertically Lagrangian" finite-volume dynamical core for global models. *Mon. Wea. Rev.,* **132,** 2293–2307.

Lindzen, R. S., and M. Fox-Rabinovitz, 1989: Consistent vertical and horizontal resolution. *Mon. Wea. Rev.,* **117,** 2575–2583.

Majda, A. J., R. V. Abramov, and M. J. Grote, 2005: *Information Theory and Stochastics for Multiscale Nonlinear Systems. CRM Monogr.,* Ser. 25, American Mathematical Society, 133 pp.

Namias, J., 1950: The index cycle and its role in the general circulation. *J. Meteor.,* **7,** 130–139.

North, G. R., T. L. Bell, R. F. Cahalan, and F. J. Moeng, 1982: Sampling errors in the estimation of empirical orthogonal functions. *Mon. Wea. Rev.,* **110,** 699–706.

Oort, A. H., and J. J. Yienger, 1996: Observed interannual variability in the Hadley Circulation and its connection to ENSO. *J. Climate,* **9,** 2751–2767.

Polvani, L. M., and P. J. Kushner, 2002: Tropospheric response to stratospheric perturbations in a relatively simple general cir-

culation model. *Geophys. Res. Lett.,* **29,** 1114, doi:10.1029/2001GL014284.

Ring, M. J., and R. A. Plumb, 2007: Forced annular mode patterns in a simple atmospheric general circulation model. *J. Atmos. Sci.,* **64,** 3611–3626.

Robinson, W. A., 1996: Does eddy feedback sustain variability in the zonal index? *J. Atmos. Sci.,* **53,** 3556–3569.

Rossby, C. G., 1939: Relation between variations in the intensity of the zonal circulation of the atmosphere and the displacements of the semi-permanent centers of action. *J. Mar. Res.,* **2,** 38–55.

Schneider, T., 2004: The tropopause and the thermal stratification in the extratropics of a dry atmosphere. *J. Atmos. Sci.,* **61,** 1317–1340.

——, and C. C. Walker, 2006: Self-organization of atmospheric macroturbulence into critical states of weak nonlinear eddy–eddy interactions. *J. Atmos. Sci.,* **63,** 1569–1586.

Son, S.-W., and L. M. Polvani, 2007: The dynamical formation of an extra-tropical tropopause inversion layer in a relatively simple general circulation model. *Geophys. Res. Lett.,* **34,** L17806, doi:10.1029/2007GL030564.

——, S. Lee, S. B. Feldstein, and J. E. Ten Hoeve, 2008: Time scale and feedback of zonal mean flow variability. *J. Atmos. Sci.,* **65,** 935–952.

Song, Y., and W. A. Robinson, 2004: Dynamical mechanisms for stratospheric influences on the troposphere. *J. Atmos. Sci.,* **61,** 1711–1725.

Taguchi, M., 2003a: Tropospheric response to stratospheric degradation in a simple global circulation model. *J. Atmos. Sci.,* **60,** 1835–1846.

——, 2003b: Tropospheric response to stratospheric sudden warmings in a simple global circulation model. *J. Climate,* **16,** 3039–3049.

——, and S. Yoden, 2002a: Internal interannual variability of the troposphere–stratosphere coupled system in a simple global circulation model. Part I: Parameter sweep experiment. *J. Atmos. Sci.,* **59,** 3021–3036.

——, and ——, 2002b: Internal interannual variability of the troposphere–stratosphere coupled system in a simple global circulation model. Part II: Millennium integrations. *J. Atmos. Sci.,* **59,** 3037–3050.

——, T. Yamaga, and S. Yoden, 2001: Internal variability of the troposphere–stratosphere coupled system simulated in a simple general circulation model. *J. Atmos. Sci.,* **58,** 3184–3203.

Thompson, D. W. J., and J. M. Wallace, 2000: Annular modes in the extratropical circulation. Part I: Month-to-month variability. *J. Climate,* **13,** 1000–1016.

Vallis, G. K., E. P. Gerber, P. J. Kushner, and B. A. Cash, 2004: A mechanism and simple dynamical model of the North Atlantic Oscillation and annular modes. *J. Atmos. Sci.,* **61,** 264–280.

Von Storch, H., and F. W. Zwiers, 1999: *Statistical Analysis in Climate Research.* Cambridge University Press, 484 pp.

Walker, C. C., and T. Schneider, 2006: Eddy influences on Hadley circulations: Simulations with an idealized GCM. *J. Atmos. Sci.,* **63,** 3333–3350.

Yin, J. H., 2005: A consistent poleward shift of the storm tracks in simulations of 21st century climate. *Geophys. Res. Lett.,* **32,** L18701, doi:10.1029/2005GL023684.