An Assessment on the Performance of IPCC AR4 Climate Models in Simulating Interdecadal Variations of the East Asian Summer Monsoon*

SUN Ying†(孙 颖) and DING Yihui (丁一汇)

National Climate Center, China Meteorological Administration, Beijing 100081

(Received September 28, 2008)

ABSTRACT

Observations from several data centers together with a categorization method are used to evaluate the IPCC AR4 (Intergovernmental Panel on Climate Change, the Fourth Assessment Report) climate models' performance in simulating the interdecadal variations of summer precipitation and monsoon circulation in East Asia. Out of 19 models under examination, 9 models can relatively well reproduce the 1979−1999 mean June-July-August (JJA) precipitation in East Asia, but only 3 models (Category-1 models) can capture the interdecadal variation of precipitation in East Asia. These 3 models are: GFDL-CM2.0, MIROC3.2 (hires), and MIROC3.2 (medres), among which the GFDL-CM2.0 gives the best performance. The reason for the poor performance of most models in simulating the East Asian summer monsoon interdecadal variation lies in that the key dynamic and thermal-dynamic mechanisms behind the East Asian monsoon change are missed by the models, e.g., the large-scale tropospheric cooling and drying over East Asia. In contrast, the Category-1 models relatively well reproduce the variations in vertical velocity and water vapor over East Asia and thus show a better agreement with observations in simulating the pattern of "wet south and dry north" in China in the past 20 years.

It is assessed that a single model's performance in simulating a particular variable has great impacts on the ensemble results. More realistic outputs can be obtained when the multi-model ensemble is carried out using a suite of well-performing models for a specific variable, rather than using all available models. This indicates that although a multi-model ensemble is in general better than a single model, the best ensemble mean cannot be achieved without looking into each member model's performance.

Key words: climate models, East Asian summer monsoon, model evaluation

1. Introduction

Simulation and prediction of the East Asian summer monsoon (EASM) circulation and precipitation is a key and challenging issue in climate research. In its most recent assessment (Randall et al., 2007), the Intergovernmental Panel on Climate Change (IPCC) indicates that the current atmosphere-ocean general circulation models (AOGCMs) have shown smaller errors in the simulation of the global monthly mean precipitation and its distribution, but for the EASM region, most models are unable to produce satisfactory simulations of monsoon-induced precipitation, due to the unique geographic location of the region and the complexity in the physical processes of the monsoon rainfall. The seasonal northward progression of major rain belts in the EASM region cannot be captured by these models and excessive monsoon rainfall is produced even with some false precipitation maximums over the central China in the model outputs. In their validation of the precipitation simulations in the Asian monsoon region, Annamalai et al. (2007) found that out of 18 AOGCMs under examination, only 6 models give reasonable monsoon precipitation on the climatological scale for the 20th century. The spatial correlation of the inter-model seasonal monsoon precipitation patterns exceeds 0.6, and better seasonal cycles are simulated by the 6 models. Out of the 6 models, 4 models produced better remote correlations between the East Asian monsoon and ENSO in the same period. In a word, although some major issues still loom in the monsoon-associated precipitation modeling,

*Supported jointly by the National Natural Science Foundation of China under Grant No. 40605020, 973 Program under No. 2006CB403604, and the State Key Project in the 11th Five-Year Plan under No. 2007BAC03A01.

†Corresponding author: sunying@cma.gov.cn.

improved simulations of the monsoon precipitation climatology have been obtained with a few new-generation climate models.

Past assessments on the EASM modeling focused mostly on the simulations of monsoon precipitation climatology (Zhao et al., 1995; Zhou and Li, 2002; Gao et al., 2004; Jiang et al., 2005; Liu et al., 2007; Feng and Fu, 2007; Tang et al., 2008; Jiang, 2008), and monsoon intraseasonal and interannual variations (Lambert and Boer, 2001; Lin et al., 2006). Few studies have performed assessments on the simulations of decadal and interdecadal variations of the summer monsoons, which have become increasingly important in current climate change studies. In the past half century, the East Asian monsoon witnessed evident decadal and interdecadal variations, which were characterized by a significant transformation of monsoon rainfall patterns and weakening monsoon circulations, with less intense southwesterlies over eastern China and impotent northward water vapor transportation (Ding and Sun, 2003; Yu et al., 2004; Yu and Zhou, 2007; Ding et al., 2008). The monsoon precipitation pattern shifted from "dry south and wet north" oppositely to "wet south and dry north" with the latter frequently showing up in the recent two decades (Zhai et al., 1999; Xu, 2001; Ding and Sun, 2003). Whether or not a model can capture the interdecadal variation of the EASM, i.e., a model's "reproductive capability" to represent the current interdecadal climate change, is not only an important indication of the model performance in the past climate simulations, but also a confidence validation of its future climate change projections.

In the IPCC AR4 (IPCC, 2007), the new-generation climate models (i.e., IPCC AR4 models, also referred to as CMIP3 models) currently used by the international research community have been extensively utilized to produce the future climate change projections. In fact, assessing the performance of these models and the uncertainties of their projections has become a hot topic in current climate studies (Meehl et al., 2007). The assessment includes evaluations on the models' physical parameterizations, overall performance and uncertainties in their climate projections. In general, to assess the overall performance of multiple models, an approach of multi-model ensembles (MME) is adopted. To some extent, the MME may offset the deviations of individual models, thus generating outputs closest to observations. The MME approach has been widely used in model assessments and climate projections (Cubasch et al., 2001). The basic assumption of MME is that outputs from each model can be considered as a close proximity to reality, and when the number of the members of MME grows big enough, the "noise" in the climate variability turns almost zero. Therefore, the MME outputs may be regarded as the best estimate of the climate change resulted from external forcing. It is emphasized that an individual model (MME member) should be capable to virtually capture the mean large-scale climate state.

In most current studies, although the capabilities to simulate the basic physical variables may vary from model to model, the weighing approach is generally not used, instead simpler arithmetic averages of multiple models are made (Cubasch et al., 2001). As assessments of model performance go in depth, some researchers have already noted that the models' performance in capturing the present climate patterns will have implications on their future climate change projections. In this sense, each MME member should have certain capability to simulate the climate mean state of some variables and their changes. As found by Knutti et al. (2006), the climate models with an intensive seasonal cycle of the surface temperature in general have greater climate sensitivity than those models with a weak surface temperature cycle. Though some more robust theoretical approaches are still being developed, an evaluation of model performance in simulating the present climate provides a way to restrict or constrain the future climate projections and associated uncertainties. Thus, it is imperative to conduct assessments of model performance for certain variables.

In this paper, based on the outputs of 19 AR4 models, the authors intend to make an assessment on these models' performance in simulating the interdecadal variation of the East Asian summer monsoon. A categorized multi-model ensemble approach is mainly used to study the differences between the EASM variations simulated by different categories of models, as well as possible causes behind the differences. The possible implications of these differences on future climate change projections will be discussed in a follow-up paper. The current paper is structured as follows: Section 2 introduces data and calculating methods used in this study; model performance in simulating the changes of the East Asian summer monsoon precipitation, large-scale circulation and water vapor are discussed in Sections 3, 4, and 5, respectively; the last section presents conclusions and discussion.

2. Data and methods

2.1 Data

Observational datasets used in this paper include:

(1) CMAP (Climate Prediction Center (CPC) Merged Analysis of Precipitation) monthly average gridded precipitation data (2.5°×2.5°) in 1979–1999 (Xie and Arkin, 1997);

(2) GPCP (Global Precipitation Climatology Project, version 2) monthly average gridded precipitation data (2.5°×2.5°) in 1979–1999 (Adler et al., 2003);

(3) China's monthly average precipitation observations at 740 stations in 1958–1999;

(4) ECMWF (European Centre for Medium-range Weather Forecasts) monthly average reanalysis data (2.5°×2.5°) in 1958–1999, http://www.ecmwf.int/research/era/Project/Plan/Project plan_TOC.html;

(5) NCEP/NCAR (National Centers for Environmental Prediction/National Center for Atmospheric Research) monthly average reanalysis data (2.5°×2.5°) in 1958–1999 (Kalnay et al., 1996).

Model data used are monthly mean data in 1958–1999 from 19 IPCC AR4 models (20C3M experiments, see Table 1). These models are characterized with

Table 1. A list of the 19 climate models examined in the present study

	Model name	Center, Country	Resolution (lon.×lat.)
1	CGCM3.1(T47)	Canadian Centre for Climate Modelling & Analysis, Canada	3.75°×∼3.75°
2	CGCM3.1(T63)	Canadian Centre for Climate Modelling & Analysis, Canada	2.8°×∼2.8°
3	CNRM-CM3	Météo-France/Centre National de Recherches Météorologiques, France	2.8°×∼2.8°
4	CSIRO	CSIRO Atmospheric Research, Australia	1.88°×∼1.88°
5	GFDL-CM2.0	US Dept. of Commerce/NOAA/Geophysical Fluid Dynamics Laboratory, United States	2.5°×2.0°
6	GFDL-CM2.1	US Dept. of Commerce/NOAA/Geophysical Fluid Dynamics Laboratory, United States	2.5°×2.0°
7	GISS-EH	NASA/Goddard Institute for Space Studies, United States	5°×4°
8	GISS-ER	NASA/Goddard Institute for Space Studies, United States	5°×4°
9	FGOALS-g1.0	LASG/Institute of Atmospheric Physics, China	2.8°×∼2.8°
10	ING-CM3.0	Institute for Numerical Mathematics, Russia	5°×4°
11	IPSL-CM4	Institut Pierre Simon Laplace, France	3.75°×2.5°
12	MIROC3.2(medres)	Center for Climate System Research (The University of Tokyo), National Institute for Environmental Studies, and Frontier Research Center for Global Change (JAMSTEC), Japan	2.8°×∼2.8°
13	MIROC3.2(hires)	Center for Climate System Research (The University of Tokyo), National Institute for Environmental Studies, and Frontier Research Center for Global Change (JAMSTEC), Japan	1.125°×∼1.12°
14	ECHAM5/MPI-OM	Max Planck Institute for Meteorology, Germany	1.88°×∼1.88°
15	MRI-CGCM2.3.2	Meteorological Research Institute, Japan	2.8°×∼2.8°
16	CCSM3	National Center for Atmospheric Research, United States	1.4°×∼1.4°
17	PCM	National Center for Atmospheric Research, United States	2.8°×∼2.8°
18	UKMO-HadCM3	Hadley Centre for Climate Prediction and Research/Met Office, United Kingdom	1.25°×1.25°
19	UKMO_HadGem1	Hadley Centre for Climate Prediction and Research/Met Office, United Kingdom	1.875°×1.25°

Table 2. Correlation coefficients between the models and observations

	Model name	Correlation coefficient of summer precipitation in 1979–1999 between the model and CMAP	Correlation coefficient of summer precipitation in 1979–1999 between the model and GPCP	Correlation coefficient of interdecadal variation (1979–1999 minus 1958–1978) of summer precipitation between the model and China station observations	Category
1	CGCM3.1(T47)	0.51	0.39	0.52	3
2	CGCM3.1(T63)	0.59	0.49	0.31	3
3	CNRM-CM3	0.84	0.85	−0.45	2
4	CSIRO	0.75	0.76	−0.10	2
5	GFDL-CM2.0	0.83	0.83	0.46	1
6	GFDL-CM2.1	0.82	0.80	−0.22	2
7	GISS-EH	0.33	0.36	−0.50	3
8	GISS-ER	0.40	0.44	−0.60	3
9	FGOALS-g1.0	0.24	0.28	−0.50	3
10	INM-CM3.0	0.76	0.70	0.15	2
11	IPSL-CM4	0.68	0.68	−0.42	3
12	MIROC3.2(hires)	0.80	0.82	0.38	1
13	MIROC3.2(medres)	0.81	0.76	0.23	1
14	ECHAM5	0.71	0.60	−0.43	3
15	MRI-CGCM2.3.2	0.60	0.61	0.43	3
16	CCSM3	0.57	0.48	−0.09	3
17	PCM	0.28	0.11	0.46	3
18	UKMO-HadCM3	0.89	0.84	−0.76	2
19	UKMO_HadGem1	0.80	0.73	−0.16	2

relatively high precision, more reasonable parameterization schemes, and updated numerical methods. Moreover, most models do not use flux adjustment schemes, and some models include interactive aerosol processes (Sun, 2005). More detailed information on the models is available at http : //www − pcmdi.llnl.gov/ipcc/aboutipcc.php.

Further information about precipitation parameterization schemes can be found in Sun et al. (2006) and Dai (2006). For the convenience, outputs from all models are linearly interpolated onto the same grid (2.5°×2.5°). In addition, the CSIRO model misses the vertical velocity field, thus the vertical velocity ensemble discussed in Section 3 does not include contributions from this model. Since the total column water vapor is not available from models UKMO_HadCM3 and UKMO_HadGem1, Section 5 excludes information from these two models.

2.2 Methods

In this paper, the multi-model ensemble mean for all the available models and that for different categories of models are used to study the differences in the simulated EASM interdecadal variation among the models. The basic idea is that the average rainfall in summer (from June to August) in the eastern China (22.5°–45°N, 110°–120°E) is treated as a reference variable for model category classification. Firstly, the models' performance in simulating multi-year average precipitation (from 1979–1999) is validated; then their performance in simulating precipitation variations (mean value of 1979–1999 minus that of 1958–1979) is examined; finally, the models are classified into 3 categories according to the assessments made in the previous two steps.

To begin with, a primary prerequisite to judge if a model is capable to reproduce the interdecadal variation of a variable should be its capability to more accurately capture the climatic multi-year average fields. Therefore, the average climate fields simulated by models are firstly validated. Taking into account the better quality of precipitation observations after 1979, due to the inclusion of satellite data, we use the CMAP and GPCP data averaged in summer

(June to August) in 1979–1999 to compare with outputs from the models. A study by Dai (2006) shows that CMAP and GPCP multi-year average precipitation fields mainly differ over ocean, and there is little difference in land. They are both used for assessing the models' performance.

In the paper, the eastern China (22.5°–45°N, 110°–120°E) is chosen as the target area. The method of correlation coefficient test is used for making comparisons between the observations and the models' outputs. The models' performance in simulating spatial precipitation distributions in the above specific area is firstly assessed. Then the models are sorted accordingly into different categories.

In Table 2, the third and fourth columns show the correlation coefficients between the simulated summer precipitation in the eastern China averaged for 1979–1999 and the CMAP/GPCP observations. It is clear that a large number of models are capable in simulating the multi-year mean summer precipitation in the target region, with the correlation coefficient as high as 0.89 (between the UKMO-HadCM3 outputs and CMAP observations). If that the correlation coefficient between the simulations and CMAP/GPCP observations exceeds 0.75 is used as a criterion for classification, 9 models meet this criterion, and they are referred to as the models that "better" simulate the climatic average precipitation field. These models are GFDL-CM2.0, MIROC3.2 (hires), MIROC3.2 (medres), CNRM-CM3, CSIRO, GFDL-CM2.1, INM-CM2.0, UKMO-HadCM3, and UKMO_HadGem1. The rest 10 models are referred to as the models that give "inadequate" simulations. Among them, some correlation coefficients between the simulations and observations are only 0.2–0.3. Apparently, these models' performance is inadequate in capturing the summer precipitation in the eastern China.

The simulated interdecadal precipitation variation is verified by comparing the difference between the two average fields, i.e., the 1979–1999 average minus the 1958–1978 mean. The reason for this is that around the end of 1970s, the atmospheric circulation and ocean in the Northern Hemisphere experienced a significant abrupt change, which was reflected in many variable fields, including the circulation over East Asia. Thus, the difference between these two temporally averaged fields may adequately represent the typical EASM interdecadal change in the past. The precipitation observations selected for comparison are the station data in China collected over a longer time, which are interpolated to a 2.5°×2.5° grid before their correlations with model outputs are calculated. Column 5 in Table 2 shows the correlations between the changes (1979–1999 minus 1958–1978) of simulated (19 models) and observed precipitation in the eastern China. It is seen that a good number of models have simulated correctly the multi-year average monsoon precipitation over the eastern China, but their performance is rather poor in capturing the interdecadal precipitation change. Simulations from a number of models are even negatively correlated with observations, i.e., the distribution of the simulated precipitation variation by these models is just opposite to the actual observations. Out of the 9 models that well simulate the climatically averaged precipitation field, only 3 models (GFDL-CM2.0, MIROC3.2 (hires), and MIROC3.2 (medres)) show a positive correlation above 0.2 with the observed precipitation, while the rest are all in negative correlations. The correlation of the climatic mean precipitation between the simulations by UKMO-HadCM3 and the observation is as high as 0.89, but it is as low as −0.76 for the precipitation variation field, almost opposite to the observed truth (figure omitted). This indicates that the models that are able to well simulate the climatic mean fields may not give a good performance in reproducing an interdecadal variation.

According to the above analyses, all models are divided into 3 categories based on the following two criterions: (1) the correlation between the simulated and observed (both CMAP and GPCP) climatic average precipitation is above 0.75 and (2) the correlation between the simulated and observed interdecadal variation of precipitation is above 0.2. Category-1 models refer to those meeting both criterions (1) and (2), Category-2 models are those meeting (1) but failing to meet (2), and Category-3 models failing to meet both (1) and (2). By this criterion, 3 models

(GFDL-CM2.0, MIROC3.2 (hires), and MIROC3.2 (medres)) fall into Category 1, 6 models (CNRM-CM3, CSIRO, GFDL-CM2.1, INM-CM2.0, UKMO-HadCM3, and UKMO-HadGem1) into Category 2, and 10 models (CGCM2.1 (T47), CGCM3.1 (T63), GISS-EH, GISS-ER, FGOALS-G1.0, IPSL-CM4, ECHAM5, MRI-CGCM2.3.2, CCSM3, and PCM) into Category 3. See Table 2 for more details.

As such, the subsequent discussion will be made on the ensemble mean of the 3 categories of models and for all 19 models, respectively. For brevity, they are hereafter referred to as the ensembles of Categories 1, 2, and 3, and the ensemble of 19 models.

3. Assessment on the performance of precipitation simulations

Figure 1 shows the 1979–1999 June-July-August (JJA) average precipitation from GPCP and CMAP, and from the ensembles of Categories 1–3 and the ensemble of 19 models. It can be seen that for the multi-year average, the difference between the different categories is not prominent in terms of the climatic mean distribution. All the ensembles can produce a distribution of precipitation featured with a decline from the eastern to western China, but they all produce a false heavy precipitation center over the central China.

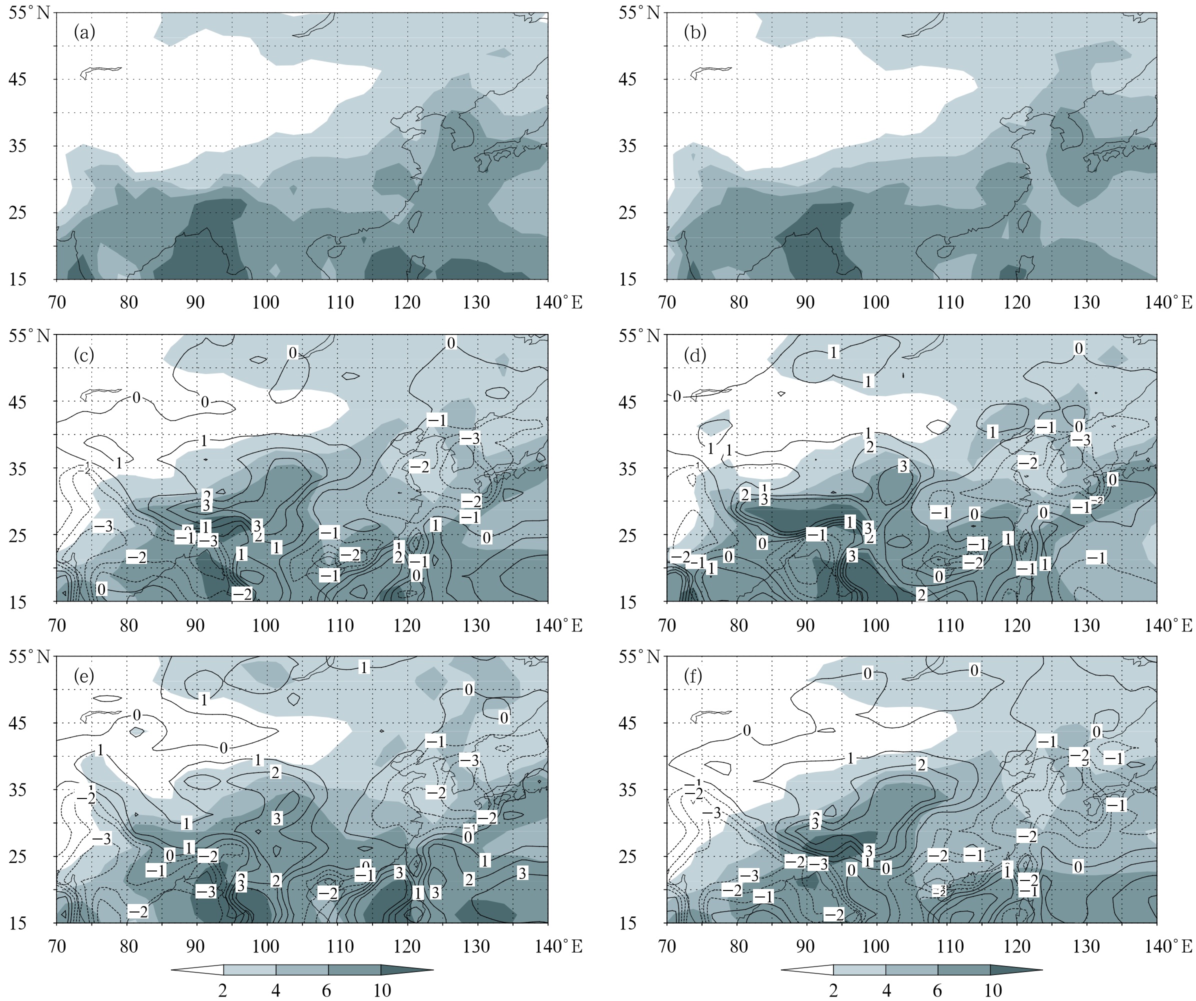

Fig.1. 1979–1999 June-July-August (JJA) mean precipitation based on (a) GPCP, (b) CMAP, (c) 19-model ensemble mean, (d) Category-1 models ensemble mean, (e) Category-2 models ensemble mean, and (f) Category-3 models ensemble mean. Shaded areas in (c), (d), (e), and (f) denote the model ensemble mean and contours are the difference between the model ensemble mean and GPCP.

However, in the key area −the eastern China, the difference in the regional averaged precipitation between the observations and those simulated by both Category-1 and -2 models is smaller than the difference between the observations and those simulated by Category-3 models, indicating that the former is "superior" to the latter in capturing the climatic average precipitation for the eastern China.

Figure 2 shows the interdecadal precipitation variation from the observations and simulations. It can be clearly seen that the precipitation variation simulated by all the model ensembles is smaller than the observations in terms of magnitude. The 19-model ensemble (Fig.2b) and Category-3 models ensemble (Fig.2e) show a distribution contrary to the real observations, e.g., reduced precipitation over South China and the Yangtze River Basin while increased precipitation across North China. The simulations by Category-2 models (Fig.2d) also differ largely from the real observations, all showing increased precipitation in most parts of China. Only the Category-1 models (Fig.2c) show a better consistency with the observations, i.e., reduced precipitation over North China and increased precipitation across the Yangtze River Basin

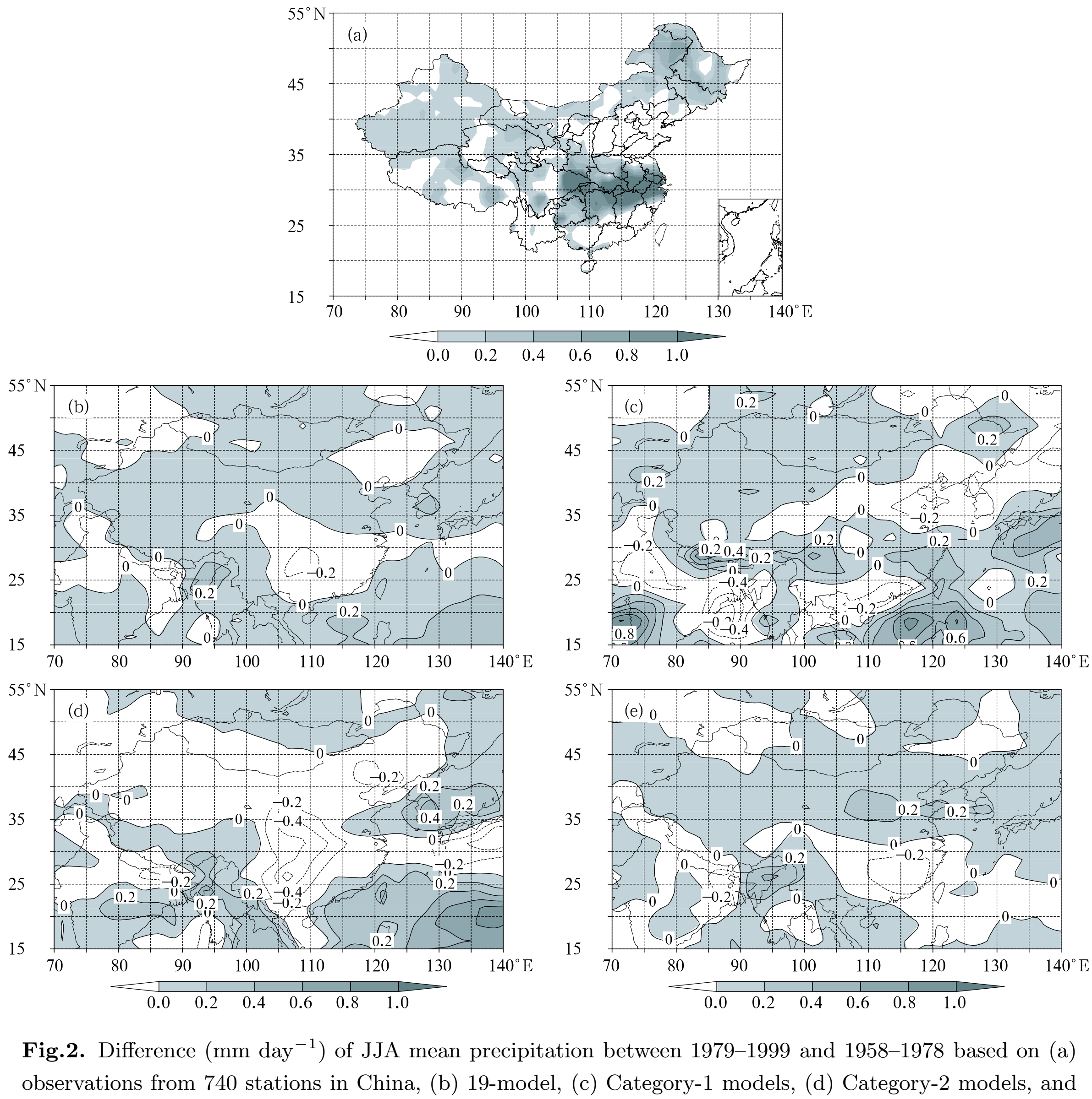

Fig.2. Difference (mm day^{-1}) of JJA mean precipitation between 1979–1999 and 1958–1978 based on (a) observations from 740 stations in China, (b) 19-model, (c) Category-1 models, (d) Category-2 models, and (e) Category-3 models ensemble mean.

with slightly decreased precipitation over South China. In addition, it is indicated that both Northeast and Northwest China are slightly wetter. Thus, Category-1 models have done a good job in simulating the interdecadal variation of the East Asian monsoon precipitation. The simulated distribution is very similar to that of the actual observations with only smaller magnitude. In this sense, we can say that the 3 models in Category 1 have stronger capability to simulate the East Asian monsoon precipitation.

In order to look out for better observations for the assessment, Fig.3 shows the distribution of the interdecadal precipitation variation from the reanalysis data of both ECMWF and NCEP. In comparison, ECMWF reanalysis precipitation is basically identical with the station observations (Fig.2a), with the only exception that the precipitation variation over the western China is greater than observations. The precipitation variation over the eastern China exhibits a "wet south and dry north" pattern. Comparatively, no matter in terms of distribution or the magnitude of the precipitation variation, NCEP reanalysis precipitation differs widely from the observations. This shows that for the East Asian region, the ECMWF reanalysis data are of better quality. Thus, large-scale circulation observations derived from the ECMWF reanalysis are to be used in the assessment of simulations of large-scale circulation changes in the next section.

4. Assessment on the performance of large-scale circulation simulations

Figure 4 shows a comparison between ECMWF reanalysis and the simulated interdecadal wind variation at 850 hPa based on the ensemble means for all models and Categories 1–3. From ECMWF data, the changes in Asia in the last 50 years are mainly characterized by prevalence of abnormal anti-cyclone outflows from the Tibetan Plateau and Mongolia, hence the abnormal northeastern air flows appear from East Asia to Bangladesh and Indian Peninsula. This indicates that the East Asian monsoon and Indian monsoon circulations become weak, which is consistent with previous findings (Ding and Sun, 2003; Yu et al., 2004). The model simulations in general are weaker than the observations, especially in that the changes of wind intensity are smaller than the reanalysis. As to the distribution of the interdecadal variation, all the model ensembles fail to produce the weakening of the large-scale Asian monsoon circulation. For East Asia, although model simulations by the Category -1 and -2 models show weakening southwesterly monsoonal flows along the south of the Yangtz River and South China, the flow patterns associated with the weakened southwesterlies, however, differ widely from observations.

Figure 5 shows the interdecadal variation of

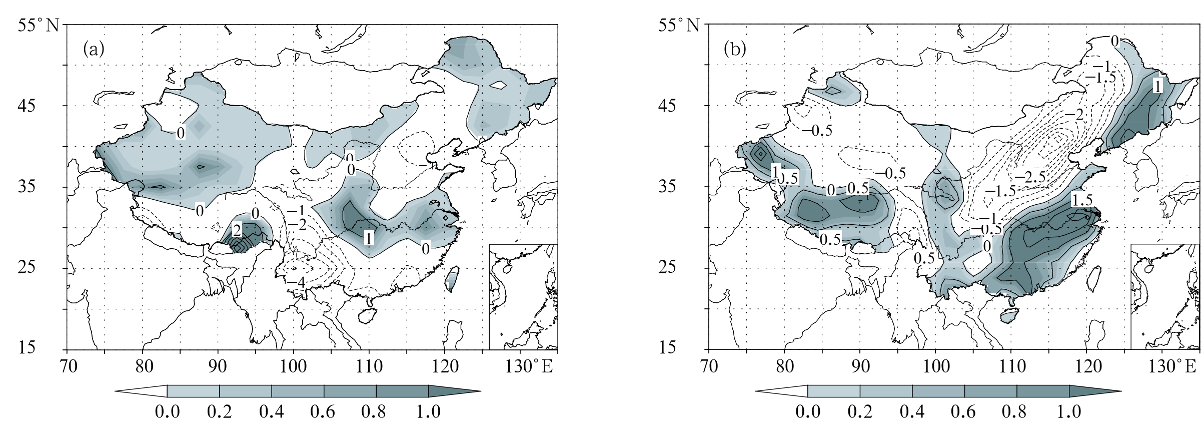

Fig.3. Difference (mm day^{-1}) of JJA mean precipitation between 1979–1999 and 1958–1978 based on (a) ECMWF and (b) NCEP/NCAR.

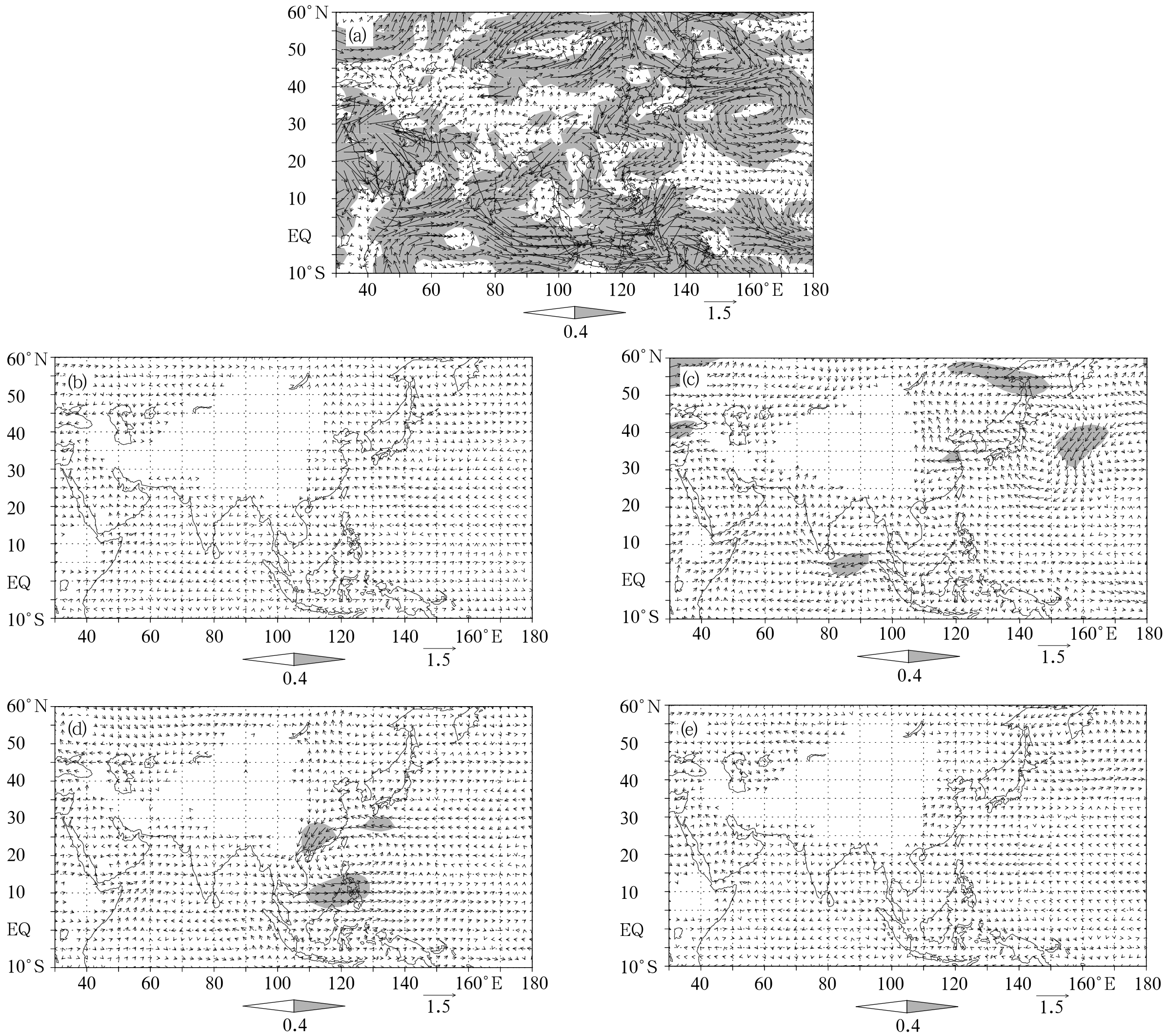

Fig.4. Difference (m s^{-1}) of JJA mean 850-hPa wind between 1979–1999 and 1958–1978 based on (a) observations from ECMWF, (b) 19-model, (c) Category-1 models, (d) Category-2 models, and (e) Category-3 models ensemble mean.

500-hPa geopotential height from the ECMWF reanalysis and from the ensemble means of 19 models and Categories 1–3 models. A negative change centered from Mongolia to Japan can be seen, indicating that the geopotential height in these areas is declining. However, all the model ensembles have failed to reproduce this pattern; instead they present a pattern with increased geopotential height across Asia, tropical Indian Ocean and the Pacific. This reflects that none of the models is capable to reproduce such an interdecadal circulation change.

Figure 6 displays the latitude-height cross-section of the interdecadal variation of vertical velocity over the eastern China (110°–120°E). It can be seen that the vertical motion has changed prominently. Figure 6a illustrates that changes in vertical velocity around 22°–27°N (South China) and 32°–40°N (North China) are positive, and therefore vertical upward motion is weakening in these areas, which is unfavorable for precipitation. By contrast, the changes in vertical motion for the latitudes around 27°–32°N (Yangtze River Basin) are negative, showing that the vertical upward motion is strengthening and favorable for precipitation. For model ensembles, all the model simulated

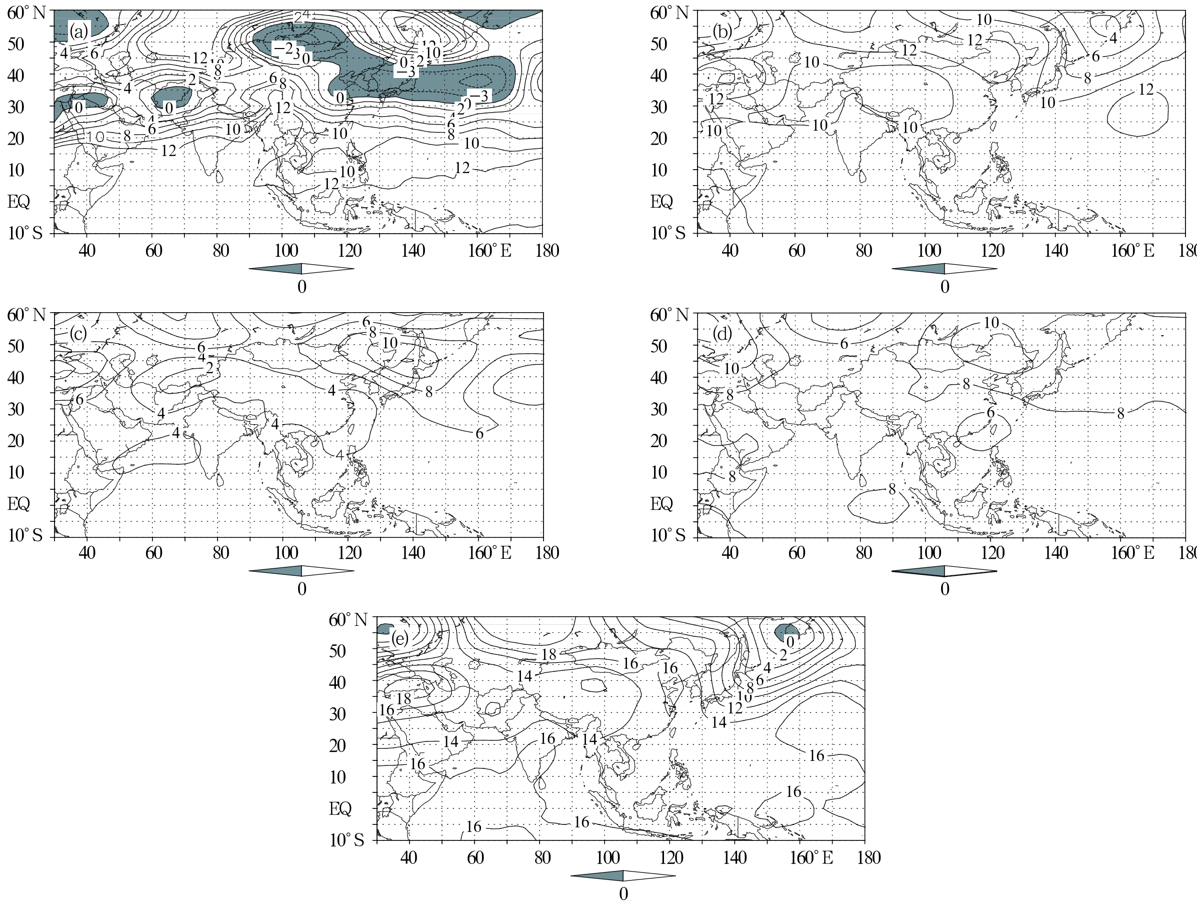

Fig.5. As in Fig.4, but for 500-hPa geopotential height (gpm).

changes are smaller than the observations in terms of magnitude. The changes in vertical motion across the Yangtze River produced by Category-2 and -3 models are positive, which is opposite to the observed changes. Only Category-1 models reproduce the negative changes over the Yangtze River areas and the positive changes over North China, which are closer to the observations, suggesting that Category-1 models well captured the characteristics of the vertical motion.

The interdecadal change of temperature shown in Fig.7 indicates possible reasons for the above-described difference between the different categories of models. ECMWF data show that in the past 20 years, an extensive large-scale atmospheric cooling occurred in the middle troposphere over Asia. It made the air column shrink, and the air pressure of the lower troposphere decreased and that of the higher troposphere increased, leading to an anomalous anti-cyclonic circulation at 850 hPa (Fig.4a), a negative height anomaly center at 500 hPa (Fig.5a), and an anomalous cyclonic circulation at higher levels (figure omitted). At the same time, to the east of the anti-cyclonic circulation, abnormal vertical upward and downward motion existed over the Yangtze River Basin and North China (Fig.6a). Model simulations show that Category-1 models display a relatively slight warming over East Asia but still fail to capture the observed large-scale cooling. The other categories of models could not reproduce the cooling in the middle troposphere, and on the contrary they reproduce an obvious warming over the same area. Therefore, from the perspective of dynamic configuration, most models lack the physical mechanism that induces the weakening of the monsoon circulation and thus are unable to reproduce the decline of the East Asian

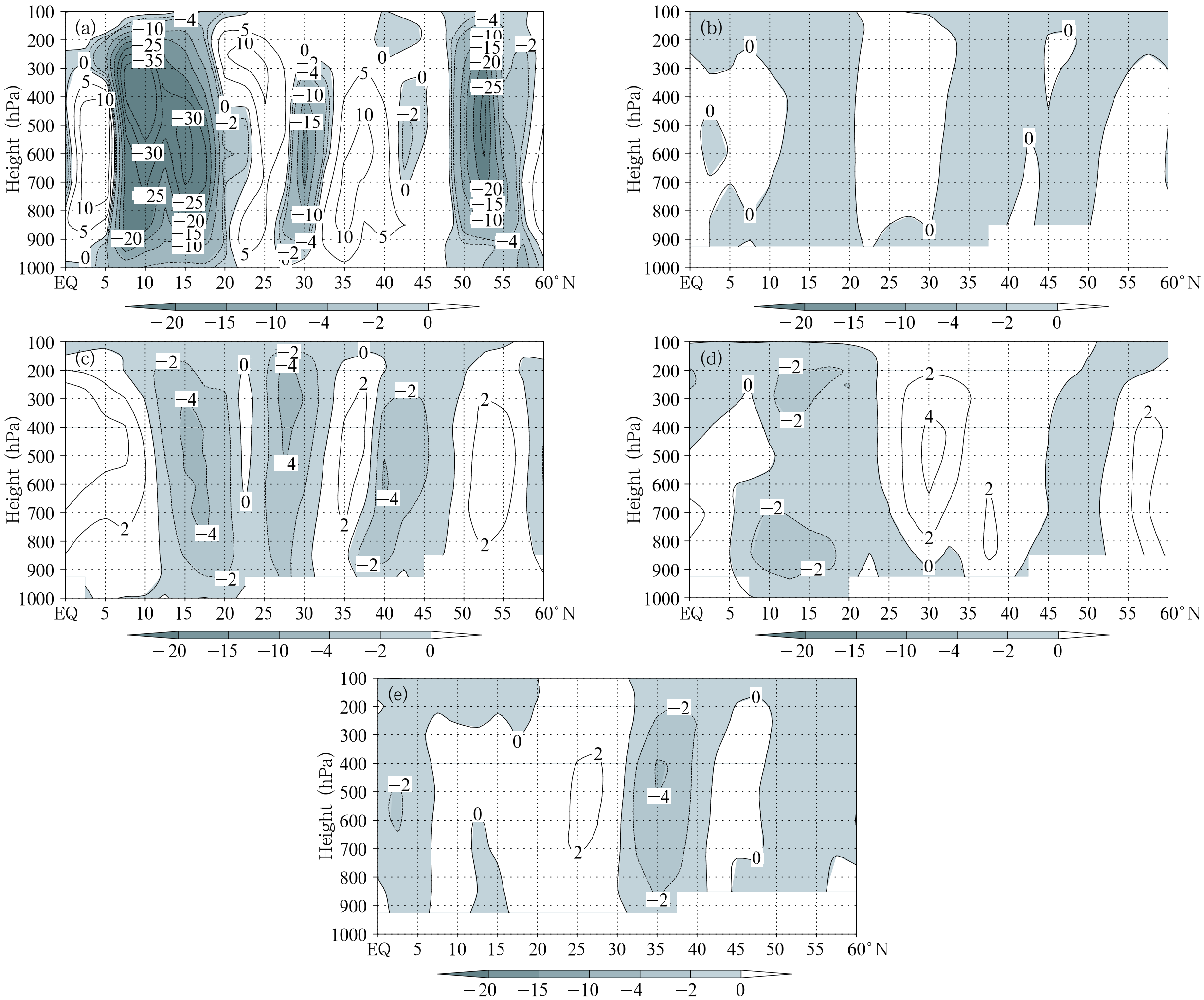

Fig.6. As in Fig.4, but for latitude-height cross-section of vertical velocity averaged between 110° and 120°E (10^{-3} Pa s^{-1}).

monsoon circulation in the past 50 years. It should be noted that some models such as GFDL-CM2.0, GFDL-CM2.0, and MIROC3.2 (medres) have given a slight cooling over East Asia (Fig.10), although the range and intensity are much weaker than the observed. It may explain why Category-1 models gave a better performance in capturing the vertical motion, while other models did not.

Some recent findings suggest that the cooling in the middle troposphere is one of the important reasons for the climate change in East Asia. In their analyses, Ding et al. (2008) pointed out that the increased snow over the Tibetan Plateau and increased sea surface temperature in the tropical Pacific may be one of the mechanisms causing the change of the land-sea thermal contrast, which in return contributes to the cooling over the middle troposphere and the weakening monsoon circulation in East Asia. However, further studies are needed to better understand such issues as whether or not the physical mechanism that triggers the cooling of the middle troposphere is related to the changes of other large-scale circulations and oceans in the Northern Hemisphere. For the models, apart from Category-1 models, the rest models are mostly incapable to simulate such a large-scale temperature change. Accordingly, these models can hardly reproduce the interdecadal change in the East Asian summer monsoon circulation.

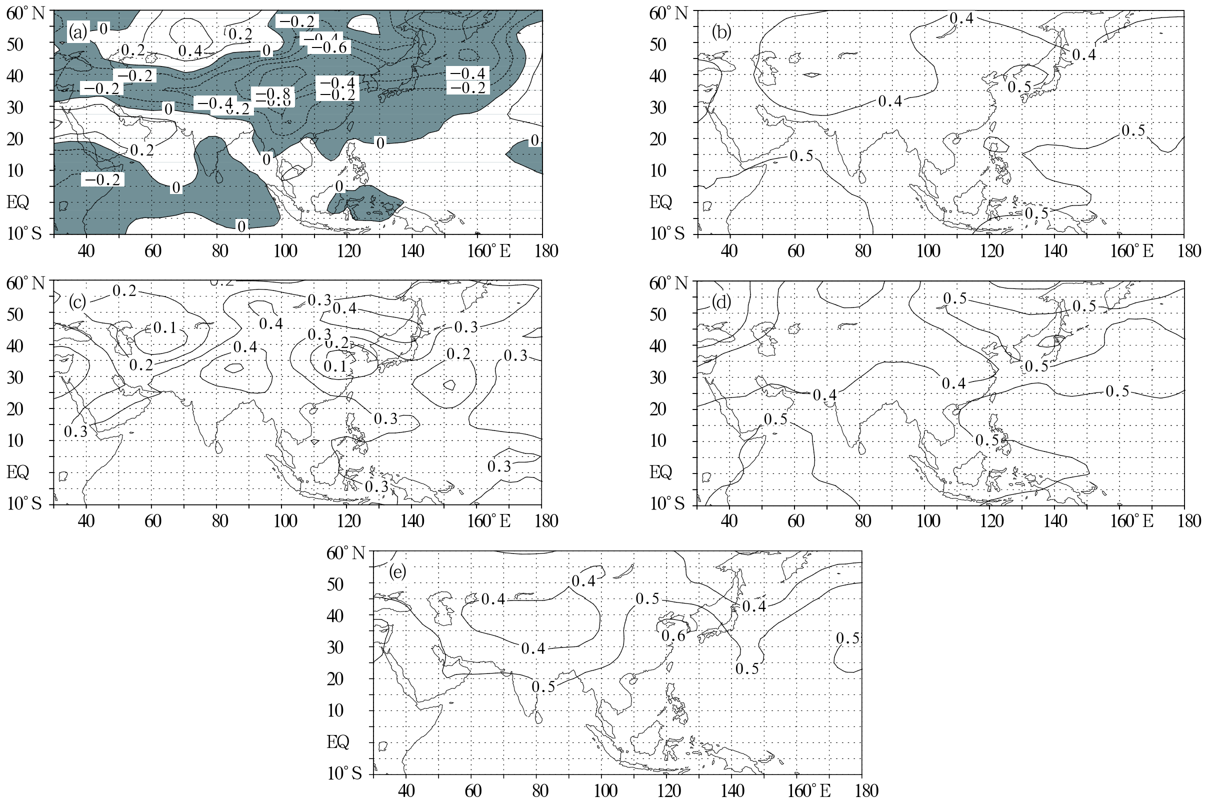

Fig.7. As in Fig.4, but for 300–500-hPa mean temperature (K).

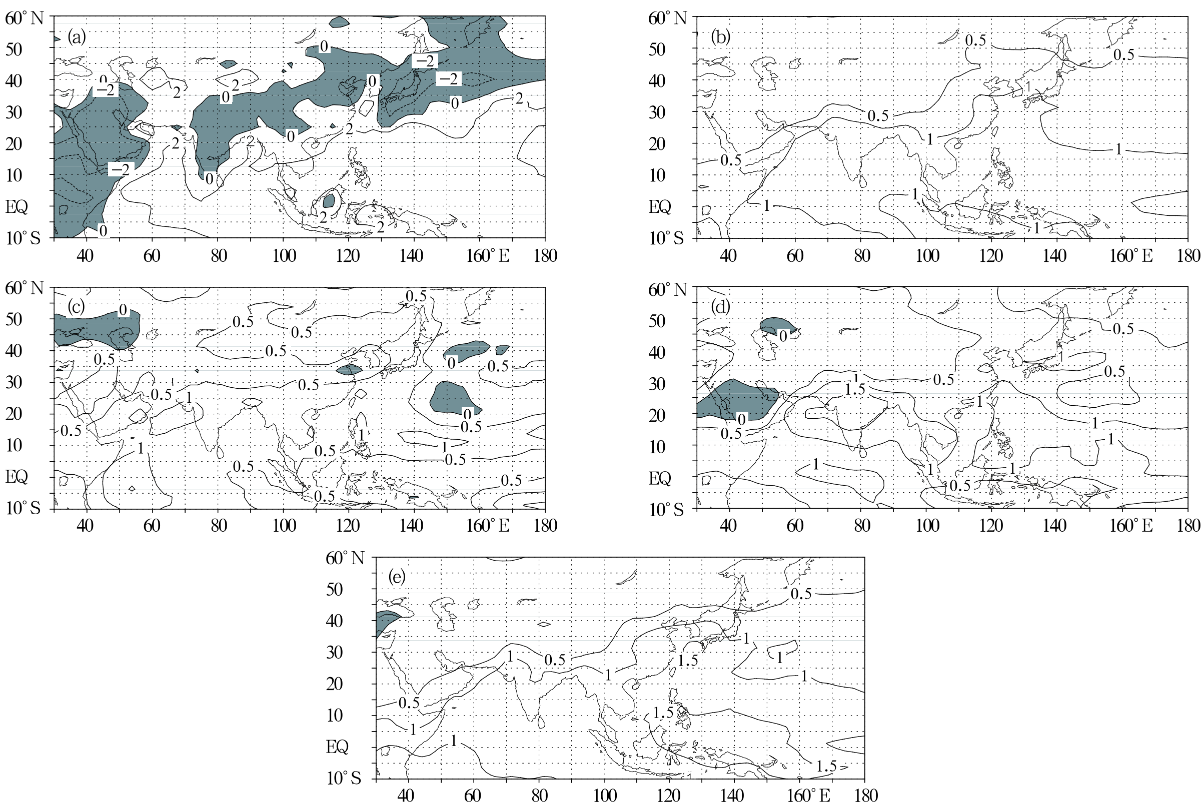

Fig.8. As in Fig.4, but for total column water vapor content (kg m^{-2}).

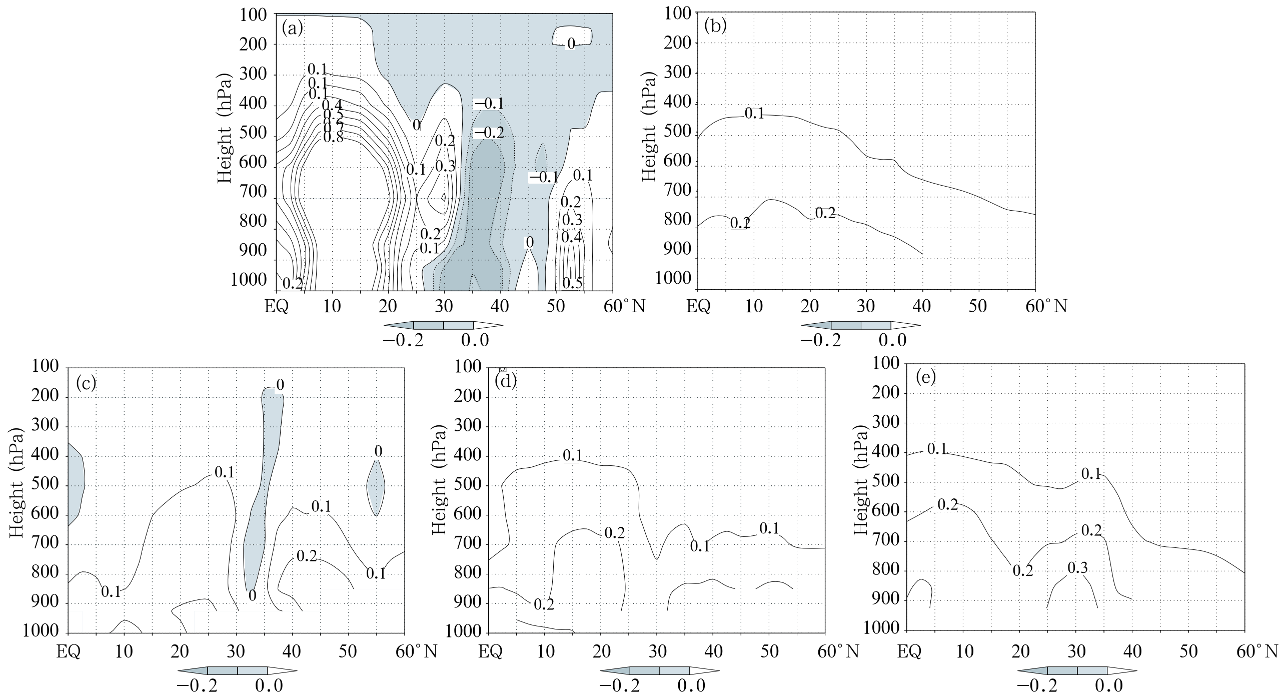

Fig.9. As in Fig.4, but for latitude-height cross-section of specific humidity averaged between 110° and 120°E (kg kg^{-1}).

5. Assessment on the performance of water vapor simulations

Water vapor changes have played a very important role in the process of precipitation formation. Figure 8 displays the interdecadal change in the total column water vapor. An evident negative center is found emerging from North China to the Indian Peninsula over the past two decades, suggesting the substantially reduced water vapor in these areas, especially over North China and Japan. This indicates that the reduced precipitation in North China is due to both the circulation change and water vapor reduction. However, in the model simulations, only Category-1 models reproduce slightly decreased water vapor over Asia and North China, while the other categories of models show increased water vapor over Asia.

The latitude-altitude cross-section (Fig.9) of the interdecadal variation of specific humidity in East Asia further illustrates the above-mentioned difference. As indicated by the zonal mean of specific humidity in 110°–120°E, the negative variation of specific humidity exists in the middle and upper troposphere in the observed fields between 30° and 40°N, extending from 800 hPa in the lower troposphere to the upper layers, showing that the water vapor over North China, is decreasing. As to the simulations, the decreasing water vapor is reproduced by Category-1 models with a negative change of water vapor seen in the troposphere between 800 and 600 hPa. However, both the range and magnitude of this negative change are much smaller than observations. The other categories of models are not able to reproduce the decreasing water vapor over the middle latitudes of East Asia, while the water vapor generally decreases as the latitudes and altitudes increase. More detailed results of individual models are discussed in the next section.

6. Conclusions and discussion

Different observational data are used in this paper to evaluate the performance of 19 CMIP3 models in simulating the East Asian summer monsoon interdecadal change. Among all the models under examination, 9 models (GFDL-CM2.0, MIROC3.2 (hires), MIROC3.2 (medres), CNRM-CM3, CSIRO,

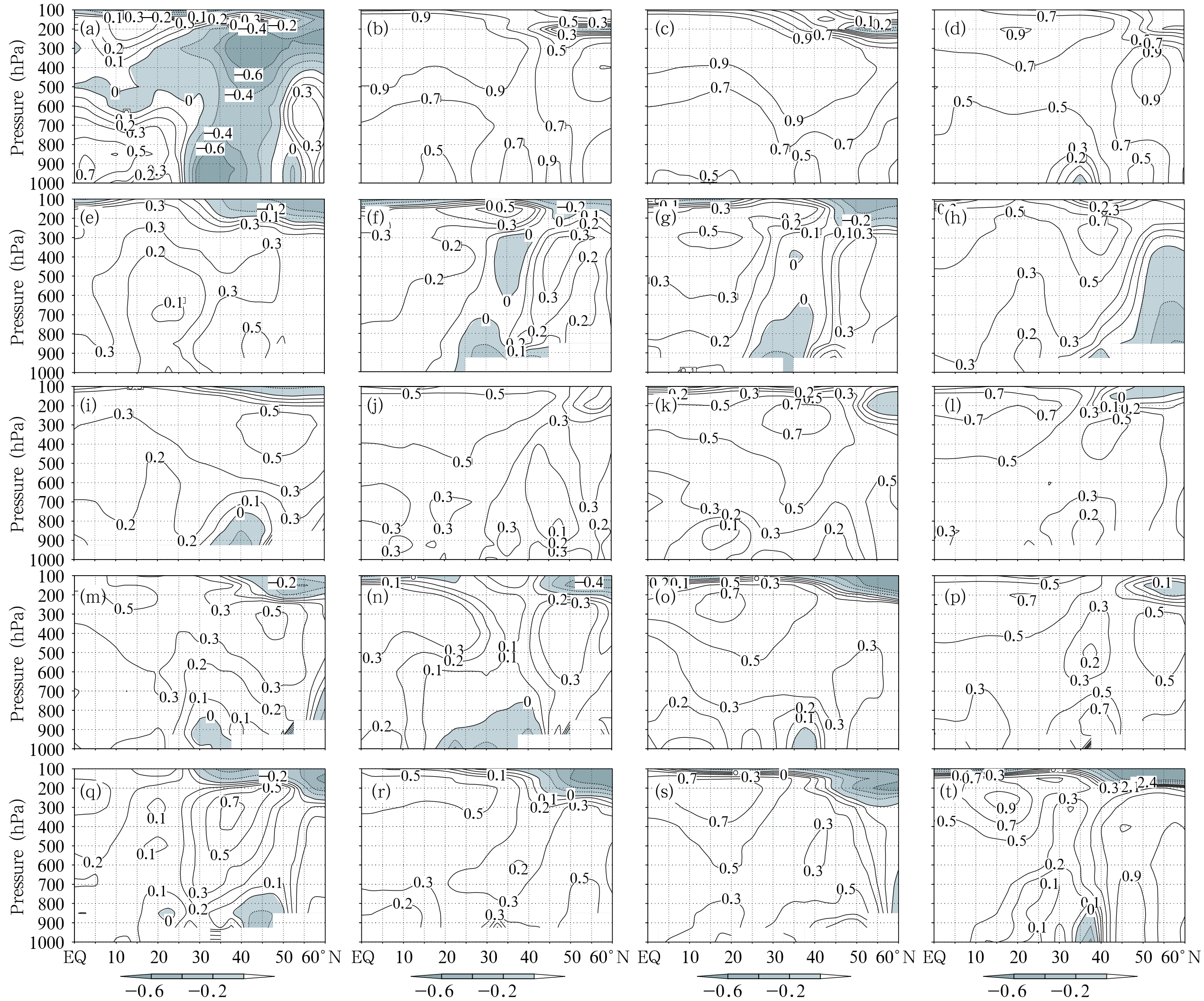

Fig.10. Latitude-height cross-sections of temperature difference (K; 1979–1999 minus 1958–1978 averages) averaged between 110° and 120°E based on ECMWF data and 19 individual model simulations.

GFDL-CM2.1, INM-CM2.0, UKMO-HadCM3, and UKMO-HadGem1) can well reproduce the multi-year average precipitation over eastern China, only 3 models (GFDL-CM2.0, MIROC3.2 (hires), and MIROC3.2 (medres)) show the capability in reproducing the interdecadal change of the EASM precipitation and circulation, while most models are not able to simulate the interdecadal change of the EASM. The analyses of large-scale circulations and water vapor fields suggest that the main reason behind the failure to capture the EASM precipitation change is the lack of related physical mechanism in the models. The insufficient dynamic and thermal forcing caused the models to deviate from the truth.

It is found that 3 well-performing models in Category 1 are able to well reproduce the variation features of the vertical motion and water vapor over East Asia. They capture the dynamic and thermal mechanisms of the precipitation change, thus reproduce the climate change featuring a "wet south and dry north" pattern over the eastern China in the past 50 years. However, the changes of temperature, vertical velocity, and water vapor in these models are apparently smaller than the observations. The outputs from individual models as given in Figs.10 and 11 clearly confirm this point. As mentioned earlier in this paper, the

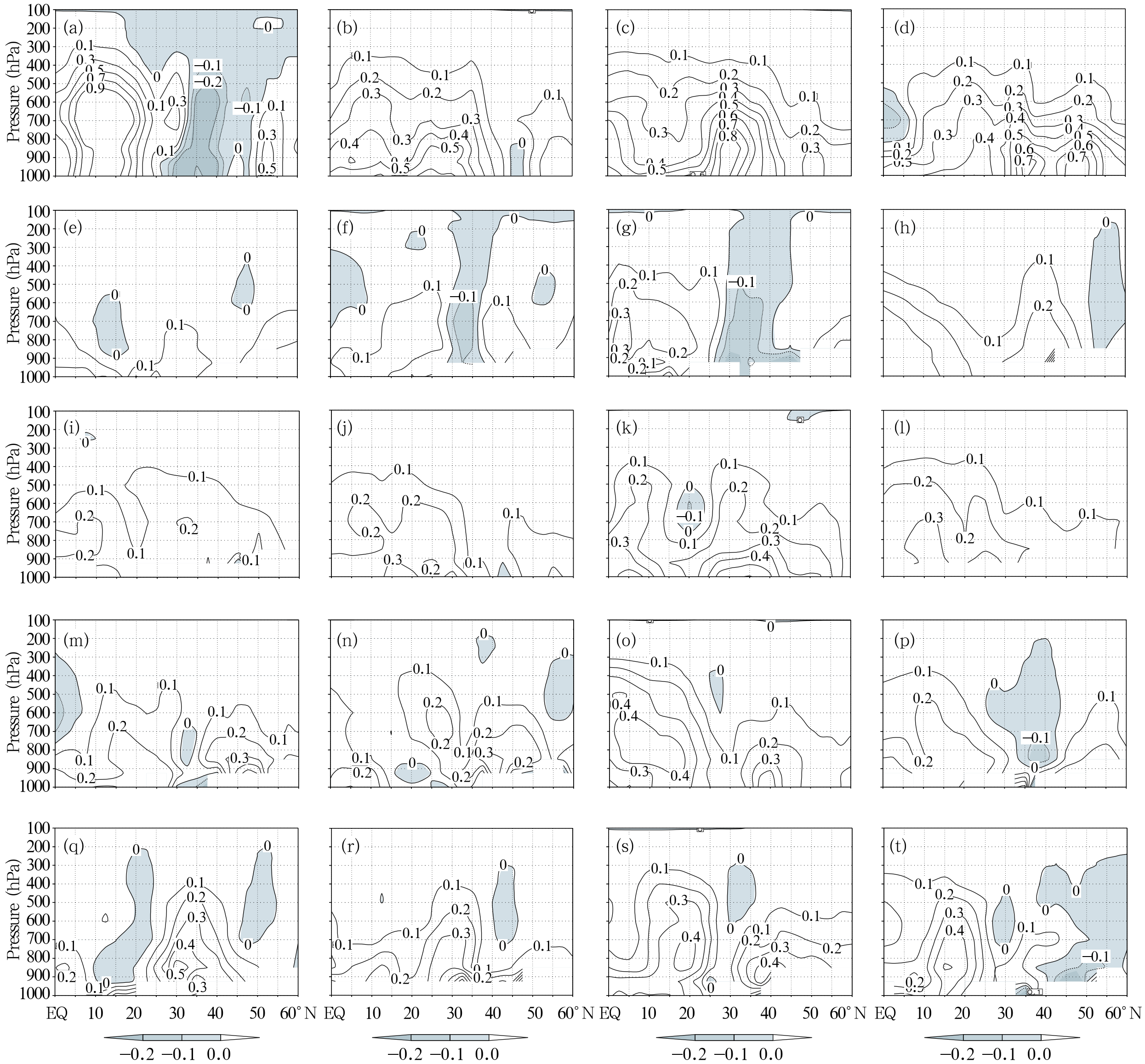

Fig.11. As in Fig.10, but for specific humidity (kg kg^{-1}).

latitude-altitude profiles of the interdecadal variation of temperature and specific humidity over East Asia (110°–122.5°E) from the ECMWF data show that a deep cooling zone exists over North China corresponding to an entire dry troposphere above it. However, except GFDL-CM2.0, GFDL-CM2.1, MIROC3.2 (hires), and MIROC3.2 (medres) models, the other models basically show warming and increasing water vapor over this region. Among these four models, the GFDL-CM2.0 and GFDL-CM2.1 have well reproduced the vertical distribution pattern, which is characterized with a drying water vapor field; the simulated cooling is smaller in range and weaker in intensity, which is more or less close to the observations as at least a slight cooling is present in the middle troposphere. The temperature change over East Asia simulated by both MIROC3.2 (hires) and MIROC3.2 (medres) is far weaker than observations. The water vapor change is seen only in MIROC3.2 (hires), but with limited cooling and drying only appearing near the surface. In comparison, GFDL-CM2.0 and GFDL-CM2.1 performed better in simulating the vertical distribution of temperature and water vapor over East Asia, but the zonal extent of the water vapor variation in

GFDL-CM2.1 is larger than the observations, thus this model failed to reproduce the zonal distribution of the precipitation variation while GFDL-CM2.0 succeeded. In this regard, the best performing model to capture the EASM variation is GFDL-CM2.0 on the whole, which also explains why its precipitation variation acquires the highest correlation coefficient with observations.

This paper also finds that the model simulations of the EASM precipitation and circulation differ largely from one model to another, due to a variety of reasons, such as the performance of the model itself, choice of parameterization schemes, different physical mechanisms in response to forcing, and so on. Furthermore, the mechanism causing the middle tropospheric cooling over East Asia may be attributed to the external forcing or, alternatively, the internal adjustment within the climate system. If the key mechanism is not inherent in the model, it will be difficult to capture the circulation change. Precipitation change is even more complicated, as both dynamic and thermodynamic factors may be involved. Even under the same forcing, the precipitation variation is much more complicated than the temperature change, since it is affected by more factors, such as the complex topography.

The assessments presented in this paper intend to make it clear that model performance in simulating variations of certain variables have great implications on the outcomes of the ensembles when different models are selected. When the ensembles are based on well-performing models, the outcome is closer to the observations. So far as the specific variables are concerned, the ensemble of the well-performing models proves to be superior to the ensemble of all available models. This shows that although a multi-model ensemble generally is better than any individual model, the assessment on the performance of the selected models that participate in the ensemble is equally important. If the MME members have good performance in simulating various variables and capturing the physical mechanisms behind the changes of these variables, such a multi-model ensemble will have a higher credibility. On the contrary, poor-performing models tend to give results that differ greatly from the observations. It is suggested that only when the physical mechanisms of climate change are better understood and when our confidence in the current and future climate predictions are improved, will both the current simulations and future climate projections reach a higher level.

Acknowledgements. We thank two anonymous reviewers for helpful comments, and Mr. Xiao Hongxian and his colleague for their help with the language. We acknowledge the international modeling groups for providing their data, 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. Thanks also go to ECMWF for providing the ERA-40 data used in this study.

REFERENCES

Adler, R. F., and Coauthors, 2003: The version-2 global precipitation climatology project (GPCP) monthly precipitation analysis (1979-present). *J. Hydrometeor.*, **4**, 1147–1167.

Annamalai, H., K. Hamilton, and K. R. Sperber, 2007: South Asian summer monsoon and its relationship with ENSO in the IPCC AR4 simulations. *J. Climate*, **20**, 1071–1083.

Cubasch, U., G. A. Meehl, G. J. Boer, R. J. Stouffer, M. Dix, A. Noda, C. A. Senior, S. Raper, K. S. Yap, and Coauthors, 2001: Projections of future climate change. *Climate Change 2001: The Scientific Basis*. Houghton J. T. et al., Eds., Cambridge Unversity Press, 525–582.

Dai, A., 2006: Precipitation characteristics in eighteen coupled climate models. *J. Climate*, **19**, 4605–4630.

Ding Yihui and Sun Ying, 2003: Long-term climate variability in China. WMO/TD No. 1172, 18–33.

Ding Yihui, Sun Ying, Wang Zunya, and Zhu Yuxiang, 2008: Interdecadal variation of the summer precipitation in China and its association with decreasing Asian summer monsoon. Part II: Possible causes of the inter decadal variability of the summer precipitation in China. *Int. J. Climatology*, accepted.

Feng Jinming and Fu Cong-Bin, 2007: Inter-comparison of long-term simulations of temperature and precipitation over China by different regional climate models. *Chinese J. Atmos. Sci.*, **31**(5), 805–814.

Gao Xuejie, Lin Wantao, Fred Kucharsky, and Zhao Zongci, 2004: A simulation of regional climate in China by using CCM3 and observed SST. *Chinese J. Atmos. Sci.*, **28**(1), 78–90.

IPCC, 2007: Summary for Poicymakers. Climate Change 2007: The Physical Science Basis. *Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change.* Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K. B. Averyt, M.Tignor, and H.L. Miller, Eds. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Jiang Dabang, 2008: Projected potential vegetation change in China under the SRES A2 and B2 scenarios. *Adv. Atmos. Sci.*, **25**(1), 126–138.

Jiang Dabang, Wang Huijun, and Lang Xianmei, 2005: Evaluation of East Asian climatology as simulated by seven coupled models. *Adv. Atmos. Sci.*, **22**(4), 479–495.

Kalnay, E., M. Kanamitsu, R. Kistler, et al., 1996: The NCEP/NCAR 40-year reanalysis project. *Bull. Amer. Meteor. Soc.*, **77**, 437–470.

Knutti, R., G. A. Meehl, M. R. Allen, and D. A. Stainforth, 2006: Constraining climate sensitivity from the seasonal cycle in surface temperature. *J. Climate*, **19**, 4224–4233.

Lambert, S. J., and G. J. Boer, 2001: CMIP1 evaluation and intercomparison of coupled climate models. *Clim. Dyn.*, **17**, 83–106.

Lin J. L., et al., 2006: Tropical intraseasonal variability in 14 IPCC AR4 climate models. Part I: Convective signals. *J. Clim.*, **19**, 2665–2690.

Liu Yanxiang, Wu Tongwen, Guo Yufu, and Yan Jinghui, 2007: Prediction research of climate change trends over North China in the future 30 years. *Acta Meteor. Sinica*, **65**(1), 45–51. (in Chinese)

Meehl, G. A., and Coauthors, 2007: The WCRP CMIP3 multimodel dataset: A new era in climate change research. *Bull. Amer. Meteor. Soc.*, 1383–1394.

Randall, D. A., R. A. Wood, S. Bony, R. Colman, T. Fichefet, J. Fyfe, V. Kattsov, A. Pitman, J. Shukla, J. Srinivasan, R. J. Stouffer, A. Sumi, and K. E. Taylor, 2007: Cilmate models and their evaluation. *Climate Change 2007: The Physical Science Basis.* Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K. B. Averyt, M. Tignor, and H.L. Miller, Eds. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Sun Ying, 2005: Introduction to analyses of climate model simulations for the IPCC AR4. *Advances in Climate Change*, **1**(4), 161–163. (in Chinese)

Sun Ying, S. Solomon, A. Dai, and R. W. Portmann, 2006: How often does it rain? *J. Climate*, **19**, 916–934.

Tang Jianping, Zhao Ming, and Su Mingkai, 2008: Numerical simulation of regional climate change under IPCC A2 scenario in China. *Acta Meteor. Sinica*, **66**(1), 13–25. (in Chinese)

Xie P., and P. A. Arkin, 1997: Global precipitation: A 17-year monthly analysis based on gauge observations, satellite estimates, and numerical model outputs. *Bull. Amer. Meteor. Soc.*, **78**, 2539–2558.

Xu Qun, 2001: Abrupt change of the mid-summer climate in central east China by the influence of atmospheric pollution. *Atmospheric Environment*, **35**, 5029–5040.

Yu Rucong and Zhou Tianjun, 2007: Seasonality and three-dimensional structure of interdecadal change in the East Asian monsoon. *J. Climate*, **20**, 5344–5355.

Yu Rucong, Wang Bin, and Zhou Tianjun, 2004: Tropospheric cooling and summer monsoon weakening trend over East Asia. *Geo. Res. Lett.*, **31**, L22212, doi: 10.1029/2004GL021270.

Zhai Panmao, Ren Fumin, and Zhang Qiang, 1999: Detection of trends in China's precipitation extremes. *Acta Meteor. Sinica*, **57**, 208–216. (in Chinese)

Zhao Zongci, Ding Yihui, Li Xiongdong, and Wang Shaowu, 1995: Evaluation of CGCM climate simulation in East Asian region? *J. Appl. Meteor. Sci.*, **6**, 9–18.

Zhou T. J., and Z. X. Li, 2002: Simulation of the East Asian summer monsoon using a variable resolution atmospheric GCM. *Clim. Dyn.*, **19**, 167–180.