

Intercomparison of the northern hemisphere winter mid-latitude atmospheric variability of the IPCC models

Valerio Lucarini · Sandro Calmanti ·
Alessandro Dell'Aquila · Paolo M. Ruti ·
Antonio Speranza

Received: 17 January 2006 / Accepted: 28 September 2006 / Published online: 14 December 2006
© Springer-Verlag 2006

Abstract We compare, for the overlapping time frame 1962–2000, the estimate of the northern hemisphere mid-latitude winter atmospheric variability within the available 20th century simulations of 19 global climate models included in the Intergovernmental Panel on Climate Change—4th Assessment Report with the NCEP-NCAR and ECMWF reanalyses. We compute the Hayashi spectra of the 500 hPa geopotential height fields and introduce an ad hoc integral measure of the variability observed in the Northern Hemisphere on different spectral sub-domains. The total wave variability is taken as a global scalar metric describing the overall performance of each model, while the total variability pertaining to the eastward propagating baroclinic waves and to the planetary waves are taken as scalar metrics describing the performance of each model phenomenologically in connection with the corresponding specific physical process. Only two very high-resolution global climate models have a good agreement with reanalyses for both the global and the process-oriented metrics. Large biases, in several cases larger than 20%, are found in all the considered metrics between the wave climatologies of most IPCC models and the reanalyses, while the span of the climatologies of the various models is, in all cases, around 50%. In particular, the travelling baroclinic waves are typically

overestimated by the climate models, while the planetary waves are usually underestimated, in agreement with what found in past analyses performed on global weather forecasting models. When comparing the results of similar models, it is apparent that in some cases the vertical resolution of the model atmosphere, the adopted ocean model, and the advection schemes seem to be critical in the bulk of the atmospheric variability. The models ensemble obtained by arithmetic averaging of the results of all models is biased with respect to the reanalyses but is comparable to the best five models. Nevertheless, the models results do not cluster around their ensemble mean. This study suggests caveats with respect to the ability of most of the presently available climate models in representing the statistical properties of the global scale atmospheric dynamics of the present climate and, a fortiori, in the perspective of modeling climate change.

1 Introduction

The operational global weather forecast models routinely used in the 1980's were well-known to suffer from serious biases in the representation of the northern hemisphere (NH) winter mid-latitude synoptic and planetary atmospheric disturbances (Tibaldi 1986). In particular, it was confirmed in a number of studies that such models on the average featured a large overestimation of the baroclinic synoptic waves (e.g. Sumi and Kanamitsu 1984; Klinker and Capaldo 1986; Siegmund 1995) and a large underestimation of the planetary waves (e.g. Wallace et al. 1983), with

V. Lucarini (✉) · A. Speranza
Dipartimento di Matematica ed Informatica,
Università di Camerino, Camerino (MC), Italy
e-mail: valerio.lucarini@unicam.it

S. Calmanti · A. Dell'Aquila · P. M. Ruti
Progetto Speciale Clima Globale,
Ente Nazionale per le Nuove Tecnologie,
l'Energia e l'Ambiente, Rome, Italy

biases sometimes of the same size as the average climatology of the observations and analyzed atmospheric fields. In many cases, such forecasting models constitute the base of the atmospheric components presently adopted in up-to-date global climate models (GCMs). Some brief notes on the history of atmospheric modeling can be found at <http://www.aip.org/history/sloan/gcm/intro.htm>. Such biases may have a relatively minor impact on reanalysis systems, since observations are used to correct the autonomous evolution of the atmospheric models. Instead, since synoptic and planetary waves provide a relevant contribution to the meridional transport of energy and momentum in the mid-latitudes, the physical processes driving such phenomena are sensible ingredients for the characterization of the climate system (Speranza 1983; Peixoto and Oort 1992), and discrepancies may be critical when interpreting the results of unconstrained climate models, possibly reducing the significance of the resulting simulations of past and future climatic conditions.

In view of the above remarks, we compare the performances of 19 GCMs, which contribute to the development of the scientific basis of the IPCC 4th Assessment Report (henceforth, IPCC-4AR), in the description of the NH mid-latitudes winter atmospheric disturbances. We do this by analysing in detail the statistical properties of the modelled 500 hPa geopotential height fields. We analyze the self-consistency of the climate models and we assess their realism by comparing their statistics with those of the NCEP-NCAR and ECMWF global reanalyses, which have been analyzed in detail elsewhere (Dell'Aquila et al. 2005). The 500 hPa geopotential height is a very powerful and widely used proxy for characterizing atmospheric variability on a wide range of spatial and temporal scales (Blackmon 1976). We perform for all models a space-time spectral analysis devised along the way of Hayashi (1971, 1979), which allows the definition of the spectral densities of different types of atmospheric disturbances as well as the separation of the spectral densities describing statistically standing and propagating waves. May (1999) has provided an example of application of Hayashi spectra techniques for the assessment of the performance of climate models. The synoptic travelling waves can be represented as high-frequency high-wavenumber (HFHW) eastward propagating spectral features, characterized by periods of order 2–7 days and by spatial scales of a few thousands of kilometers. These waves can be associated with the release of available energy driven by conventional baroclinic conversion (Blackmon 1976; Speranza 1983; Wallace et al. 1988), so that they

are often referred to as baroclinic waves. On the other hand, planetary waves, which interact with orography (Charney and DeVore 1979; Charney and Straus 1980; Buzzi et al. 1984; Benzi et al. 1986) and are catalyzed by the sub-tropical jet (Benzi and Speranza 1989; Ruti et al. 2006), play a dominant role in the low-frequency low-wavenumber (LFLW) spectral region of standing waves, whose characteristic time and space scales are in the interval 10–45 days and 7,000–15,000 km, respectively (Hansen and Sutera 1986).

To support the tantalizing aim of planning practical strategies for the improvement of numerical climate models, the project for climate model diagnostics and intercomparison (PCMDI: <http://www.pcmdi.llnl.gov>) has gathered the output of climate models produced worldwide into a single server and solicited the provision of simple scalar metrics of model performances. We underline the possibility of distinguishing two distinct, albeit related, kinds of metrics.

The comparison of bulk thermodynamic quantities defining the climate state, such as the tropospheric average temperature, tropospheric average specific humidity, variance of geopotential height, allows the definition of *global metrics* which may be considered as robust diagnostic tools which are representative of the overall properties of the system. Nevertheless, such an approach does not allow for the disentanglement of the role of each one of the vast range of distinct physical processes. In order to capture the differences in the representation of specific physical processes, it is necessary to use specialized diagnostic tools—that we may call *process-oriented metrics*—as indexes for model reliability. This approach may be helpful in clarifying the distinction between the performance of the models in reproducing *diagnostic* and *prognostic* variables of the atmospheric system.

In the present analysis, we introduce the global and process oriented metrics by defining ad hoc integral measures over different spectral sub-domains of the computed Hayashi spectral densities. The estimate of the total wave variability of the field is chosen as a global scalar metric assessing the overall statistical properties of different types of atmospheric disturbances. The estimate of the variability pertaining to the eastward propagating HFHW waves and to the standing LFLW waves are taken as scalar metrics, accounting for the model representation of specific physical processes and specifically related to the capability in describing the travelling baroclinic waves and planetary waves, respectively. We maintain that by applying this method, the reliability of GCMs in the simulation of some basic climatic processes can be critically addressed.

This paper is organized as follows. In Sect. 2 we describe the datasets and sketch the method employed for analysis. In Sect. 3 we compare the model performances in specific sub-regions of the full space-temporal spectrum; the conclusions are drawn in Sect. 4.

2 Data and methods

The 19 IPCC models considered in this study are listed in Table 1 along with their main features. Although the analysis is applied to an atmospheric variable, we report also a few details about the ocean component in order to stress that the results are representative of the response of a complex system which can not be reduced to the behavior of its single components (Lucarini 2002). Indeed, we will see that for some models (e.g. the GISS model), the same atmosphere over different oceans results in substantially different atmospheric variability. The PCMDI datasets consists of various simulations of different scenarios, including the reconstruction of past and present climate, control runs with fixed atmospheric compositions, and future climate projections. In this work we consider the 20th century simulations, where the time frame considered is 1962–2000. The output data of the INGV-SXG model have been obtained directly from the INGV modeling group, since unfortunately up to September 2006 they have not been available from the PCMDI server. All models attempt the simulation of the climate of this period by setting as time-varying input the observed concentrations of trace gases, such as CO_2 and ozone.

The IPCC models output for 1962–2000 is compared with observations by considering two major reanalysis datasets: the reanalysis produced by the National Center for Environmental Prediction (NCEP), in collaboration with the National Center for Atmospheric Research (NCAR) (Kistler et al. 2001), and one released by the European Center for Mid-Range Weather Forecast (ECMWF) (Simmons and Gibson 2000).

2.1 Geostrophic approximation and latitudinal averaging

Since the goal of this study is to diagnose the mid-latitude winter atmospheric variability of GCMs, along the lines of Dell’Aquila et al. (2005), we should use the December–January–February (DJF) daily values of geopotential height at 500 hPa data averaged over the latitudinal belt 30°N – 75°N , where the bulk of mid-latitude atmospheric waves activity is observed (Blackmon 1976; Speranza 1983).

Unfortunately, this field is not one of the standard daily three dimensional (3D) field outputs of the IPCC models, which comprise only zonal and meridional wind speed, air temperature and specific humidity. In principle, the geopotential height could be rigorously computed from temperature and specific humidity using the hydrostatic relation and the equation of state for air. Such an approach requires, however, knowledge of two time-varying 3D fields, namely temperature and specific humidity. Also the knowledge of boundary terms such as the constant 2D field of surface height and the time-varying 2D field of surface pressure is needed. The latter field is not readily available in the IPCC-4AR dataset, but could be reconstructed with suitable interpolations from the available sea-level pressure data.

In view of the large computational resources needed and the uncertainties on the surface boundary term, we have instead retrieved the 500 hPa meridional wind speed. In fact, in the geostrophic approximation, the meridional wind is related to the zonal gradient of the geopotential height by:

$$f(\phi)v(p, \phi, \lambda, t) = \frac{g}{R \cos(\phi)} \frac{\partial z(p, \phi, \lambda, t)}{\partial \lambda} \quad (1)$$

where t is time, p , ϕ , and λ , respectively, the pressure, the latitude and the longitude, $f(\phi) = 2\Omega \sin(\phi)$, v indicates the meridional component of the wind velocity vector, g is the gravity acceleration, R is the radius of the Earth, and z is the geopotential height. This approach requires much smaller computer resources, has a simple model-independent implementation, and is *local* in the sense that it involves only the relevant $p = 500$ hPa level. Of course the geostrophic relation is only an order Rossby number (about 0.05–0.10 in our case) approximation, but is well suited for the mid-latitudes (Peixoto and Oort 1992; Holton 1992), where we restrict our analysis. Finally, since, geostrophy justifies the possibility of extracting information about the dynamics from the geopotential height field, such an approach is consistent with the goal of this study.

We can then obtain the following relation between the longitudinal derivative of the zonally averaged geopotential height and the meridional velocity (Peixoto and Oort 1992):

$$\frac{\partial \langle z(p, \phi, \lambda, t) \rangle_{\phi_1}^{\phi_2}}{\partial \lambda} = \frac{R}{g} \langle f(\phi)v(p, \phi, \lambda, t) \cos(\phi) \rangle_{\phi_1}^{\phi_2} \quad (2)$$

where the area-weighted latitudinal average of the generic function $\psi(p, \phi, \lambda, t)$ is defined as follows:

Table 1. An overview of the IPCC models

	Basic Information			Atmosphere					Ocean		
	Model (Reference)	Institution	Trunc. (Lon x Lat)	Z	Chemistry and Aerosols	Clouds and Convection	Albedo	Numerics and Adjustments	Lon x Lat. (Equator)	Vertical	
1.	BCCR-BCM2.0 Furevik et al. (2003)	Bjerknes Center for Climate Research, Norway	T63	h31	m_{A_p} , da_{A_p} , u_{A_p} , s_{A_p}	S: Ricard and Royer (1993) C: Bougeault (1985)	O_o , I_m , L_m	L, sI	$1.5^\circ \times 1.5^\circ$ ($1.5^\circ \times 0.5^\circ$)	35	
2.	CNRM-CM3 Salas-Méjia et al. (2005)	Météo France, France	T63	h45	O_o , m_{A_p} , da_{A_p} , u_{A_p} , s_{A_p}	S: Ricard and Royer (1993) C: Bougeault (1985)	-	L, sI	$2^\circ \times 2^\circ$ ($2^\circ \times 0.5^\circ$)	31z	
3.	CGCM3-1(T47)		T47(T63)	z31	-	C: Zhang and McFarlane(1995)	-		$1.85^\circ \times 1.85^\circ$ ($1.4^\circ \times 0.94^\circ$)	29z	
4.	CGCM3-1(T63) Kim et al. (2002)	CCCma, Canada	T63	h18	s_{A_p}	S: Roystayn (1997) C: Gregory & Rowntree (1990)	L_p	L, Lp	$1.875^\circ \times 0.84^\circ$	31z	
5.	CSIRO-Mk3.0 Gordon et al. (2002)	CSIRO, Australia	T63	h31	s_{A_p}	S: Lohmann and Roeckner (1996) C: Nordeng (1994)	I_m , L_m	L, sI/Lp	$1.5^\circ \times 1.5^\circ$	40z	
6.	ECHAM5/MPI-OM Jungelaus et al. (2006)	Max Planck Inst., Germany	T63	h19	s_{A_p}	S: Sundqvist (1979) C: Nordeng (1994)	I_m , L_m	L, sI/Lp, F	$2.8^\circ \times 2.8^\circ$ ($0.5^\circ \times 0.5^\circ$)	20z	
7.	ECHO-G Min et al. (2005)	MIUB, METRI, and M&D, Germany/Korea	T30	26	s_{A_p}	S: Kiehl et al. (1998), Rasch and Kristjansson(1998) C: Zhang & McFarlane(1995), Hack (1994)	-	O, sI, HWnC	$1^\circ \times 1^\circ$	H	
8.	FGOALS-g1.0 Yu et al. (2004)	LASG, China	$2.8^\circ \times 2.8^\circ$	h24	m_{A_p} , da_{A_p} , u_{A_p} , s_{A_p}	S: Roystayn (2000), Tiedke(1993) C: Moorthi and Suarez (1992)	-	C (CM2.0) L (CM2.1)	$1^\circ \times 1^\circ$ ($1^\circ \times 1/3^\circ$)	Z	
9.	GFDL-CM2.0	GFDL, USA	h12	h19	s_{A_p}	-	I_m , L_m	U, Lp	$4^\circ \times 3^\circ$	16	
10.	GFDL-CM2.1 Delworth et al. (2005)	NASA-GISS, USA	$4^\circ \times 3^\circ$	20	m_{A_p} , da_{A_p} , u_{A_p} , s_{A_p}	S: Lohmann and Roeckner (1996) C: Nordeng (1994)	I_o , L_m	U	$2^\circ \times 2^\circ$ (EH) $4^\circ \times 5^\circ$ (ER)	(EH) z (ER)	
11.	GISS-AOM Lucarini and Russell (2002)	NASA-GISS, USA	T106	h19	s_{A_p}	S: Volodin and Diansky (2004) C: Betts (1986)	O_o , I_m , L_m	L, sI	$2^\circ \times 2^\circ$ ($2^\circ \times 0.5^\circ$)	31z	
12.	GISS-EH		$4^\circ \times 5^\circ$	21	s_{A_p}	-	I_p , L_m	C, sI, F	$2.5^\circ \times 2^\circ$	33	
13.	GISS-ER Schmidt et al. (2005)	NASA-GISS, USA	T106	h19	s_{A_p}	S: Bony and Emanuel (2001) C: Grandpeix et al. (2004)	I_m , L_m	C, Lp	$2^\circ \times 2^\circ$ ($2^\circ \times 0.5^\circ$)	31z	
14.	INGV-SXG Gualdi et al. (2006)	INGV, Italy	$5^\circ \times 4^\circ$	h30	-	S: Le Treut and Li (1991) C: Pan and Randall (1998)	I_m , L_m	L, Lp	$1.4^\circ \times 1.4^\circ$ ($1.4^\circ \times 0.5^\circ$)	43h	
15.	INM-CM3.0 Volodin and Diansky (2004)	Inst. of Num. Math., Russia	$2.4^\circ \times 3.75^\circ$	h30	m_{A_p} , da_{A_p} , u_{A_p} , s_{A_p}	S: Yukimoto et al. (2001) C: Randall and Pan (1993)	I_m , L_m	S, Lp, F	$2.5^\circ \times 2.0^\circ$ ($2.5^\circ \times 0.5^\circ$)	23z	
16.	IPSL-CM4 Marti et al. (2005)	IPSL, France	T42	h30	s_{A_p}						
17.	MIROC3.2(hires)		T106(hires)	56(hires)							
18.	MIROC3.2(medres) K-1 mod. Dev. (2004)	CCSR/NIES/FRCGC, Japan	T42(medres)	20(medres)							
19.	MRI-CGCM2 Yukimoto and Noda (2002)	Meteorological Research Institute, Japan	T42	h30	s_{A_p}						

The horizontal resolution is expressed in terms of truncation (T) for spectral models (with $T47 \approx 2.8^\circ$, $T63 \approx 2.8^\circ$, $T63 \approx 2.8^\circ$, $T63 \approx 2.8^\circ$); for Z coordinates, the letter before the number of levels indicates whether the vertical coordinate is height (z), pressure normalized with surface pressure (σ) or hybrid (h); in the chemistry column, the chemical species included in the model are specified. The string xA_y indicates whether the included aerosols are marine ($x = m$), desertic ($x = d$), urban ($x = u$) or sulfates ($x = s$) and if they are imposed as constants ($y = c$), prescribed according to climatology ($y = p$) or modelled ($y = m$). For clouds (S) and convection (C) the relevant references are indicated. The treatment of albedo is reported for oceanic albedo (O_y), ice albedo (I_y) and land surface albedo (L_y) that can be constant ($y = c$) or modeled ($y = m$); some details about the numerics are also specified, in particular it is specified whether the advection scheme for heat and moisture is or centered (C), upstream (U), spectral (S) or semi-Lagrangian (L); if the time stepping is explicit – Leapfrog (Lp) – or semi-implicit (sl) and whether flux-adjustment is employed (F). For the oceanic component we report the horizontal resolution (resolution at the equator in parentheses if it is different from the rest of the domain); the number of vertical levels along with an indication of the vertical coordinate that can be depth (z), depth normalized with the maximum local depth (σ), hybrid (h) or density (d). For further information, refer to the PCMDI web site : <http://www.pcmdi.llnl.gov>

$$\langle \psi(p, \phi, \lambda, t) \rangle_{\phi_1}^{\phi_2} = \frac{\int_{\phi_1}^{\phi_2} \psi(p, \phi, \lambda, t) \cos(\phi) d\phi}{\int_{\phi_1}^{\phi_2} \cos(\phi) d\phi} \quad (3)$$

We rename the function $\langle z(p, \phi, \lambda, t) \rangle_{\phi_1}^{\phi_2}$ as $Z(p, \lambda, t)$ and the function $R/g(f(\phi)v(p, \phi, \lambda, t) \cos(\phi))_{\phi_1}^{\phi_2}$ as $V(p, \lambda, t)$. We then have:

$$\frac{\partial Z(p, \lambda, t)}{\partial \lambda} = V(p, \lambda, t), \quad (4)$$

which constitutes the baseline of the later described spectral approach.

2.2 Hayashi spectra

The variability of a 1D + 1D (space-time) field in terms of waves of different periods and zonal wavenumbers can be effectively described by means of the space-time Fourier decomposition introduced by Hayashi (1971, 1979). Straightforward space-time decomposition will not distinguish between standing and travelling waves: a standing wave will give two spectral peaks corresponding to travelling waves moving eastward and westward at the same speed and with the same phase. The problem can only be circumvented by making assumptions regarding the nature of the wave. One approach relies on assuming complete statistical coherence between the eastward and westward components of standing waves and on attributing the incoherent part of the spectra to real travelling waves (Fraedrich and Bottger 1978; Hayashi 1979). By computing the cross-spectra and the coherence of the signal, the method allows for separating statistically the propagating and the standing components of the atmospheric waves. We then adopt a very similar procedure to that described in Dell’Aquila et al. (2005) to compute the Hayashi spectra $H_a^n(k_j, \omega_m)$, where the index n refers to the year; the subscript a can take the values of T , S , E , W , corresponding to the total (T), standing (S), eastward propagating (E), and westward propagating (W) components of the spectrum; ($k_j = 2\pi j$, $\omega_m = 2\pi m/\tau$) represent the discrete set of wavelengths and frequencies over which the spectrum is computed, and $\tau = 90d$ is the length of the winter. We emphasize that customarily, in our terms, Hayashi spectra are represented by plotting the quantity $k_j \omega_m 2\pi/\tau H_a^n(k_j, \omega_m)$, so that equal geometrical areas represent equal variance in log–log plots. With this definition, the Hayashi spectra presented in this paper are expressed in m^2 , as done in Blackmon

(1976) and Speranza (1983), and can be compared to those given in Dell'Aquila et al. (2005) after a multiplication by $1/8 \times 86400$ s.

By considering the basic properties of transformation of the Fourier series with respect to the derivation, we have the following relations between the Hayashi spectra of the function $Z(p, \lambda, t)$ and of the function $V(p, \lambda, t)$:

$$H_a(k_j, \omega_m)|_Z = k_j^2 H_a(k_j, \omega_m)|_V. \quad (5)$$

Therefore, in this work, for each dataset, we first compute the Hayashi space-time spectra of the quantity $V(p, \lambda, t) = R/g(f(\phi)v(p, \phi, \lambda, t) \cos(\phi))_{\phi_1}^{\phi_2}$ with $p = 500$ hPa, ϕ_1 and ϕ_2 set to the grid-points closest to 30°N and 75°N , respectively, and then obtain the Hayashi spectra of the corresponding latitudinally averaged geopotential height by using Eq. 5.

In order to evaluate the model performances in different spectral sub domains, we introduce the following integral quantity:

$$E_a^n(\Omega) = \sum_{m=m_1, j=j_1}^{m=m_2, j=j_2} H_a^n(k_j, \omega_m), \quad (6)$$

where $a = T, S, E, W$, n indicates the winter; the integration extremes, $m_{1,2}$ and $j_{1,2}$, determine the spectral region of interest, with $\Omega = [\omega_{m_1}, \omega_{m_2}] \times [k_{j_1}, k_{j_2}]$. The quantity $E_a^n(\Omega)$ introduced in Eq. 6 represents the portion of variance of the spectrum associated with a given subdomain Ω and with a given winter n and is expressed in units of m^2 . The averaging process defined in Eq. 6 overcomes the well-known instability of the direct Fourier transform, in particular in describing small scale spectral features. Moreover, the intercomparison of the statistical properties of quantities such as those defined in Eq. 6 obtained for various models allow us to test the models agreement for the waves occurring on the subdomain Ω . Following basic statistical arguments, we estimate the average intra-seasonal atmospheric variability in the spectral subdomain Ω as:

$$\bar{E}_a(\Omega) = \frac{1}{N} \sum_{n=1}^N E_a^n(\Omega), \quad (7)$$

where N is the number of years in the averaging process. The interannual variability of the signal $E_t^n(\Omega)$ is described in terms of its standard deviation:

$$\sigma_{E_a(\Omega)} = \sqrt{\frac{1}{N-1} \sum_{n=1}^N (E_a^n(\Omega) - \bar{E}_a(\Omega))^2}. \quad (8)$$

The two quantities $\bar{E}_a(\Omega)$ and $\sigma_{E_a(\Omega)}$ characterize the climatology of the atmospheric waves occurring in the spectral subdomain Ω . If, for a given Ω , the confidence interval of two models, defined e.g. by $[\bar{E}_a(\Omega) - 2\sigma_{E_a(\Omega)}, \bar{E}_a(\Omega) + 2\sigma_{E_a(\Omega)}]$ do not overlap, we can say that the two time series of the corresponding waves are not statistically consistent.

It is possible to test the reliability of the Hayashi spectra of the reconstructed geopotential height and to estimate the essentially model-independent biases introduced by the geostrophic approximation by including in this study the analysis of the variability of the readily available 500 hPa geopotential height of the NCEP-NCAR and ECMWF reanalyses, thoroughly studied in Dell'Aquila et al. (2005).

All computations are implemented in a MATLAB[®] routine available at http://www.clima.casaccia.enea.it/staff/cal manti/ipcc_hayashi_v.m.

3 Results

The space-time spectra are computed for each of the 39 winters included in our datasets spanning the time frame 1962–2000. Figure 1a–d shows the various components of the 39-winters averages of the spectra computed in each DJF period from the *geostrophically reconstructed* 500 hPa geopotential height for the NCEP reanalysis dataset averaged over the latitudinal band 30°N – 75°N . Figure 1a shows the total power spectrum; Fig. 1b shows the power spectrum related to standing waves; Fig. 1c shows the power spectrum related to eastward propagating waves; Fig. 1d shows the spectrum of the westward propagating waves. The average spectrum for the eastward propagating waves has a sloping lobe which is the possible signature of a dispersion relation $\omega = \omega(k)$ defining the kinematic properties of the—mainly—baroclinic waves. Note that the spectral width of the lobe is mainly related to the fact that we actually observe waves that are characterized by a life-cycle including growth and decay besides simple propagation. We will deal elsewhere with a better spectral characterization of this type of signal. In contrast, the average spectrum of westward propagating variance, mainly due to long and low-frequency waves, does not feature a similarly clean and legible structure, suggesting that specific phenomena with well-defined propagation properties, such as the Branstator–Kushnir waves (Branstator 1987; Kushnir 1987; Branstator and Held 1995), may not be distinguished by this method.

For comparison, we show in Fig. 2 the Hayashi spectra computed with the 500 hPa geopotential height for the NCEP reanalysis dataset averaged over the

Fig. 1 Climatological average over 39 winters of Hayashi spectra for the geostrophically reconstructed 500 hPa geopotential height (relative to the latitudinal belt 30°N–75°N) from NCEP data. The Hayashi spectra have been obtained multiplying the spectra by $k\omega \cdot \tau/2\pi$. The units are m^2

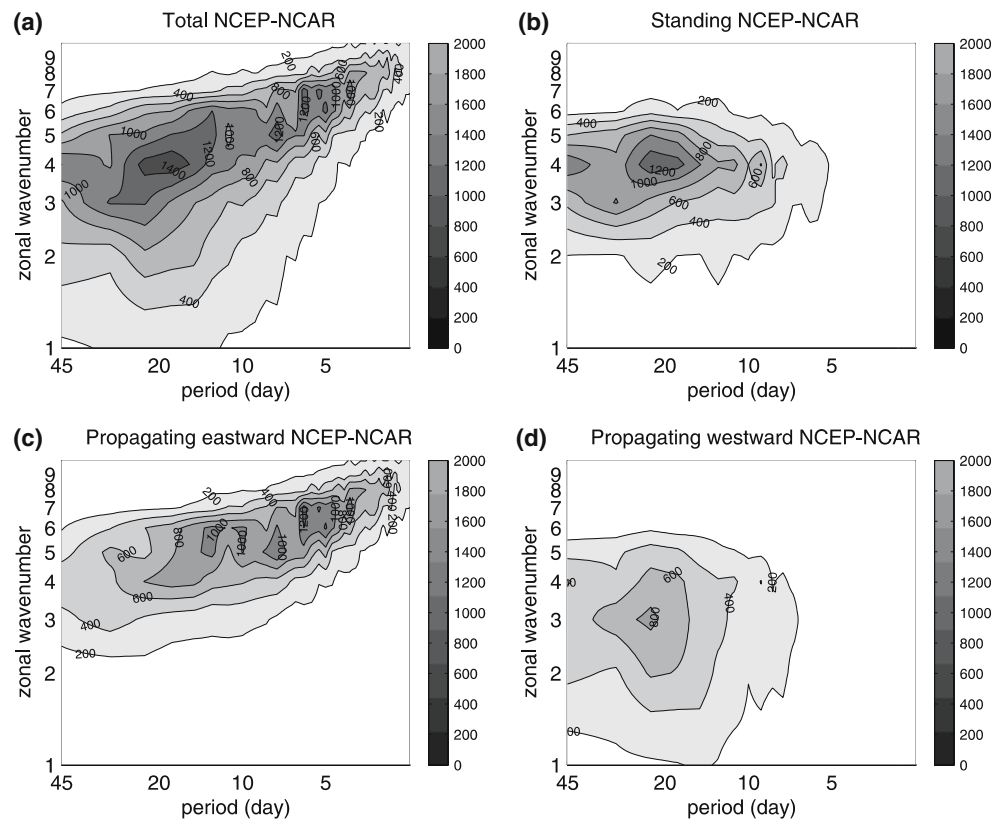


Fig. 2 Climatological average over 39 winters of Hayashi spectra for the 30°N–75°N averaged 500 hPa geopotential height from NCEP data. The Hayashi spectra have been obtained multiplying the spectra by $k\omega \cdot \tau/2\pi$. The units are m^2

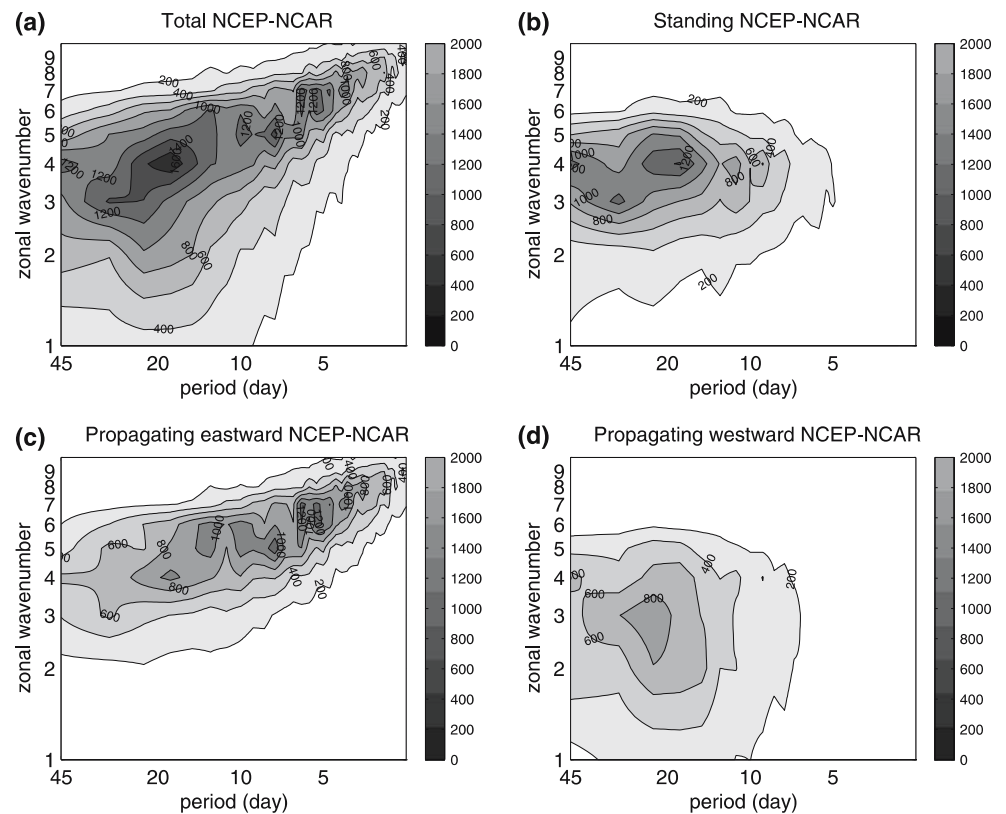
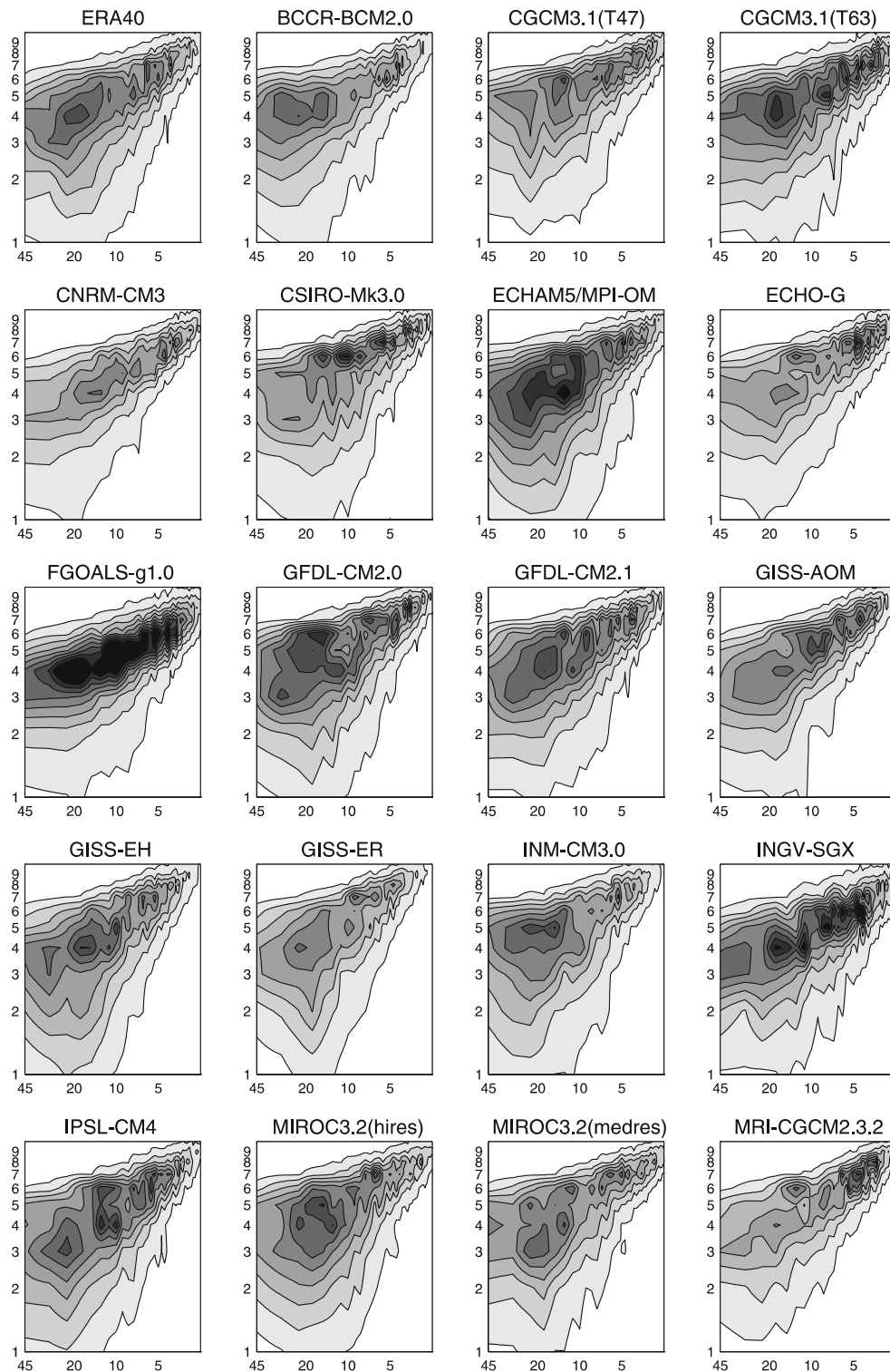


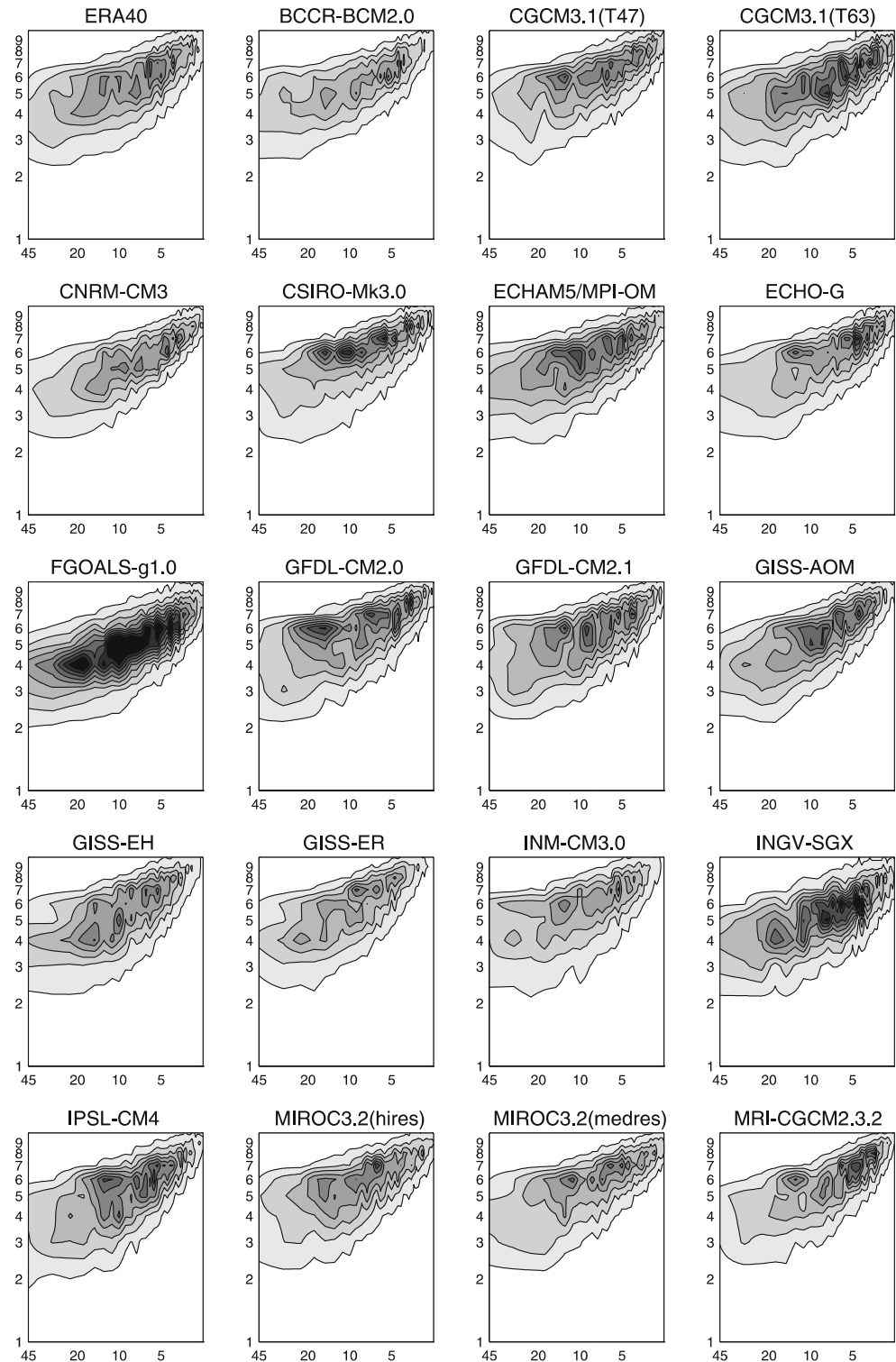
Fig. 3 Climatological averages over 39 winters of the total Hayashi spectra for the geostrophically reconstructed 30°N – 75°N averaged 500 hPa geopotential height from the ERA40 reanalysis and the 19 GCMs, as indicated in the panels. Hayashi spectra are multiplied by $k\omega \cdot \tau/2\pi$ for representation purposes (see text). Darker shades of grey denote higher values and isolines intervals are 200 m^2 . Period (in days) in abscissas and wavenumber in ordinates



latitudinal band 30°N – 75°N . A similar figure, related to the time-frame 1958–2002, has been already shown and discussed in Dell’Aquila et al. (2005). The Hayashi spectra of the 500 hPa geopotential height (Fig. 2) have consistently slightly higher values than those computed with the reconstructed field, thus suggesting

that the geostrophic reconstruction filters out some atmospheric variability, as can be reasonably expected. We obtain similar results by reconstructing geopotential height field from ERA40 reanalysis. However, the qualitative structure of the spectra is maintained. Since the effect of adopting the geostrophic approximation is

Fig. 4 As in Fig. 3 but for the eastward propagating component

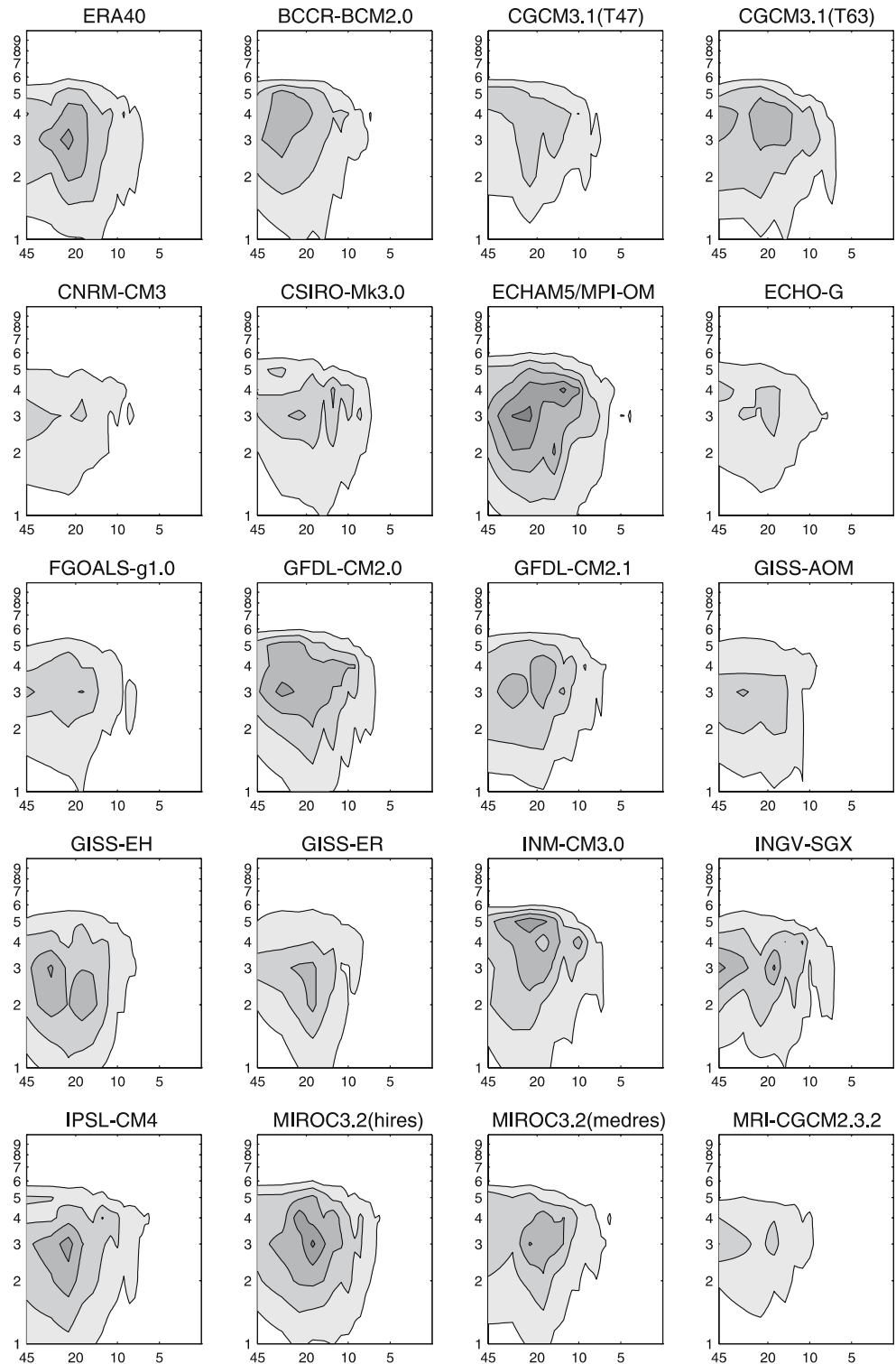


essentially model-independent, we assume our approach to be reasonable and consider the Hayashi spectra of the latitudinally averaged and geostrophically reconstructed 500 hPa geopotential height fields of the various models as good estimates for the spectra of the exact fields.

The 39 year-average of the winter Hayashi spectra of the IPCC GCMs reported in Table 1 are presented in Figs. 3, 4, 5 and 6.

At first glance, all models spectra are qualitatively similar to those of the NCEP and ERA40 datasets. In particular, a large portion of the total variance is con-

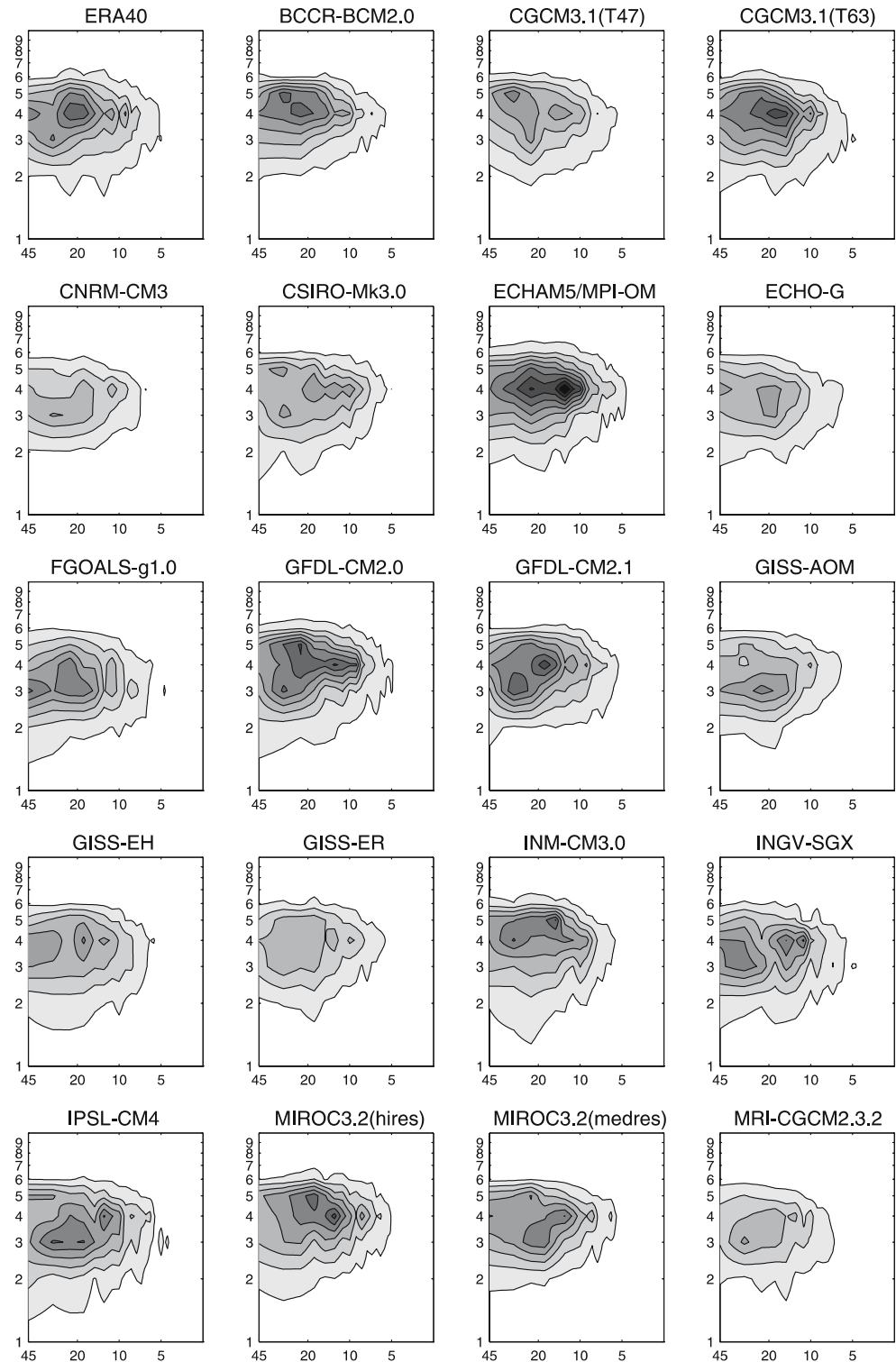
Fig. 5 As in Fig. 3 but for the westward propagating component



centrated in the LFLW domain, and can be related mostly to standing waves and to westward propagating waves. The HFHW domain, corresponding mainly to synoptic disturbances, contains a smaller portion of the total variance, and is almost exclusively related to eastward propagating waves.

Discrepancies stem from a closer comparison of Fig. 1 and Figs. 3, 4, 5 and 6. For example, some models, such as ECHAM5/MPI-OM and FGOALS-g1.0, tend to overestimate the atmospheric variability for the eastward propagating waves (Fig. 4). Other models [e.g. CGCM3.1 (T42), the CNRM-CM3, the

Fig. 6 As in Fig. 3 but for the standing component



CSIRO-Mk3.0, the ECHO-G, the entire family of GISS GCMs, the MRI-CGCM2.3.2 models)] underestimate the power spectrum of the atmospheric signal for the standing and westward propagating waves at most time and space scales (Figs. 5, 6). Focusing on the standing waves (Fig. 5) and looking in more detail at

the spectra, we find that only the IPSL-CM4, MIROC, and CGCM3.1(T63) models feature a peak at wave-number 4 and period of about 20 days, as the two reanalyses. This feature is not in agreement with the *common wisdom*, coming from the linear Rossby wave theory, that the low-frequency resonant wave has

wavenumber 3, thus suggesting the inadequacy of such a simple conceptual model when climatic properties are considered (Holton 1992). Of course, a rigorous statistical assessment of the significance of such spectral feature is required in order to draw more robust conclusions (see discussion below), but nevertheless it has to be noted that, since the geometrical domain is truly periodic, the assessment of the wavenumber decomposition of the field is more reliable than the frequency analysis.

It is clear that visual inspection of the Hayashi spectra, though instructive and useful, is not a viable way for objectively intercomparing the models and auditing their performances with respect to the reanalyses. For any nonparametric distribution, it is in general a hard task to deconvolve in a statistically rigorous way *real* signal (if any) from noise for a given 2D spectral density, i.e. to assess the significance of single peaks and dips, and it is still harder to assess whether two densities are statistically consistent. Note that for the computation of the space-time spectra shown in Figs. 1, 2, 3, 4, 5 and 6, some general hypothesis have been formulated as to the contribution of noise o the observed signal (Hayashi 1979). A possible strategy for assessing the robustness of the

observed spectral features has been discussed by Hansen et al. (1989), who compared different spectral decompositions and pointed out relevant discrepancies only for the standing component of the spectrum. Nevertheless, in this study, instead of focusing on the details of the spectra, we adopt a *coarse grain* approach and degrade the retrieved information contained in the Hayashi spectra by suitable averages, obtaining, as a *trade-off*, the possibility of objectively interpreting and statistically assessing the results. We choose this approach in order to fulfill one of the main requests of PCMDI, which is the provision of simple scalar metrics for the evaluation of model performances.

In the following sections, we summarize the most relevant information contained in each panel of Figs. 3, 4, 5 and 6 into a few well-defined numbers more robustly characterizing the skill of each model. We then consider the averaging procedure defined in Eq. 6, and we introduce two different but complementary approaches, the first one evaluating the overall performances of the various models at all time and space scales (*global metrics*), the second one assessing the abilities of the models in reproducing the correct features of the baroclinic and planetary waves (*process oriented metrics*).

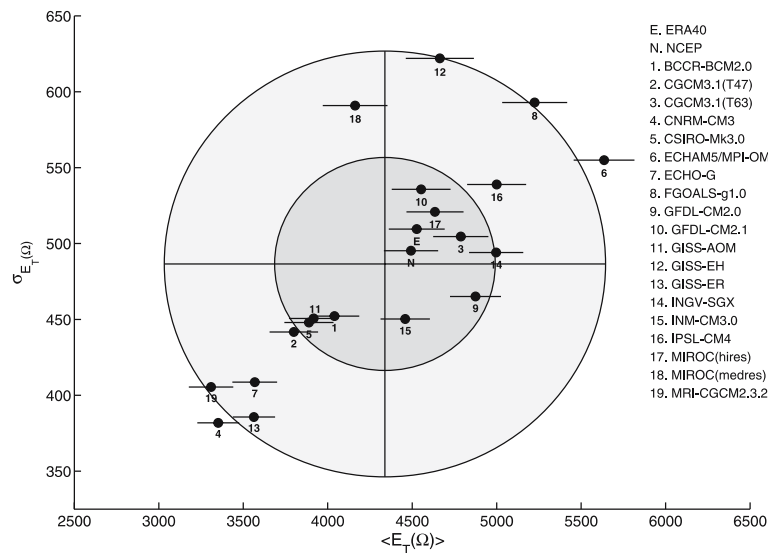


Fig. 7 Mean value $\bar{E}_T(\Omega)$ and interannual standard deviation variance $\sigma_{E_a(\Omega)}$ of the total spectrum integrated over the whole spectral domain for the 2 reanalyses and the 19 GCMs. The letters indicate the data computed from the NCEP-NCAR reanalysis (N), from the ERA40 reanalysis (E). The shaded areas represent the dispersion of data; the center of the ellipses is the ensemble average; the semi-axes of the inner ellipse are equal to the variance of data in the corresponding direction; the

semi-axes of the outer ellipse correspond to twice the variance. For each dot, the horizontal error bars represent the 95% confidence intervals of the estimate $\bar{E}_T(\Omega)$ and half-width is $2\sigma_{E_a(\Omega)}/\sqrt{n}$, where $n = 39$. The vertical error bar is not reported to maintain the readability of the figure. See Sect. 3.1 for a discussion of the statistical significance of the variance at 95% confidence level

3.1 Global metric

A global scalar metrics can be introduced by integrating power spectrum corresponding to each one of the 39 winter seasons over the full frequency–wavenumber domain. We then consider the quantity $E_T^n(\Omega)$ introduced in Eq. 6 where Ω is set to be the full frequency–wavenumber domain and n is the index running over the winters.

In Fig. 7 we show the climatological average $\bar{E}_T(\Omega)$ of the integral of the full spectrum versus its interannual variability $\sigma_{E_T(\Omega)}$, estimated according to Eqs. 7 and 8, respectively. Note that NAO effects are automatically included in the evaluation of the variability and should not be considered as special in any sense, for the purposes of this study. In this scatter diagram the abscissa represents the model average intra-seasonal variability, while the ordinate represents the interannual variability of the model intra-seasonal variability. The models ensemble average is also shown. In general, similar scatter diagrams, relating the two most relevant statistical moments of any chosen variable, are the simplest representation of a given climate. All of the time series $E_T^n(\Omega)$ from which we derive the statistical moments reported in Fig. 7 are compatible with the null hypotheses of white noise, featuring values of the lagged correlation smaller than 0.3 for all time lags equal or greater than 1 year. Therefore, for each model the confidence interval for the estimate of the climatological mean is centered on the quantity $\bar{E}_T(\Omega)$ (the estimator can be assumed to be unbiased) and its half-width can be approximated as $2\sigma_{E_T(\Omega)}/\sqrt{N} \approx 0.32\sigma_{E_T(\Omega)}$, thus being proportional to the ordinate of the corresponding representative point. If the confidence intervals of two models do not overlap, we can say that their climatological means are not statistically consistent. Note that if we had longer runs, we could restrict progressively the confidence interval of the climatological mean. Whereas the length of the considered simulations allows for a reliable statistical interpretation of the model discrepancies in the description of intra-seasonal variability (the x -axis of the scatter diagram), the same is not true for the interannual variability. In Fig. 7 we do not show the vertical error bar, because that would make the scatter diagram unreadable. However, by treating the variance of $\bar{E}_T(\Omega)$ as a χ^2 -distributed random variable with $N - 1$ degrees of freedom ($N = 39$ is the length of the time series) and by choosing a confidence level of 95%, we obtain a confidence interval spanning around 30% of the variance itself. Therefore, the length of the time series is not enough for a statistical assesment of the models

discrepancies in the description of midlatitude interannual variability.

The reconstructed latitudinally averaged 500 hPa geopotential height fields of the two reanalyses have similar intra-seasonal variability. If we consider the *observed* 500 hPa geopotential height fields (here not shown), in agreement with what results from the visual inspection of the Hayashi spectra, we have that in both cases the intraseasonal and interannual atmospheric variability is increased by a constant factor around 10%. The reliability of the approach of considering a geostrophically balanced reconstruction of the 500 hPa geopotential height field is then confirmed.

Apparently, if we consider the model ensemble (centre of the ellipses in Fig. 7), we have a good agreement with observations: both reanalyses lie well within 1 SD (inner ellipsis) from the ensemble average. However, the models are widely spread over the plane space spanned by the two variables, with a typical bias of about 15% with respect to the reanalyses.

In general, the biases on the intraseasonal and interannual variability are positively linearly correlated: for larger average signals the variability tends to be larger, so that the model spread in Fig. 7 is definitely not isotropic.

The models that are statistically consistent with the reanalyses in terms of the total wave activity are the high-resolution version of the MIROC, the GFDL-CM2.1 and the INM3.0 model. The T63 version of the CGCM3.2 also performs well, featuring a slightly too large intraseasonal variability. Some models, such as FGOALS1.0 and ECHAM5/MPI-OM, feature very large positive biases ranging over 20% for the intraseasonal variability. Some other models, usually of relatively low-resolution, such as CNRM-CM3 GISS-ER, ECHO-G and MRI-CGCM2.3.2, feature over 20% negative biases. The GISS-EH model has a very high value of interannual variability. Although the statistical significance is low (see discussion above) it is likely that this model has serious flaws in describing the atmospheric interannual variability at mid-latitudes correctly.

In some cases, it is possible to track the improvements occurring between different versions or set-ups of the same climate model. Thus, e.g. the GISS-EH model (which includes an isopycnal ocean component) has a better representation of the intra-seasonal atmospheric with respect to the GISS-ER which has a z -coordinate ocean model. However, the two models have considerably different interannual variability, which is in both cases different from that of the reanalysis. The CGCM3.1 model is presented in two versions which are identical except for the horizontal

resolution (T47, corresponding to about 3.75° resolution, and T63 corresponding to about 2.8° resolution). The CGCM3.1(T63) improves the representation of both the intra-seasonal variability and of the interannual variability with respect to CGCM3.1(T47), although the improvements concerning the interannual variability are not statistically significant. Similar improvements are observed between MIROC(hires) and MIROC(medres) and between GFDL-CM2.1 and GFDL-CM2.0. The MIROC model is presented with two different horizontal as well as vertical resolutions (T42L20 for the medres version and T106L56 for the hires version). The medres version shows already quite good performances and is among the best models. However, a substantial improvement is observed when switching to higher resolution. The GFDL-CM2.0 and GFDL-CM2.1 models are very similar. In particular they share the same horizontal as well as vertical resolution, but in the GFDL-CM2.1 model some numerical techniques are improved with respect to GFDL-CM2.0. For example, CM2.0 uses centred spatial differencing in the horizontal, whereas in CM2.1 the horizontal discretisation is performed with a flux-form semi-Lagrangian method. Finally, let us note that the models with a better representation of the intraseasonal variability and with the values of $\sigma_{E_T(\Omega)}$ closer to those of the reanalyses display also the most realistic ENSO variability as shown in Van Oldenburg et al. (2005). It is then likely that longer model simulations would not alter substantially the general picture sketched in Fig. 7, although they would allow for a more robust statistical assessment of the results concerning the models interannual variability.

We underline that for each dataset the time-series $E_T^n(\Omega)$ summarizes the combined effect of all types of atmospheric disturbances acting on all time and space scales in the winter n . Therefore, a metric based on the

Table 2 Definition of four regions in the Hayashi spectra of the winter atmospheric variability; the symbol d is used as shorthand for ‘day’

Spectral properties	$j_1 = 2, j_2 = 4$	$j_1 = 6, j_2 = j_{\max}$
$m_1 = 2, m_2 = 9$	$\Omega = \text{LFLW}$	$\Omega = \text{LFHW}$
$m_1 = 13, m_2 = 45$	$\Omega = \text{HFLW}$	$\Omega = \text{HFHW}$

Low frequency low wavenumber (LFLW); high frequency low wavenumber (HFLW); low frequency high wavenumber (LFHW); high frequency high wavenumber (HFHW). Low Frequency relates to periods from 10 to 45 days; High Frequency relates to periods from 2 to 7 days; Low Wavenumber relates to length scales larger than 1,000 km; High wavenumber relates to length scales ranging from a few tens to hundreds of kilometers. The values $m_2 = 45, j_2 = j_{\max}$ correspond to the Nyquist frequency and wavenumber for the model-dependent data resolution

statistical properties of $E_T^n(\Omega)$ cannot distinguish between noise and the contributions of waves of different natures due to different physical processes, and plain noise and merely gives (thanks to Parseval’s theorem, essentially) the total variability (or energy) of the signal $Z(p = 500 \text{ hPa}, \lambda, t)$. In order to gain more insight into how well the various models perform, the results presented in Fig. 7 must then be supplemented with the information provided by more specific and detailed (in a spectral sense) measures of the wave signal.

3.2 Process-oriented metrics

In order to construct *process-oriented* metrics, i.e. those pointing at the diagnosis of specific dynamical processes, we consider a decomposition of the frequency–wave number domain based on the known phenomenology of planetary scale atmospheric perturbations. Following Dell’Aquila et al. (2005), we consider two spectral subdomains, presented in Table 2.

The first is the HFHW subdomain, which includes periods from 2 to 7 days ($13 \leq m \leq 45$) and zonal wavenumbers $j \geq 6$ (length scales ranging from a few hundreds of kilometres to 5,000 km). These spectral components can be associated with the release of available energy driven by conventional baroclinic conversion (Blackmon 1976; Speranza 1983; Wallace et al. 1988), so that they are often referred to as baroclinic waves.

The second is the LFLW subdomain, which includes periods from 10 to 45 days ($2 \leq m \leq 9$) and zonal wavenumbers $2 \leq j \leq 4$ (length scales larger than 7,000 km). These waves are mostly generated by the flow interaction with orography (Charney and DeVore 1979; Charney and Straus 1980; Buzzi et al. 1984; Benzi et al. 1986) and are catalyzed by the sub-tropical jet (Benzi and Speranza 1989; Ruti et al. 2006). Note that the selection of the range of zonal wavenumbers corresponds to the definition of the Wave Amplitude Indicator (WAI), which has been used as a proxy for the analysis of the low frequency variability of the atmosphere (Benzi et al. 1986; Benzi and Speranza 1989; Ruti et al. 2006). The WAI provides a synthetic picture of the ultra-long planetary waves and includes the wavelengths that are suitable for orographic resonance, since an approximate mode of zero phase velocity (resonance) is around 3 (Malguzzi and Speranza 1981).

For each year of a given dataset, we then provide a bulk measure of the planetary standing waves and of the eastward propagating baroclinic waves by computing the quantities $E_E^n(\Omega_{\text{HFHW}})$ and $E_S^n(\Omega_{\text{LFLW}})$ as defined in Eq. 6, respectively, where the two

Ω -domains are prescribed as above, n is the index running over the winters, and the lower indexes S and E refer to standing and eastward propagating components, respectively. The quantities $E_E^n(\Omega_{\text{HFHW}})$ and $E_S^n(\Omega_{\text{LFLW}})$ can then be used to characterize the capabilities of each model of correctly describing two different dynamical processes of the atmosphere. Note that, starting from the analysis of the physical processes of the real atmosphere, the spectral subdomain employed for the definition of the scalar metric have been selected once for all models, in order assess how well the models compare to the reference provided by the two reanalyses.

We focus on the average (climatological) description provided by each model of the planetary standing variance and of the baroclinic eastward propagating variance. We then consider for each model the mean quantities $\bar{E}_E(\Omega_{\text{HFHW}})$ and $\bar{E}_S(\Omega_{\text{LFLW}})$. Similar to the case of the global metric, the time series $E_E^n(\Omega_{\text{HFHW}})$ and $E_S^n(\Omega_{\text{LFLW}})$ are statistically consistent with a white noise model (no significant internannual lagged correlation) for all models as well for the two reanalyses. Therefore, the standard error associated with the respective metrics, i.e. the quantities $\sigma_{E_E(\Omega_{\text{HFHW}})}/\sqrt{N}$ and $\sigma_{E_S(\Omega_{\text{LFLW}})}/\sqrt{N}$, is a good estimate for the uncertainty associated with the climatological mean of the corresponding atmospheric variance.

For all models and for the reanalyses the cross-correlation of the time series $E_E^n(\Omega_{\text{HFHW}})$ and $E_S^n(\Omega_{\text{LFLW}})$ is not significantly different from zero, so that they can be considered, at least approximately, as independent variables. Moreover, for all datasets, we have that the two time series $E_E^n(\Omega_{\text{HFHW}})$ and $E_T^n(\Omega)$ are essentially uncorrelated, while $E_S^n(\Omega_{\text{LFLW}})$ and

$E_T^n(\Omega)$ have statistically significant positive correlation whose value is around 0.5. This implies that, for all models, whereas a certain signature of the total wave activity signal is time-wise retained in the standing variance the contribution of the baroclinic high-frequency variance seems to be somehow decoupled.

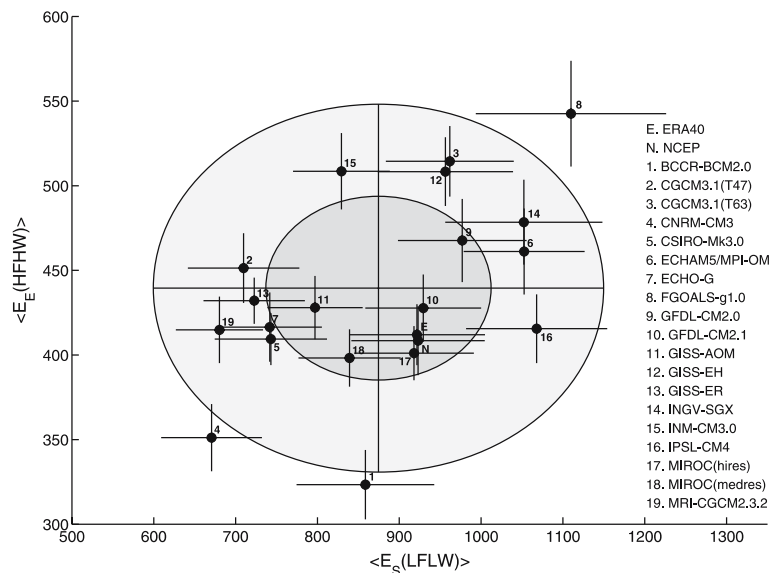
In Fig. 8 we show for all datasets the climatological average $\bar{E}_S(\Omega_{\text{LFLW}})$ versus the climatological average $\bar{E}_E(\Omega_{\text{HFHW}})$, and indicate for both directions the corresponding standard error.

When considering the ERA40 and NCEP reanalyses datasets, we observe that the two reanalyses are virtually indistinguishable, within the confidence level provided by the error bars, and they could be brought into even closer agreement by dropping the pre-satellite period (Dell'Aquila et al. 2005). However, the discrepancies among IPCC models by far exceed the discrepancies between the two different *dynamical interpolations* of the available observations.

Such a *process oriented* metrics is complementary to the global metrics presented in the previous subsection, since it is less prone to the limitation that the errors originating in different regions of the spectrum may average out and hide major model deficiencies. Instead, considering the *process oriented* metrics, it is easier to bring the models biases into light.

A general (with very few exceptions) occurrence for all IPCC models is that they overestimate the variability in the HFHW subdomain. Regarding the LFLW subdomain, more than half of the models tend to underestimate the corresponding variability. As an overall result, the model ensemble (center of the ellipses in Fig. 8) overestimates the HFHW variability and underestimates the LFLW variability. This closely

Fig. 8 Scatter plot of $\bar{E}_S(\Omega_{\text{LFLW}})$, mean value of the integral over the low-frequency low-wavenumber spectral subdomain of the standing waves (abscissas), versus $\bar{E}_E(\Omega_{\text{HFHW}})$, mean value of the integral over the high-frequency high-wavenumber subdomain (ordinates) of the eastward propagating waves. For each dot the horizontal (vertical) error bar gives the 95% confidence level of the estimate $2\sigma_{E_S(\Omega_{\text{LFLW}})}/\sqrt{n}$ where $n=39$. Shading is analogous to Fig. 7



resembles the results of many diagnostic studies performed in the past on global weather forecast models (Tibaldi 1986), which are the ancestors of the atmospheric components of GCMs.

Moreover, if we consider the confidence intervals of the statistical estimator of the climatological means we again have that very few models are statistically consistent with the reanalyses with a reasonable significance. The models that are closer to the reanalyses are the MIROC (hires and medres) model and the GFDL2.1 model. In these cases, the error bars have relevant overlaps with the area defined by the error bars of the NCEP and ERA40 reanalyses. It is rather interesting to note that not only the MIROC(hires) model and the GFDL2.1 model are the best in both the global and the process-oriented metrics, but they are also the only two with similar performances in both metrics: some models that perform well in the global metric perform poorly in the process-oriented ones, and *vice versa*.

By using the process-oriented metrics, the models having the worst performances are the CNRM-CM3 model, which underestimates both the low frequency and the high frequency, the FGOALS model, which overestimates the variability in both spectral sub-domains, and the BCCR-BCM model, which seriously underestimates the high-frequency variability.

We also underline that the results of the process-oriented metrics are quite robust with respect to slightly changing the spectral subdomain, so that when perturbing the boundaries of Ω_{HFHW} and Ω_{LFLW} , the resulting picture is similar to Fig. 8. This rules out the possibility that our results might be strongly affected by noisy boundary effects in the spectral filtering.

Unfortunately, there is no unique way to attribute the biases of LFLW and HFHW variability to specific features that are common to a class of models. Instead, it is interesting to compare the relative changes in the performances of models presented with different configuration. This might help in the identification of specific strategies for model improvements.

In the case of the CGCM3.1 models, an increased horizontal resolution (T63) leads to a better statistics of planetary-scale standing waves. However, the performance on small scale disturbances is better in the low resolution version (T47). Thus, increasing the horizontal resolution alone leads to no automatic overall substantial improvement of the model performance. A further example in the same direction is the comparison of the models ECHAM5/MPI-OM and INGV-SXG, that share the same atmospheric component with different horizontal and vertical resolution. In this case, the model with higher horizontal resolu-

tion (INGV-SXG, T106 instead of T63) but lower vertical resolution (L19 instead of L31), has worse performance with regard to eastward propagating baroclinic waves, while retaining similar performance with regard to standing planetary waves.

As mentioned above, two of the versions of the GISS model share the same atmosphere component over different oceans. In this case, the use of vertical density coordinates (GISS-EH) appears to improve the statistics of the standing waves, whose level is too low in the case of z -coordinates (GISS-ER). Instead, the performance with regard to eastward propagating waves is worse.

In the case of the MIROC model, increased resolution alone improves the model performance, but in this case both the horizontal *and* the vertical resolutions are increased.

Also in the context of the process-oriented metrics, the improvement from GFDL-CM2.0 to GFDL-CM2.1 is substantial, with the latter in good agreement with the reanalyses. The experience with the GFDL models demonstrates how the increase of computer power (i.e. of resolution) may not be the only pathway to model improvements. Instead, the employment of more accurate discretisation techniques may make a substantial contribution.

4 Conclusions

This study wishes to propose possible strategies and actual tools for auditing climate models. In particular, we have focused on assessing the degree of agreement with the reanalyses, and the degree of mutual coherence, of some of the most well-known coupled GCMs in the description of the statistical properties of the atmospheric disturbances in the mid-troposphere. We maintain that such an analysis provides more insight into the performances of the coupled GCMs than the comparison of properties such as surface temperature, because the internal mechanisms of the atmosphere are taken into consideration.

We have considered the variability of the 500 hPa geopotential height field, derived from the NCEP and ERA40 reanalyses for the time frame 1962–2000 and in the twentieth century control run of the IPCC GCMs. We compute the Hayashi spectra of the 500 hPa geopotential height fields and introduce an integral measure of the variability observed in the NH on different spectral sub-domains. The total wave variability is taken as a *global metric* describing the overall performance of each model, while the variability related to the eastward propagating baroclinic waves and to the

standing planetary waves are taken as *process-oriented metrics*, aimed at measuring the model capability of describing the corresponding physical process.

In such a context, we obtain the striking result that large biases, in most cases larger than 15%, are found in the metrics between the atmospheric waves climatology of most IPCC models and the reanalyses. In all cases, the span of the climatologies of the various models is around 50% of the climatology of the reanalyses. In particular, when considering the total variability of the wave fields of the GCMs, we find that the biases on the intraseasonal and interannual variability are positively linearly correlated, i.e. for larger average signals the variability tends to be larger. When considering the process-oriented metrics, we have the confirmation that the baroclinic waves are typically overestimated by the climate models, while the planetary waves are usually underestimated. This closely resembles the results of many diagnostic studies performed in the past on global weather forecast models (Tibaldi 1986). The model ensemble obtained by arithmetic averaging the results of all models is, for both the global and process oriented metrics, biased with respect to the reanalyses, but is comparable to the best 5 models. Nevertheless, the models results do not cluster around their ensemble mean. Only for two models—GFDL-CM2.1 and MIROC(hires)—the climatologies of the wave activity are statistically consistent with those of the reanalyses, both for the global and process-oriented metrics. The fact that these two models are the only two performing well in both metrics suggest that they outperform the other models in the representation of the mid-latitude winter atmospheric dynamics of the NH. The span of performance of the GCMs in the representation of the tropospheric waves is a rather relevant issue, particularly because all models are expected to agree reasonably in other diagnostics such as the average vertical temperature profiles, average surface temperature, precipitation and so on (see e.g. the results presented in the TAR).

The general pictures obtained with the *global* and the *process-oriented* metrics, are essentially different, as most of the models showing good agreement with reanalyses for one metric do not perform similarly well for the other one.

In particular, the substantial improvement detected in the *global metric* for a higher resolution version of a given model may indeed mask the minor improvements (if any) in describing a specific process, as in the case of the two CGCM3.1 GCMs. Also, the INM-CM3.0 model, which seems rather close to observations when using the *global metric*, fails to correctly describe both the LFLW and HFHW regions of the

spectrum of mid-latitude atmospheric variability. The two models ECHAM5/MPI-OM and INGV-SXG share the same atmospheric component with different horizontal and vertical resolution. In this case, on the other hand, the model with higher horizontal resolution (INGV-SXG, T106 instead of T63) but lower vertical resolution (L19 instead of L31), has worse performances on eastward propagating baroclinic waves, while retaining a similar performance on standing planetary waves, but is largely better for the global metrics.

The outputs of the *process-oriented metrics* may suggest some of the priorities for planning strategies for model improvements. From this perspective, we find that the increase of horizontal resolution alone has no substantial effect on our estimate of model performances, while the increase of horizontal *and* vertical resolution brings (e.g. MIROC(hires)) into rather close agreement with observations. An increased vertical resolution could be useful to better mimic the vertical structure of the ultra-long waves, in particular the orographic baroclinic standing perturbations (Buzzi et al. 1984). The numerical scheme improvement also has a positive impact on model performance (GFDL models). In particular, the use of semi-Lagrangian advection schemes for tracers seems to be an important requirement for model reliability.

Somewhat unexpectedly, in the case of the GISS-ER and GISS-EH models, the characteristics of the adopted ocean model also seems to play a critical role in determining the bulk of the atmospheric variability. We may guess that what could be relevant in this case might be the differences in the representation of the surface ocean temperature in the Northern Atlantic region, which delicately depends on the representation of the thermohaline circulation. Among the three GISS models, GISS-AOM displays superior performances.

In order to obtain more detailed information regarding the representation of the physical processes occurring in the different spectral regions, the process-oriented metrics proposed in this work should be complemented with a careful investigation in the space-time domain. Such an analysis should look at large scale horizontal spatial structures and wave source regions, wave dynamics, and may determine that, for example, a model populates correctly a considered spectral subdomain, but for the wrong physical reasons. The statistics of mid-latitude cyclones occurrences and intensities is a good example of this kind of *dual* analysis with respect to what is proposed in this paper.

Moreover, since the adopted approach requires a latitudinal averaging operation, such that all the lati-

tudinal features of the fields are lost, we foresee the necessity of a more detailed—more differentiated in latitude - analysis in order to more thoroughly evaluate the models.

This study, although preliminary, suggests caveats concerning the ability of most of the presently available climate models in describing the statistical properties of the global scale atmospheric dynamics of the present climate and, *a fortiori*, in the perspective of climate change. One of the possible outcomes of this study may be the provision of quantitative information needed to *weight* model reliability when considering models ensemble results, e.g. in the context of the IPCC reports. A positive result of this study is that, in spite of all the (sometimes rather large) biases between the datasets in the climatological means of the signals $E_E^n(\Omega_{\text{HFHW}})$, $E_S^n(\Omega_{\text{LFLW}})$, and $E_T^n(\Omega)$, corresponding to the activity of the respective class of waves, some statistical properties are well represented in all models. In particular, for all models we have that all the signals considered have rather weak time-lagged self-correlation, that the cross-correlation of $E_E^n(\Omega_{\text{HFHW}})$ and $E_S^n(\Omega_{\text{LFLW}})$ is rather weak, and that, whereas the correlation of $E_S^n(\Omega_{\text{LFLW}})$ and $E_T^n(\Omega)$ is positive and statistically significant, the correlation of $E_E^n(\Omega_{\text{HFHW}})$ and $E_T^n(\Omega)$ is rather weak, so that the baroclinic waves have, as opposed to the planetary standing waves, a very weak signature in terms of the total wave activity.

On the other hand, the GFDL-CM2.1 and MIROC(hires) models, being able to reproduce with some degree of confidence the statistical properties of wave activity in the atmosphere, seem to be the best candidates for more detailed studies on atmospheric circulation regimes (Ruti et al. 2006), which will be the subject of future study. Among the several other in-depth analyses which can follow from the results presented here, we would like to mention three future paths. In the context of the global properties of the atmosphere, it seems relevant to study the links between the degree of the models mutual coherence and realism in the description of the mid-latitudes atmospheric variability, and in the representation of ENSO (Van Oldenburg et al. 2005), which seems critical in the set-up of the regimes of the low-frequency mid-latitudes planetary waves (Ruti et al. 2006). In the context of the understanding of climate change, it seems relevant to study the mutual coherence of the GCMs with regard to the statistics of the atmospheric waves considered in the present analysis in their sensitivity to CO² doubling. We also wish to mention that performing a similar analysis of the Southern Hemisphere mid-latitudes faces the problem of the large

disagreement between the two reanalyses (Dell'Aquila et al. 2006).

A further, more comprehensive, thermodynamically oriented, intercomparison study can be performed, along some of the lines proposed in the WCRP/WGNE Workshop on the Second Phase of the Atmospheric Model Intercomparison Project (Gleckler 2004), by analyzing the Lorenz cycle and estimating the efficiency of the conversion of the available potential energy of the atmosphere in the various climate models. We believe that this might be a physically robust diagnostic to be used in the study of the climate machine and in the climate models quality-check.

The computations described in this paper are implemented in a MATLAB[®] routine available at http://www.clima.casaccia.enea.it/staff/cal manti/ipcc_hayashi_v.m.

Acknowledgments We acknowledge the international modeling groups for providing their data for analysis, the Program for Climate Model Diagnosis and Intercomparison (PCMDI) for collecting and archiving the model data, the JSC/CLIVAR Working Group on Coupled Modeling (WGCM) and their Coupled Model Intercomparison Project (CMIP) and Climate Simulation Panel for organizing the model data analysis activity, and the IPCC WG1 TSU for technical support. The IPCC Data Archive at Lawrence Livermore National Laboratory is supported by the Office of Science, U.S. Department of Energy. A special thank to George Boer, Seita Emori, Silvio Gualdi, Kenneth Lo, Gavin Schmidt and Daniel Robitaille for supplying some of the datasets that, due to technical problems, were not available on the on the PCMDI servers when we were working on this paper. The authors are also grateful to F.W. Gerstengarbe, U. Böhm, M. Claussen, R. Pielke Sr., G. Browning, K. Fraederich, G. Russell, A. Tsonis, B. Hoskins, and L. Smith for encouragement and useful comments. We acknowledge the useful comments of two anonymous referees. V.L. acknowledges the support of ISAC-CNR and CINFAI (HYDROCARE project).

References

- Benzi R, Speranza A (1989) Statistical properties of low frequency variability in the Northern Hemisphere. *J Clim* 2:367–379
- Benzi R, Malguzzi P, Speranza A, Sutera A (1986) The statistical properties of general atmospheric circulation: observational evidence and a minimal theory of bimodality. *Q J R Meteorol Soc* 112:661–674
- Betts AK (1986) A new convective adjustment scheme. Part I. Observational and theoretical basis. *Q J R Met Soc* 112:677–691
- Blackmon ML (1976) A climatological spectral study of the 500 mb geopotential height of the Northern Hemisphere. *J Atmos Sci* 33:1607–1623
- Bony S, Emanuel KA (2001) A parameterization of the cloudiness associated with cumulus convection; evaluation using TOGA COARE data. *J Atmos Sci* 58:3158–3183
- Bougeault P (1985) A simple parameterization of the large-scale effects of cumulus convection. *Monsoon Weather Rev* 113:2108–2121

- Branstator G (1987) A striking example of the atmospheres leading traveling pattern. *J Atmos Sci* 44:2310–2323
- Branstator G, Held IM (1995) Westward propagating normal modes in the presence of stationary background waves. *J Atmos Sci* 52:247–262
- Buzzi A, Trevisan A, Speranza A (1984) Instabilities of a baroclinic flow related to topographic forcing. *J Atmos Sci* 41:637–650
- Charney JG, DeVore JG (1979) Multiple flow equilibria in the atmosphere and blocking. *J Atmos Sci* 36:1205–1216
- Charney JG, Straus DM (1980) Form-drag instability, multiple equilibria and propagating planetary waves in the baroclinic, orographically forced, planetary wave system. *J Atmos Sci* 37:1157–1176
- Dell'Aquila A, Lucarini V, Ruti PM, Calmanti S (2005) Hayashi spectra of the northern hemisphere mid-latitude atmospheric variability in the NCEP-NCAR and ECMWF reanalyses. *Clim Dyn* 25:639–652, DOI:10.1007/s00382-005-0048-x
- Dell'Aquila A, Calmanti S, Ruti PM, Lucarini V (2006) Southern hemisphere mid-latitude atmospheric variability of the NCEP-NCAR and ECMWF reanalyses. *J Geophys Res* (in press)
- Delworth TL, Broccoli AJ, Rosati A, Stouffer RJ, Balaji V, Beesley JA, Cooke WF, Dixon KW, Dunne J, Dunne KA, Durachta JW, Findell KL, Ginoux P, Gnanadesikan A, Gordon CT, Gries SM, Gudgel R, Harrison MJ, Held IM, Hemler Rsikan S, Horowitz LW, Klein SA, Knutson TR, Kushner PJ, Langenhorst AR, Lee HC, Lin SJ, Lu J, Malyshev SL, Milly PCD, Ramaswamy V, Russell J, Schwarzkopf MD, Shevliakova E, Sirutis JJ, Spelman MJ, Stern WF, Winton M, Wittenberg AT, Wyman B, Zeng F, Zhang R (2005) GFDL's CM2 global coupled climate models – Part 1: formulation and simulation characteristics. *J Clim* 19:643–674
- Fraedrich K, Bottger H (1978) A wavenumber frequency analysis of the 500 mb geopotential at 50°N. *J Atmos Sci* 35:745–750
- Furevik T, Bentsen M, Drange H, Kindem IKT, Kvamsto NG, Sorteberg A, (2003) Description and evaluation of the Bergen climate model: ARPEGE coupled with MICOM. *Clim Dyn* 21:27–51
- Gleckler PJ (Ed) (2004) Proceedings of the WCRP/WGNE Workshop on the Second Phase of the Atmospheric Model Intercomparison Project (AMIP2), Meteo-France, Toulouse, pp 11–14 November 2002, August 16, 2004, UCRL-PROC-209115)
- Gordon HB, Rotstayn LD, McGregor JL, Dix MR, Kowalczyk EA, O'Farrell SP, Waterman LJ, Hirst AC, Wilson SG, Collier MA, Watterson IG, Elliott TI (2002) The CSIRO Mk3 Climate System Model (Electronic publication). Aspendale: CSIRO Atmospheric Research. (CSIRO Atmospheric Research technical paper; no. 60). 130 pp. (http://www.dar.csiro.au/publications/gordon_2002a.pdf)
- Grandpeix J-Y, Phillips V, Tailleux R (2004) Improved mixing representation in Emanuel's convection scheme. *Q J R Meteorol Soc* 130:3207–3222
- Gregory D, Rowntree PR (1990) A mass flux convection scheme with representation of cloud ensemble characteristics and stability dependent closure. *Monsoon Weather Rev* 118:1483–1506
- Gualdi S., Scoccimarro E, Navarra A (2006) Changes in Tropical Cyclone Activity due to Global Warming: Results from a High-Resolution Coupled General Circulation Model, submitted to *J. Climate*
- Hack JJ (1994) Parameterization of moist convection in the NCAR Community Climate Model (CCM2). *J Geophys Res* 99:5551–5568
- Hansen AR, Sutera A (1986) On the probability density distribution of planetary-scale atmospheric wave amplitude. *J Atmos Sci* 43:3250–3265
- Hansen AR, Sutera A, Venne DE (1989) An examination of midlatitude power spectra: evidence for standing variance and the signature of El Niño. *Tellus* 41(A):371–384
- Hayashi Y (1971) A generalized method for resolving disturbances into progressive and retrogressive waves by space Fourier and time cross-spectral analysis. *J Meteorol Soc Jap* 49:125–128
- Hayashi Y (1979) A generalized method for resolving transient disturbances into standing and travelling waves by space-time spectral analysis. *J Atmos Sci* 36:1017–1029
- Holton JR (1992) *An Introduction to Dynamic Meteorology*. Academic, San Diego
- Jungclaus J, Botzet M, Haak H, Keenlyside N, Luo J-J, Latif M, Marotzke J, Mikolajewicz U, Roeckner E (2006) Ocean circulation and tropical variability in the AOGCM EC-HAM5/MPI-OM. *J Clim* 19:3952–3972
- K-1 model developers (2004) K-1 coupled model (MIROC) description, K-1 technical report, 1, Hasumi H, Emori S (eds.), Center for Climate System Research, University of Tokyo, 34pp
- Kiehl JT, Hack JJ, Bonan G, Boville BA, Williamson D, Rasch P (1998) The National Center for Atmospheric Research Community Climate Model: CCM3. *J Clim* 11:1131–1149
- Kim S-J, Flato GM, Boer GJ, McFarlane NA (2002) A coupled climate model simulation of the Last Glacial Maximum, Part 1: transient multi-decadal response. *Clim Dyn* 19:515–537
- Kistler R, Kalnay E, Collins W, Saha S, White G, Woollen J, Chelliah M, Ebisuzaki W, Kanamitsu M, Kousky V, van den Dool H, Jenne R, Fiorino M (2001) The NCEP-NCAR 50-year reanalysis: monthly means CD-ROM and documentation. *Bull Amer Meteor Soc* 82:247–267
- Klinker E, Capaldo M (1986) Systematic errors in the baroclinic waves of the ECMWF model. *Tellus* 38A:215–235
- Kushnir Y (1987) Retrograding wintertime low-frequency disturbances over the north Pacific-ocean. *J Atmos Sci* 44:2727–2742
- Le Treut H, Li X-Z (1991) Sensitivity of an atmospheric general circulation model to prescribed SST changes: feedback effects associated with the simulation of cloud optical properties. *Clim Dyn* 5:175–187
- Lohmann U, Roeckner E (1996) Design and performance of a new cloud microphysics parameterization developed for the ECHAM4 general circulation model. *Clim Dyn* 12:557–572
- Lucarini V (2002) Towards a definition of climate science. *Int J Environ Pollut* 18:409–414
- Lucarini V, Russell GL (2002) Comparison of mean climate trends in the northern hemisphere between National Centers for Environmental Prediction and two atmosphere-ocean model forced runs. *J Geophys Res* 107 (D15), 10.1029/2001JD001247
- Malguzzi P, Speranza A (1981) Local Multiple Equilibria and Regional Atmospheric Blocking. *J Atmos Sci* 9:1939–1948
- Marti O, et al (2005) The new IPSL climate system model: IPSL-CM4, Tech. rep., Institut Pierre Simon Laplace des Sciences de l'Environnement Global, IPSL, Case 101, 4 place Jussieu, Paris, France
- May W (1999) Space-time spectra of the atmospheric intraseasonal variability in the extratropics and their dependency on

- the El Niño/Southern Oscillation phenomenon: model versus observation. *Clim Dyn* (1999) 15:369–387
- Min S-K, Legutke S, Hense A, Kwon W-T (2005) Internal variability in a 1000-year control simulation with the coupled climate model ECHO-G. Part I. Near-surface temperature, precipitation and mean sea level pressure. *Tellus* 57A:605–621
- Moorthi S, Suarez MJ (1992) Relaxed Arakawa-Schubert: A parameterization of moist convection for general circulation models. *Monsoon Weather Rev* 120:978–1002
- Nordeng TE (1994) Extended versions of the convective parameterization scheme at ECMWF and their impact on the mean and transient activity of the model in the tropics. Technical Memorandum No. 206, European Centre for Medium-Range Weather Forecasts, Reading, United Kingdom
- Pan D-M, Randall DA (1998) A Cumulus Parameterization with a Prognostic Closure. *Q J R Meteorol Soc* 124:949–981
- Peixoto JP, Oort AH (1992) *Physics of climate*, Chap. 1. American Institute of Physics, New York
- Randall DA, Pan D-M (1993) Implementation of the Arakawa-Schubert cumulus parameterization with a prognostic closure. In: Emanuel K, Raymond D (eds) *Cumulus parameterization, a meteorological monograph*. American Meteorological Society, pp 137–144
- Rasch PJ, Kristjánsson JE (1998) A comparison of the CCM3 model climate using diagnosed and predicted condensate parameterizations. *J Clim* 11:1587–1614
- Ricard J-L, Royer J-F (1993) A statistical cloud scheme for use in a AGCM. *Ann Geophysicae* 11:1095–1115
- Rotstayn LD (1997) A physically based scheme for the treatment of stratiform clouds and precipitation in large-scale models. I: Description and evaluation of the microphysical processes. *Q J R Meteorol Soc* 123:1227–1282
- Rotstayn LD (2000) On the “tuning” of the autoconversion parameterizations in climate models. *J Geophys Res* D105:15495–15507
- Ruti PM, Lucarini V, Dell’Aquila A, Calmanti S, Speranza A (2006) Does the subtropical jet catalyze the midlatitude atmospheric regimes? *Geophys Res Lett* 33, L06814, doi:10.1029/2005GL024620
- Salas-Méllia D, Chauvin F, Déqué M, Douville H, Gueremy JF, Marquet P, Planton S, Royer JF, Tyteca S (2005) Description and validation of the CNRM-CM3 global coupled model, submitted to *Clim Dyn*
- Schmidt GA, Ruedy R, Hansen JE, Aleinov I, Bell N, Bauer M, Bauer S, Cairns B, Canuto V, Cheng Y, DelGenio A, Faluvegi G, Friend AD, Hall TM, Hu Y, Kelley M, Kiang NY, Koch D, Lacis AA, Lerner J, Lo KK, Miller RL, Nazarenko L, Oinas V, Perlwitz J, Perlwitz J, Rind D, Romanou A, Russell GL, Sato M, Shindell DT, Stone PH, Sun S, Tausnev N, Thresher D, Yao M-S (2005) Present day atmospheric simulations using GISS Model E: Comparison to in-situ, satellite and reanalysis data. *J Clim* 19:153–192
- Siegmund P (1995) The generation of available potential energy: a comparison of results from a general circulation model with observations. *Clim Dyn* 11:129–140
- Simmons AJ, Gibson JK (2000) The ERA-40 Project Plan, ERA-40 Project Report Series No. 1, European Centre for Medium-Range Weather Forecasts, Reading
- Speranza A (1983) Deterministic and statistical properties of the westerlies. *Paleogeophysics* 121:511–562
- Sumi A, Kanamitsu M (1984) A study of systematic errors in a numerical prediction model, Part I: General aspects of the systematic errors and their relation with the transient eddies. *J Met Soc Japan* 62:234–251
- Tibaldi S (1986) Envelope orography and maintenance of the quasi-stationary circulation in the ECMWF global models. *Adv Geophys* 29:339–373
- Tiedke M (1993) Representation of clouds in large-scale models. *Mon Wea Rev* 121:3040–3061
- Van Oldenburg GJ, Philip S, Collins M (2005) El Niño in a changing climate: a multi-model study. *Ocean Science, Discussions* 2:267–298
- Volodin EM, Diansky NA (2004) El-Niño reproduction in coupled general circulation model of atmosphere and ocean. *Russ Meteorol Hydrol* 12:5–14
- Wallace JM, Tibaldi S, Simmons AJ (1983) Reduction of systematic forecast errors in the ECMWF model through the introduction of an envelope orography. *Q J R Meteorol Soc* 109:683–717
- Wallace JM, Lim GH, Blackmon ML (1988) Relationship between cyclone tracks, anticyclone tracks and baroclinic waveguides. *J Atmos Sci* 45:439–462
- Yu Y, Zhang X, Guo Y (2004) Global coupled ocean-atmosphere general circulation models in LASG/IAP. *Adv Atmos Sci* 21:444–455
- Yukimoto S, Noda A (2002) Improvements of the Meteorological Research Institute Global Ocean–atmosphere Coupled GCM (MRI-CGCM2) and its climate sensitivity, Tech. Rep. 10, NIES, Japan
- Yukimoto, et al, Noda A, Kitoh A, Sugi M, Kitamura Y, Hosaka M, Shibata K, Maeda S, Uchiyama T (2001) The new Meteorological Research Institute coupled GCM (MRI-CGCM2) – Model climate and variability - Pap. Meteorol Geophys 51, 47–88. *Meteor Geophys* 51:47–88
- Zhang GJ, McFarlane NA (1995) Sensitivity of climate simulations to the parameterization of cumulus convection in the CCC-GCM. *Atmos-Ocean* 3:407–446