Diagnostic metrics for evaluation of annual and diurnal cycles

Bin Wang · Hyung-Jin Kim · Kazuyoshi Kikuchi ·
Akio Kitoh

Abstract Two sets of diagnostic metrics are proposed for evaluation of global models' simulation of annual and diurnal cycles of precipitation. The metrics for the annual variation include the annual mean, the solstice and equinoctial asymmetric modes of the annual cycle (AC), and the global monsoon precipitation domain and intensity. The metrics for the diurnal variation include the diurnal range, the land–sea contrast and transition modes of the diurnal cycle (DC), and the diurnal peak propagation in coastal regions. The proposed modes for the AC and DC represent faithfully the first two leading empirical orthogonal functions and explain, respectively, 82% of the total annual variance and 87% of the total diurnal variance over the globe between 45°S and 45°N. The simulated AC and DC by the 20-km-mesh MRI/JMA atmospheric general circulation model (AGCM) are in a wide-ranging agreement with observations; the model considerably outperforms any individual AMIP II GCMs and has comparable performance to 12-AMIP II model ensemble simulation measured by Pearson's pattern correlation coefficient. Comparison of four versions of the MRI/JMA AGCM with increasing resolution (180, 120, 60, and 20 km) reveals that the 20-km version reproduces the most realistic annual and diurnal cycles. However, the improved performance is not a linear function of the resolution. Marked improvement of the simulated DC (AC) occurs at the resolution of 60 km (20 km). The results suggest that better represented parameterizations that are adequately tuned to increased resolutions may improve models' simulation on the forced responses. The common deficiency in representing the monsoon domains suggests the models having difficulty in replicating annual march of the Subtropical Highs that is largely driven by prominent east-west land–ocean thermal contrast. Note that the 20-km model reproduces realistic diurnal cycle, but fails to capture realistic Madden-Julian Oscillation.

1 Introduction

Improvement of weather forecast and climate prediction models requires systematic evaluation and tracking of the models' performance. The evaluation is concerned with fundamental processes. Two types of process-oriented diagnostics are generally involved in the evaluation. One is the forced response of the climate system to external forcing such as solar radiation, which is primarily reflected on the annual cycle (AC) and diurnal cycle (DC). Another type of process involves internal feedback processes within the atmosphere or the coupled climate system, such as Madden-Julian Oscillation (MJO), El Nino-Southern Oscillation (ENSO), and other modes of climate variability. Recent MJO simulation diagnostics designed by US CLIVAR MJO working group is one of examples (Waliser et al. 2009).

Evaluation of the forced response is of essential importance and complementary to the evaluation of the internal feedbacks. High-quality simulation of the AC and

B. Wang (✉)
Department of Meteorology and International Pacific Research Center, School of Ocean and Earth Science and Technology, University of Hawaii at Manoa, Honolulu, HI 96922, USA
e-mail: wangbin@hawaii.edu

H.-J. Kim · K. Kikuchi
International Pacific Research Center, School of Ocean and Earth Science and Technology, University of Hawaii at Manoa, Honolulu, HI 96922, USA

A. Kitoh
Meteorological Research Institute, Tsukuba, Japan

DC may have a positive impact on simulation of climate variability (e.g. Sperber and Palmer 1996; Slingo et al. 2003), especially teleconnections that heavily depend on climatological basic states. Both the AC and DC involve a large number of radiative, dynamical, and thermal processes, hence the realism with which the basic annual and diurnal characteristics can be replicated in GCMs provides a critical indicator for assessing the models' skill. Thus, evaluations of the GCM performance have been intensively focused on the annual and diurnal cycles (Kang et al. 2002; Collier and Bowman 2004; Dai and Trenberth 2004; Ploshay and Lau 2010 among others).

Variety of ways to measure the models' performance on the annual and diurnal cycles has been used in the past. The most popular approach is the Fourier harmonic analysis (e.g. Horn and Bryson 1960; Hastenrath 1968; Hsu and Wallace 1976; Dai and Wang 1999). This method provides a convenient vector form for displaying characteristics of the first and second Fourier harmonics. However, when the cyclic response has significant asymmetry with respect to the mean, the Fourier analysis can potentially distort individual harmonics' amplitudes. For instance, typical monsoon precipitation often has a short and intense summer rainy season and a long and "flat" dry season; the Fourier analysis, in this case, may yield a fictitious semi-annual harmonic with considerable amplitude whereas reduce the amplitude of the annual harmonic (Wang 1994). Another approach is the phase-amplitude characteristic analysis without Fourier decomposition (Mokhov 1985). With this approach, the primary annual range and the apparent semiannual range (for bimodal variation) can be objectively defined to depict realistic magnitudes of the annual and semiannual variations; and the phase propagations can be displayed in a way similar to description of wave propagation (Wang 1994). A common weakness of the aforementioned approaches is associated with the difficulty to quantify the degree of success on the simulated diagnostic fields (e.g. the dial vector for each harmonics or the contours of phase lines).

In this study, we propose a new way of evaluation of the forced responses with a specific focus on designing metrics and specification of their objective measures. Such a systematic diagnostic package can facilitate quantifying models' scientific quality and uncertainties, comparing models' differences and revealing models' shortcomings. For convenience of discussion, we will focus on precipitation field, which is of particular interest as it represents a major source of diabatic heating that drives the tropical general circulation. The principles behind designing the metrics, however, are applicable to other fields.

We first describe the data and models used in this study in Sect. 2. Section 3 discusses the dominant modes of the annual and diurnal variations, which provide a clear motivation for the proposed metrics. Section 4 describes diagnostic metrics and objective measures used for evaluation of the models' simulation of the annual and diurnal cycles. Sections 5 and 6 provide a detailed assessment of the performance of an extremely high resolution AGCM, 20-km-mesh Meteorological Research Institute/Japan Meteorological Agency (MRI/JMA) AGCM. For inter-model comparison, the MRI/JMA model with four different resolutions and twelve GCMs that participated in Intergovernmental Panel for Climate Change (IPCC) fourth assessment (AR4) experiment (the second phase of the Atmospheric Model Intercomparison Project, AMIP II) were also examined. The cause of the models' common deficiencies is discussed. Section 7 examines the impact of the model resolution on the performance. Finally, summary and discussion are presented in the last section.

2 Data and models

Global Precipitation Climatology Project (GPCP) (Adler et al. 2003) dataset (1979–2008), the Climate Prediction Center Merged Analysis of Precipitation (CMAP) (Xie and Arkin 1997), and two datasets derived from 10 years (1998–2007) of Tropical Rainfall Measuring Mission (TRMM) observation were used for validation. The TRMM 3G68 version 6 data (information can be found at ftp://trmmopen.gsfc.nasa.gov/pub) provide precipitation at $0.5° \times 0.5°$ resolution based on TRMM swath data alone. The TRMM 3B42 version six data provide higher spatial resolution ($0.25° \times 0.25°$) and three hourly data, which were created by blending four types of passive microwave data including TRMM Microwave Image (TMI) and infrared (IR) data and carefully calibrated using monthly rain gauge data (Huffman et al. 2007). As shown by Dai et al. (2007) and Kikuchi and Wang (2008), the 3B42 data have better spatial patterns due to large sampling size while the 3G68 data (precipitation radar observation) provide more reliable diurnal phase information. The 3B42 data were estimated from TMI and IR, both are strongly affected by the presence of anvil clouds including non-precipitating cirrus type which develop after precipitating deep convective clouds. As such the 3B42 diurnal phase tends to lags that derived from the 3G68.

The MRI/JMA AGCM with 4 progressively higher resolutions (equivalent to 180, 120, 60, and 20 km) were evaluated. The highest resolution of the model, TL959L60, has 60 layers in vertical and triangular truncation at 959 with a linear Gaussian grid in horizontal, which is equivalent to a 20-km-mesh. This hydrostatic and spectral model was developed from a global numerical weather prediction model (JMA 2002) and has been used for both medium-range forecasts at JMA (Murakami and Matsumura 2007)
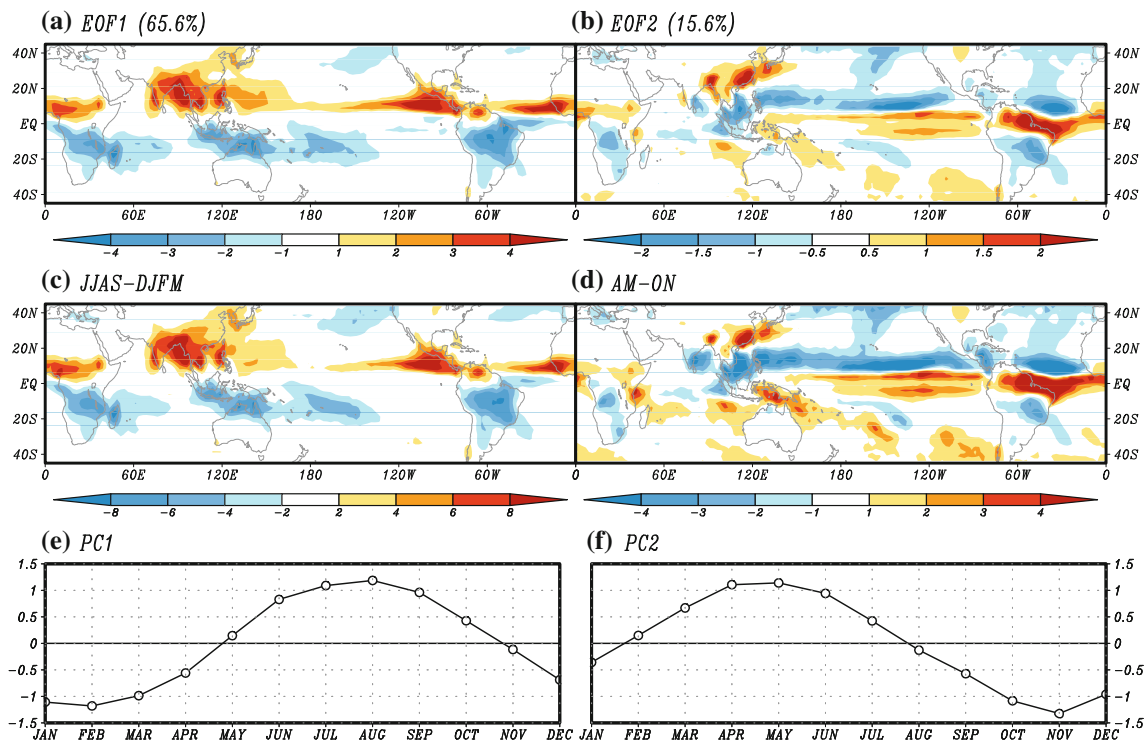
Fig. 1 **a**, **b** The spatial patterns of the first two EOF modes of the climatological monthly mean precipitation rate (mm/day). **c** The solstice mode of the annual variation (AC1) as described by the June–September (*JJAS*) minus December–March (*DJFM*) mean precipitation rate. **d** The equinoctial asymmetric modes (AC2) as described by the April–May minus the October–November mean precipitation rate. **e**, **f** The corresponding principal components PC1 and PC2. Here, GPCP data were used

and climate projections at MRI (Mizuta et al. 2006; Oouchi et al. 2006; Kusunoki et al. 2006; Kitoh and Kusunoki 2008). A detailed description of the model can be found in Mizuta et al. (2006). The 180-, 120- and 60-km versions used the same physics package except for resolution-dependent parameters. The 20-km model was specially tuned to reduce biases. The model was integrated for 25 years (1979–2003) using HadISST1.1/sea ice original $1.0° \times 1.0°$ data (Rayner et al. 2003) but interpolated to fit each version's resolution. To facilitate multi-model inter-comparison, twelve (six) AGCMs that participated in the AMIP II were used for AC (DC) evaluation. The period for AC evaluation spans from 1980 to 1999 for both observations and model simulations. But for DC evaluation, since 3-hourly data are required, the DC climatology was derived from a 10-year period (1998–2007) for TRMM, 25-year (1979–2003) for MRI/JMA AGCM, and only 1-year (2000) data for the AMIP II models. In the AMIP II, AGCMs have been run with observed lower boundary conditions of SST and sea ice concentration and their simulations were verified by the Program for Climate Model Diagnosis and Intercomparison (PCMDI) quality control processes. For fair comparison of the AC (DC) simulation among the models a common grid system of $2.5° \times 2.5°$ ($2.0° \times 2.0°$) was applied to the all model

outputs with bi-linear interpolation. However, evaluation of the 20-km version model used high resolution ($0.25° \times 0.25°$) TRMM data.

3 Dominant modes of the annual and diurnal variations

Different from Fourier harmonic and amplitude-phase analyses, the method for description and evaluation of the AC and DC here is based on empirical orthogonal function (EOF) analysis of spatiotemporal characteristics of the AC and DC. It has been shown that for both AC (Wang and Ding 2008) and DC (Kikuchi and Wang 2008), the amplitude and phase propagation can be faithfully represented by the first two leading EOF modes.

Figure 1a, b present spatial patterns of the first two leading EOFs of climatological monthly mean precipitation; the two EOFs account for, respectively, 66 and 16% of the total annual variance of precipitation. Both EOFs describe the annual cycle (Fig. 1e, f). The first EOF features an inter-hemispheric contrast with the maximum and minimum occurring in July and February, respectively, which reflects the atmospheric response to the solar forcing in solstice seasons with a 1-to-2-month phase delay. This leading EOF was termed as solstice mode of the AC (Wang
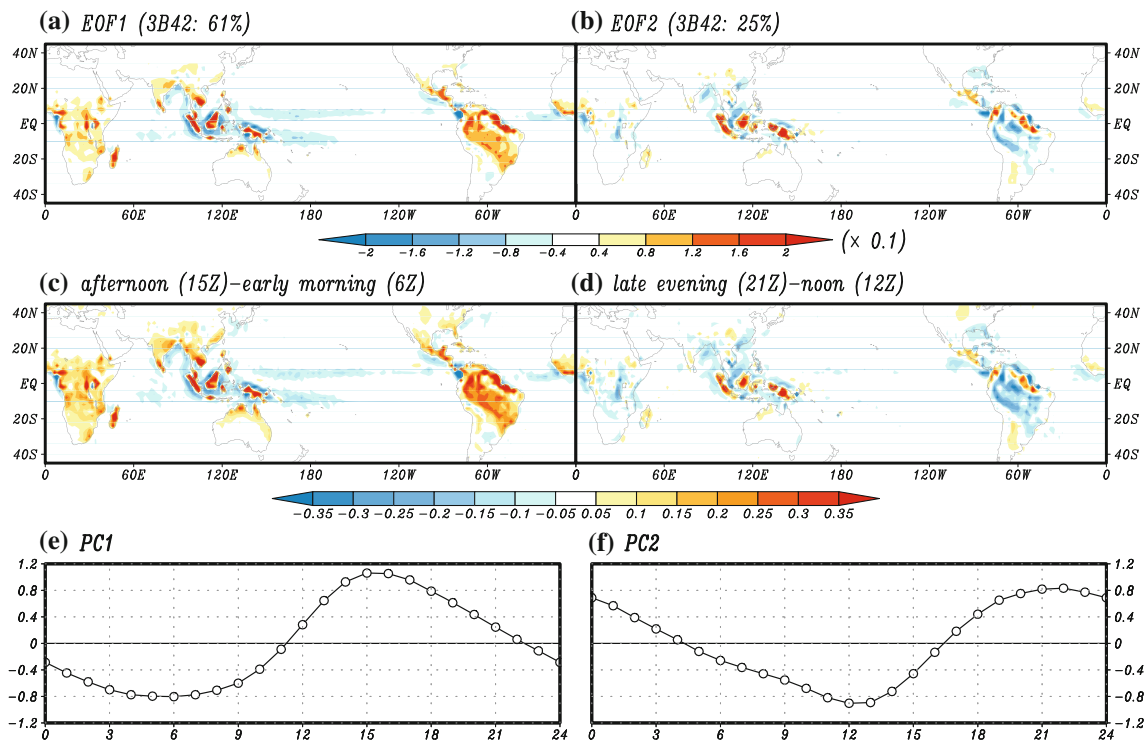
Fig. 2 **a**, **b** The spatial patterns of the first two EOF modes of the climatological diurnal precipitation rate (mm/h). **c** The land–sea contrast mode of the diurnal variation (DC1) as described by afternoon (*15Z*) minus early morning (*6Z*) precipitation rate. **d** The transition mode (DC2) as described by late evening (*21Z*) minus noon (*12Z*) precipitation rate. **e**, **f** The corresponding principal components PC1 and PC2. Here, TRIMM 3B42 and 3G68 were used (See the text for details)

and Ding 2008). The second EOF has the maximum and minimum occurring around April and October, respectively, reflecting a spring-fall asymmetric response, which was called the equinoctial asymmetric mode (Wang and Ding 2008).

Comparison of Fig. 1a, c indicates that the first EOF (the solstice mode) can be captured extremely well by the June through September (JJAS) mean minus December through March (DJFM) mean precipitation rate. Similarly, the second EOF (the equinoctial asymmetric mode) can be realistically represented by the April–May minus the October–November mean precipitation rate (Fig. 1b, d). These annual differences offer simple measures for gauging the AC.

For precipitation DC, the spatial patterns and the corresponding principal components of the first two leading EOFs are shown in Fig. 2. The two diurnal EOFs explain 61 and 25% of the total diurnal variance. Note that the observed spatial patterns (Fig. 2a, b) were derived from 3B42 dataset and the principal components (Fig. 2e, f) were derived from 3G68 dataset. The reason for use mixed datasets is that the EOFs derived from the two datasets are essentially the same modes, but the 3B42 data provide a better spatial pattern due to its dense spatial coverage while

the 3G68 data provide a more accurate diurnal phase (See Sect. 2 for details).

The EOF1 of DC represents a contrasting land–sea regime that features an afternoon peak over land with large amplitude and an early morning peak over ocean with moderate amplitude. Thus, the first EOF will be called land–sea contrast mode. The afternoon peak over land is particularly pronounced in South America, the Maritime continent, and equatorial Africa near Lake Victoria (Fig. 2a). This land afternoon peak is due to solar heating-induced afternoon maximum convective available potential energy that favors moist convection and showery precipitation (e.g. Dai 2001). The early morning peak is found primarily over the oceanic convergence zones in the Pacific, Atlantic, and Indian Oceans. The EOF2 mode depicts a DC with a maximum in late evening (Hour 21) and a minimum at noon (Hour 12). Note that the main peak of the EOF2 lags that of the EOF1 by about 6 h (a quarter of diurnal cycle, Fig. 2e, f) and this phase shift indicates a propagation feature of the diurnal cycle by combination of the two modes. For convenience, we refer the second EOF as a (complementary) transition mode. The salient propagation of diurnal cycle peak is seen along the land–sea boundaries of the Maritime Continent, the Indian

subcontinent, northern Australia, the west coast of America extending from Mexico to Ecuador, the west coast of equatorial Africa, and Northeast Brazil (Fig. 2b).

Figure 2a, c indicate that the land–sea contrast mode (EOF1) can be realistically expressed by the diurnal difference between afternoon (Hour 15) and early morning (Hour 6). Similarly, the EOF2 pattern (transition mode) can be very well represented by the diurnal difference between late evening (Hour 21) and noon (Hour 12) (Fig. 2b, d). Note also that since the diurnal phases depicted by the 3B42 lag those depicted by the 3G68 by about 3 h, in construction of Fig. 2c, d using 3B42 data, this phase shift had been taken into account. We will use these diurnal differences as measures for evaluating DC.

4 The metrics for evaluation of the annual and diurnal cycles

4.1 The annual cycle

A set of core diagnosis maps is designed to evaluate models' performance on global precipitation, including (1) annual mean (AM), (2) the solstice mode, i.e., JJAS (June–September) minus DJFM (December–March) mean (AC1), (3) the equinoctial asymmetric mode, i.e., AM (April–May) minus ON (October–November) mean (AC2), (4) global monsoon precipitation domain, and (5) monsoon precipitation intensity.

We used global monsoon domain and intensity as additional diagnostic measures because delineation of monsoon regime requires integrated information about the annual mean rainfall, the amplitude of annual range, and the local seasonal distribution of the rainfall (Wang 1994). An objective measure of monsoon precipitation intensity is extremely useful because the strengths of the global and regional monsoons vary on variety of time scales ranging from interannual to geological. It has been demonstrated that the global monsoon precipitation intensity is a lucid measure of global climate variations (Liu et al. 2009a). Therefore, examination of competence of the models in capturing global monsoon precipitation domain and the monsoon precipitation intensity is another desirable attribute.

Since the contrast between rainy summer and dry winter is an essential characteristic of monsoon climate, global monsoon precipitation domain is delineated as the annual range of precipitation rate exceeds a threshold of 2.5 mm/day, where

Annual range = local summer minus winter mean

precipitation rate (1)

Here we define the annual range of precipitation by MJJAS minus NDJFM in NH and the reversal in SH following Wang and Ding (2008). The suggested criterion means the annual range exceeds 375 mm. In observation, use of this criterion distinguishes the monsoon precipitation regime very well from the adjacent dry regime (desert, trade wind, and Mediterranean regimes) and the equatorial perennial rainfall regime. The annual range defined here can also offer a simple measure for monsoon strength. However, the annual range generally decreases with latitude, to better depict the degree of concentration of precipitation in summer, we used the following normalized annual range to measure monsoon strength which preserves the "relativity" in describing monsoon characters:

Monsoon precipitation intensity (MPI)
= (Annual range)/(Annual mean) (2)

The MPI is computed over each model grid. The global monsoon intensity can be measured by area-weighted average of MPI at each grid within the global monsoon domain (Wang and Ding 2006). In a similar way, the Northern Hemisphere (NH) and Southern Hemisphere (SH) monsoon precipitation intensities can be defined. This definition can be also extended to quantify interannual variation of the global monsoon precipitation.

4.2 The diurnal cycle

The core diagnosis fields for assessing performance on precipitation diurnal cycle include (a) diurnal range (amplitude), (b) the land–sea contrast mode, i.e., Hour 15 minus Hour 6 precipitation rate (DC1), (3) the transition mode, i.e., Hour 21 minus Hour 12 precipitation rate (DC2), and (4) the diurnal peak phase diagram to depict propagation of the DC in coastal regions.

4.3 Objective measures

For objective quantification of the models' performance and to facilitate multi-model comparison, statistical measures are applied to the proposed metric fields. Pearson pattern correlation coefficient (PCC) (Wilks 1995) was used to gauge the degree of similarity in the spatial patterns between the observed and simulated fields. The similarity of co-variability does not indicate if the simulation is of the right magnitude, hence is not a strict measure of accuracy. For this reason, domain-averaged (with areal weight) root mean square error (RMSE) was used to measure typical simulation errors. To appraise performance on the monsoon precipitation domains, we have used the threat score: the number of hit grids divided by the sum of hit, missed, and false-alarm grids (Wilks 1995). Here the hit grid means the grid at which simulation and observation agree with each other; the missed grid means an observed grid being
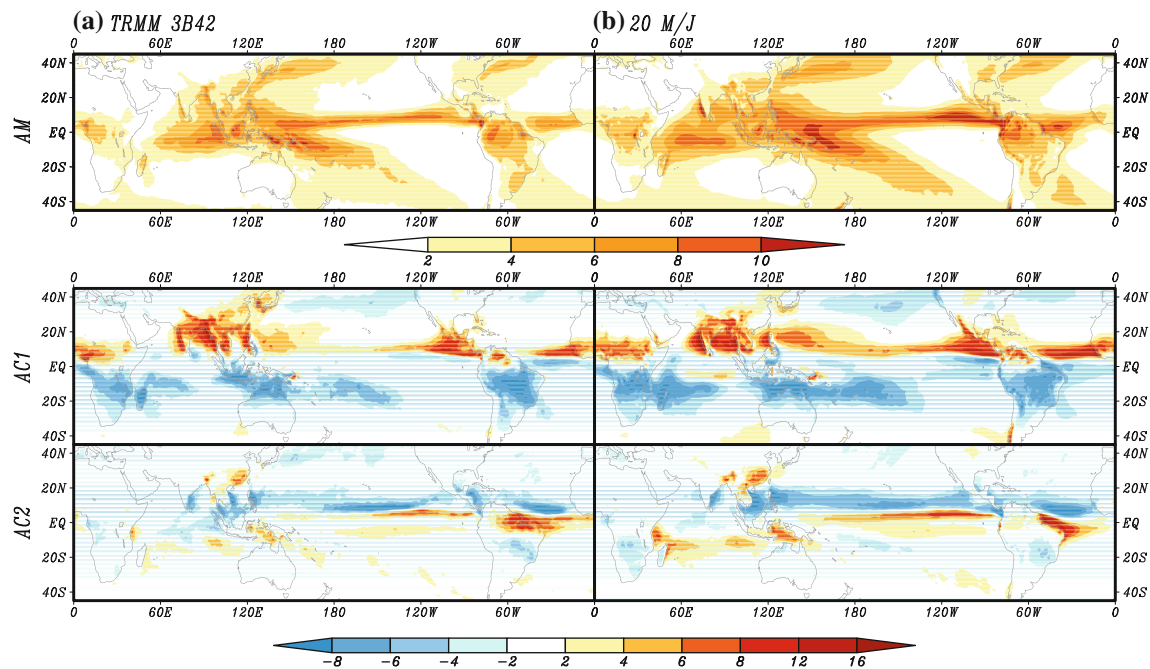
Fig. 3 a Observed annual mean precipitation rate (*top*), the solstice mode (*AC1*) as depicted by JJAS minus DJFM mean precipitation rate (*middle*), and the equinoctial asymmetric mode (*AC2*) as depicted by April–May (*AM*) minus October–November (*ON*) mean precipitation rate (*bottom*). **b** The model counterparts obtained from 20-km-mesh MRI/JMA GCM forced by the observed SST. The units are mm/day. The horizontal resolution is set to 0.25° × 0.25° in line with the original resolution of TRMM 3B42 data

missed in the simulation; and the false-alarm grid means a grid that is recognized by the model but not by observation. The threat score varies from 0 to 1 with 0 being the worst score. A higher score means a better match between observation and simulation.

5 Evaluation of annual variation

5.1 The 20-km MRI/JMA AGCM

Figure 3 presents evaluation of the annual mean precipitation and the solstice and equinoctial modes of the AC against the TRMM 3B42 data set. Here, the spatial resolution of 0.25° × 0.25° was used for detailed comparison. The simulated spatial pattern of the annual mean field agrees well with the observed (top panel), in particular the position of the Pacific intertropical convergence zone (ITCZ) and South Pacific convergence zone (SPCZ), the local maximum precipitation in the northeast coasts of Arabian Sea and Bay of Bengal as well as along the Himalayas. However, the magnitude is generally larger in the model than the observation; in particular excessive precipitation appears in the far western Indian Ocean, Atlantic ITCZ and Central America. The simulated two AC modes, referred to as 'annual cycle 1 (AC1)' and 'annual cycle 2 (AC2)', also bear a close resemblance to the

corresponding observed counterparts in terms of spatial patterns and time evolutions (middle and bottom panels). The pattern correlation coefficients (PCC) of the simulated AC1 and AC2 with their corresponding observed counterparts are 0.86 and 0.78, respectively. But the rainfall variation center of the AC1 located at the Western Ghats of India is extended farther to the west and the intensity is overestimated by about a factor of 2. The annual cycle of the ITCZ is also too strong, which is often seen in other coupled climate models (e.g. Kim et al. 2008). The modeled AC2 spatial pattern exaggerates the amplitude over the tropical regions, especially over the Pacific and Atlantic Oceans. The model's bias in simulation of the AC2 mode appears to be similar to those in other models in which horizontal resolution is considerably coarser than the 20-km (e.g. Kim et al. 2008; Lee et al. 2010), suggesting that the cause of the deficiencies are not totally due to resolution. In the western Hemisphere, the spring-fall rainfall contrast is primarily related to seasonal variations in the equatorial cold tongues in the Pacific and Atlantic. In the eastern Hemisphere, on the other hand, the wind departure associated with the seasonal changes of land–ocean thermal contrast and inertia is the primarily reason for the spring-fall asymmetry (Wang and Ding 2008). The common bias in simulation of AC2 implies that the models have deficiencies in capturing these processes.
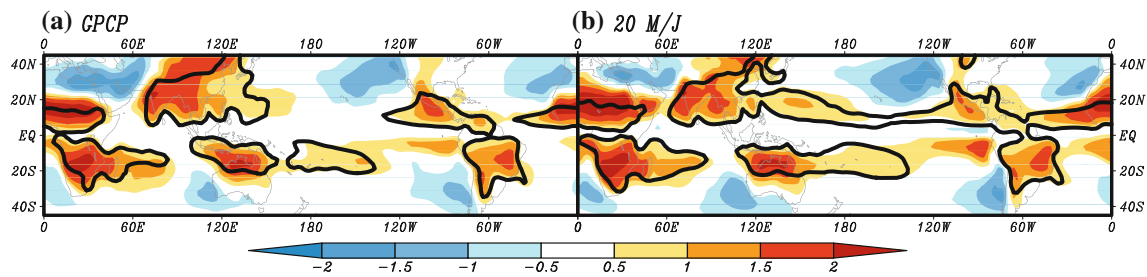
Fig. 4 a Observed monsoon precipitation domain defined by using the GPCP data by the annual range greater than 2.5 mm/day (outlined by the *thick solid lines*) and monsoon precipitation intensity (*color shadings*). **b** The model counterparts of **a** derived from the 20-km-mesh MRI/JMA GCM. The negative intensity means a wet winter and dry summer (typical Mediterranean climate)

Figure 4 compares the observed (left panel) and simulated (right panel) monsoon precipitation domain and intensity. The GPCP data are used as observed data (so is in the following Figs. 5, 6, 7) because remapping the outputs of coarser resolution model onto the TRMM's high spatial resolution might lead to misinterpretation of the inter-model comparison. The observed global monsoon precipitation domain includes six major continental monsoon regions. Note that most of the regional monsoon embraces adjacent marginal seas and oceanic regions. Thus, the oceanic monsoon regions are extensions of the continental monsoons and integral parts of the regional monsoon systems. In the subtropical mid-South Pacific, however, a pure oceanic region exists that has similar seasonal distribution as the monsoon regime. But, due to lack of land–ocean thermal contrast, this oceanic region will not be considered as a typical monsoon region. The simulated monsoon precipitation domain is in a general agreement with observation on the global scale, especially over the SH (Fig. 4b). But, regionally, there are marked differences. For East Asian monsoon domain, for example, the model failed to capture substantial portions of southeastern China and the East China Sea; meanwhile, the monsoon domain over Philippine Sea is exaggerated. As a result, the oceanic monsoon domain over the western North Pacific is extended all the way to the eastern North Pacific and eventually connected with the Central-North American monsoon (see Sect. 5.3 for further discussion). In addition, the separation between the Indonesia–Australia monsoon and the subtropical mid-South Pacific regime is unclear. These problems seem to be related to the exaggerated AC1 (Fig. 3b), and the errors in the simulated seasonal distribution of precipitation.

5.2 Multi-model intercomparison

Figure 5 compares various models' performance against observation on the yearly mean, annual cycle (AC1 and AC2), and monsoon precipitation index in terms of PCC

and domain-averaged RMSE normalized by the observed spatial standard deviation over the globe between 45°S and 45°N. In order to estimate observational uncertainty, we have computed the PCC and RMSE between the two observed (estimated) datasets, i.e., CMAP and GPCP rainfall. The PCCs are 0.96 for both the annual mean and the first and second annual variation mode. The RMSE is 0.40, 0.34 and 0.33, respectively, for the annual mean, and the first and second mode of annual variations. It means that the patterns between GPCP and CMAP are highly similar, but the amplitudes of these fields differ (with CMAP having larger values). The fidelity of the annual mean and annual cycle modes reproduced in the models' simulations should be considered with this uncertainty of satellite observation in mind.

It is first noted that the PCC and RMSE for the AMIP II models tend to have a linear relationship. This linearity is significant for the AC1 and MPI with a statistical confidence level higher than 95%, suggesting that an AGCM with a higher PCC tends to have a lower RMSE. Secondly, the spatial similarity between simulation and observation is worst for AC2 and best for the annual mean without any exception. Similarly, the precipitation biases are largest in the AC2 and, in general, smallest in the annual mean. The same is true for the 20-km MRI/JMA AGCM simulation. It is also of interest to note that the spread of PCC among the 12 models is smallest for the annual mean and largest for the AC2.

Results presented in Fig. 5 show that the MRI/JMA 20-km resolution model generally outperforms the lower-resolution AGCMs in comparison. In terms of PCC, in particular, the 20-km MRI/JMA model is comparable to the 12-model's multi-model ensemble (MME) simulation. Here the MME was obtained by a simple arithmetic average among the 12 AMIP models. The 20-km model has lager RMSE than that of the MME when compared to GPCP data, but given the uncertainty in the estimated intensity between CMAP and GPCP, no conclusion can be made. Overall, the MME is a very effective way to improve

Fig. 5 Evaluation of the
performance of the 12 AGCMs'
and four versions of the MRI/
JMA AGCMs' simulation on
climatology: **a** annual mean
precipitation rate, **b** the solstice
mode (JJAS minus DJFM), **c** the
equinoctial asymmetric mode
(AM minus ON), and
d monsoon precipitation
intensity. The *abscissa* and
ordinates are pattern correlation
coefficient (*PCC*) and domain-
averaged RMSE normalized by
the observed spatial standard
deviation, respectively. Note
that 120 and 60 M/J are
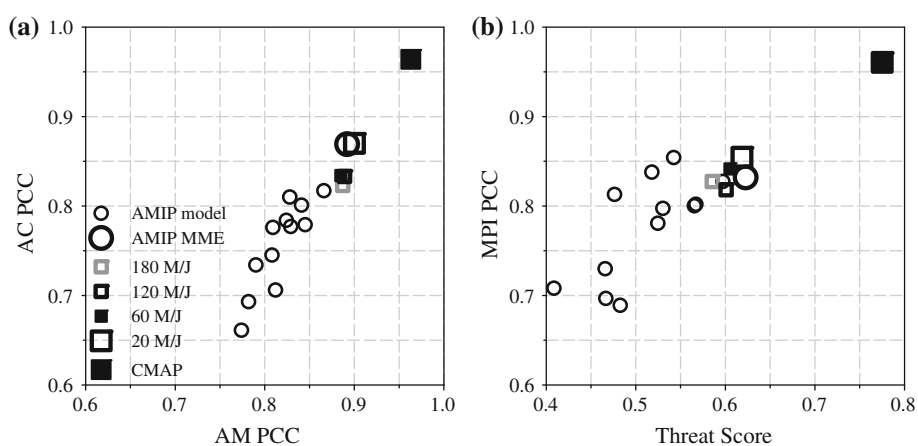overlapped in **a**. The domain
used is 0–360E, 45S–45N
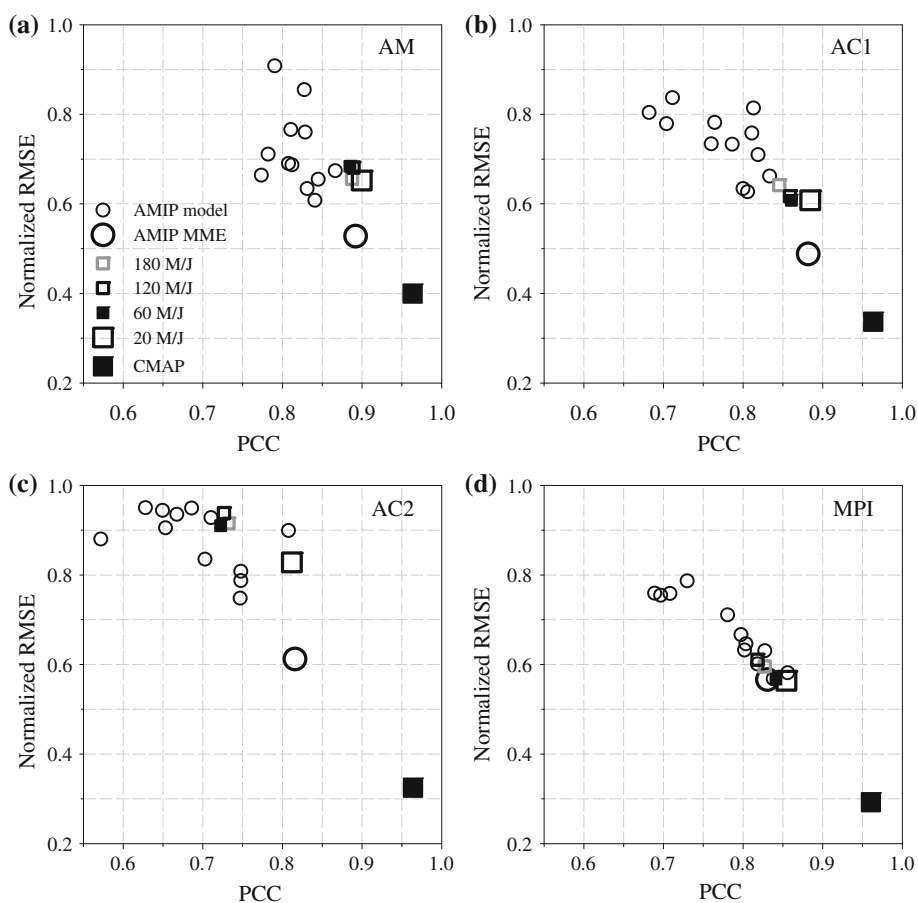
Fig. 6 Overall assessment of
the 12 AGCMs' and four
versions of the MRI/JMA
AGCMs' performance. **a** PCCs
of annual mean (*abscissa*)
versus annual cycle (*ordinate*),
and **b** threat score of monsoon
domain (*abscissa*) versus PCC
of the climatological monsoon
precipitation intensity (MPI).
The domain used is 0–360E,
45S–45N

model's performance. However, the advantage of high
resolution model lays in its strength in resolving extreme
events.

Figure 6a, b summarize multi-model performance in
simulating the annual precipitation modes and global
monsoon precipitation, respectively. The PCC of annual
cycle in Fig. 6a is constructed from the combined PCCs of
the AC1 and AC2 weighted by their corresponding frac-
tional variance (e.g., 0.66 for AC1 and 0.16 for AC2 in
GPCP). Thus, the annual cycle PCC reflects combined

performance on the AC1 and AC2 simulation. Result of
Fig. 6 indicates that, by means of these statistical param-
eters, the performance of a model is quantifiable and, more
importantly, discernable from one model to another. The
prominent linearity ($\gamma^2 = 0.76$) shown in Fig. 6a indicates
a close linkage of PCCs between the annual mean and
annual cycle of the modeled precipitation. Similarly, but to
a lesser degree, the simulated monsoon precipitation
domain (threat score) and intensity (MPI PCC) are also
correlated ($\gamma^2 = 0.51$, Fig. 6b). Results of Fig. 6 confirm

Fig. 7 Comparison of the performance of 12 AGCMs and four versions of the MRI/JMA AGCM on simulation of monsoon domains (threat score, *abscissa*) and monsoon precipitation intensity (PCC, *ordinate*) for seven regional monsoons and the NH and SH monsoons (**b–j**). **a** Shows each regional monsoon domain used for PCC and threat score calculation. The shading represents the observed monsoon precipitation domain
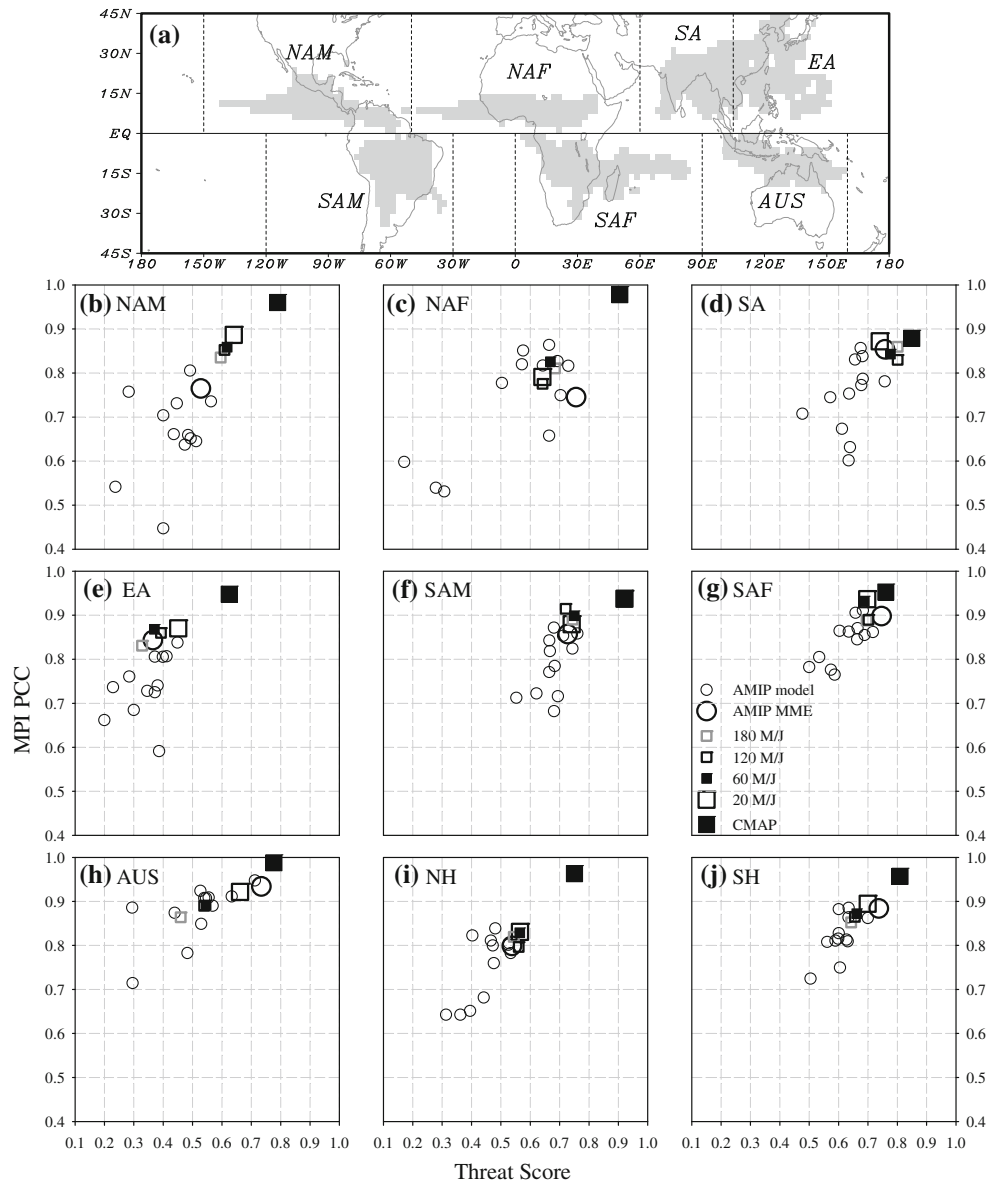
that the 20-km MRI/JMA model considerably outperforms all individual AMIP II models and is comparable to the 12-model's MME simulation.

5.3 A higher-level evaluation: regional aspect of monsoon precipitation

The method for evaluation of global monsoon domain and intensity is equally applicable to evaluation of the regional monsoon precipitation intensity. In view of the distinctive features between the Indian monsoon and East Asian monsoon (Tao and Chen 1987; Wang et al. 2001), the Asian monsoon system is further divided into two sub-systems: the south Asian (or Indian) and East Asian (including western North Pacific) monsoon by an artificial

boundary along the east periphery of the Tibetan Plateau at 105°E (Fig. 7a).

Figure 7 compares performance of 12 AMIP II model's and their MME as well as various versions of the MRI/JMA model on simulation of regional monsoon domains and corresponding monsoon precipitation intensity. Here, the GPCP monsoon precipitation domain is uniformly applied to all models' data sets to calculate regional monsoon precipitation intensity. Comparison is also made between CMAP and GPCP to show the uncertainties between the two observed precipitation datasets. Overall, the monsoon precipitation intensities (the PCC of MPI) are significantly better simulated than the monsoon domains (the threat score) for all models. The 20-km MRI/JMA AGCM shows a better or comparable skill compared to the

Table 1 Climatological area-weighted mean monsoon precipitation intensity (annual range/annual mean) over each regional monsoon domain that was determined using GPCP data (see Fig. 7a for global (GLO) and regional monsoon domains)

	NAM	NAF	SA	EA	SAM	SAF	AUS	NH	SH	GLO
GPCP	1.05	1.52	1.57	0.93	1.20	1.41	1.31	1.24	1.31	1.27
CMAP	1.09	1.56	1.62	1.01	1.22	1.42	1.33	1.29	1.33	1.31
20 M/J	0.85	1.46	1.44	0.74	1.17	1.43	1.13	1.09	1.27	1.17
MME	0.60	1.15	1.39	0.68	1.29	1.34	1.03	0.93	1.24	1.06

MME of 12 AMIP models in all regional monsoons. In general, the monsoon domains simulated by the models have better performance on three SH regional monsoons and the South Asian monsoon. The threat scores of the MME for these four regions range from 0.73 to 0.76. However, East Asian and North American monsoon domains turn out to be poorly simulated: The MME threat score is 0.37 and 0.53 for East Asian and North American monsoon, respectively.

The causes of such failure deserve further analysis. A part of the monsoon domain over East Asia is commonly missed in the models, whereas over the western North Pacific and eastern US the domains are often false-alarmed resulting in the relatively poor threat scores over those two particular regions (figure not shown). East Asia and eastern North America are the regions where pronounced east-west land–ocean thermal contrast exists. The models seem to have difficulties to capture the correct annual march of the circulation systems, especially the Subtropical High under such an east-west thermal contrast settings (Kang et al. 2002). The east-west land–sea thermal contrast and the topographic effect (Tibetan High and Rocky mountains) have fundamental influences on the strength and position of the western North Pacific and North Atlantic Subtropical Highs during boreal summer (Wu and Liu 2003; Wu et al. 2009; Wang et al. 2008). These two systems are critical for correctly reproducing seasonal distribution of monsoon precipitation.

The climatological monsoon precipitation intensities for each regional monsoon and global monsoon are also computed using GPCP, CMAP, 20-km MRI/JMA AGCM, and 12 models' MME mean. The results are presented in Table 1. The observational uncertainty measured by the difference of the area-averaged MPI between CMAP and GPCP is about 3% over the globe, with CMAP having larger values in all regional monsoon domains. The largest difference (9%) is seen in the East Asia monsoon region. Of particular interest is that the area-averaged MPI of the MME is significantly underestimated by about 5–40% depending on the regions with an exception of the South America monsoon. Further examination reveals that the underestimated MPIs can be explained by the excessive

local winter precipitation in the models. Over the monsoon domains, the overestimation of the local winter rainfall surpasses that in the local summer rainfall. Accordingly, the annual range is reduced, yet the annual mean is increased, thereby MPIs are reduced significantly. For MRI/JMA AGCM, both the annual range and annual mean are over-predicted. But the increase in annual mean is larger, resulting in decreased MPI but to a lesser degree compared to other models.

An interesting fact is that the MME does so well in capturing the monsoon domain and the two annual cycle modes (Figs. 5, 6), but poorly in representing the monsoon precipitation intensity (Table 1). This indicates that the MME of the AMIP II models do well in matching patterns but not so well in intensity. The low-resolution models appear handicapped by their predominantly relatively coarse resolution which usually means less intense simulations.

6 Evaluation of the diurnal variation of precipitation

Figure 8 compares the diurnal range of precipitation that are observed (TRMM 3B42) and simulated in the 20-km MRI/JMA model. Again, the spatial resolution of $0.25° \times 0.25°$ is used. The overall spatial distribution of diurnal amplitude simulated in the model is realistic, especially the observed large amplitude over the Maritime Continent, Central America, northeast coast of Brazil, and equatorial West African coast. However, the model tends to overestimate the amplitude over land such as the tropical western South America, the equatorial central Africa and South Asia. The regions of overestimated amplitude tend to coincide with the regions of excessive annual precipitation.

Figure 8 also compares the first two leading modes of diurnal rainfall variations between the observed and simulated in the 20 km MRI/JMA model. It is impressive that the simulated spatial patterns of both the land–sea contrast mode as measured by the Hour 15 minus Hour 6 precipitation rate and the transition mode measured by the Hour 12 minus Hour 21 precipitation rate are in good agreement with observations in terms of both the fractional variance
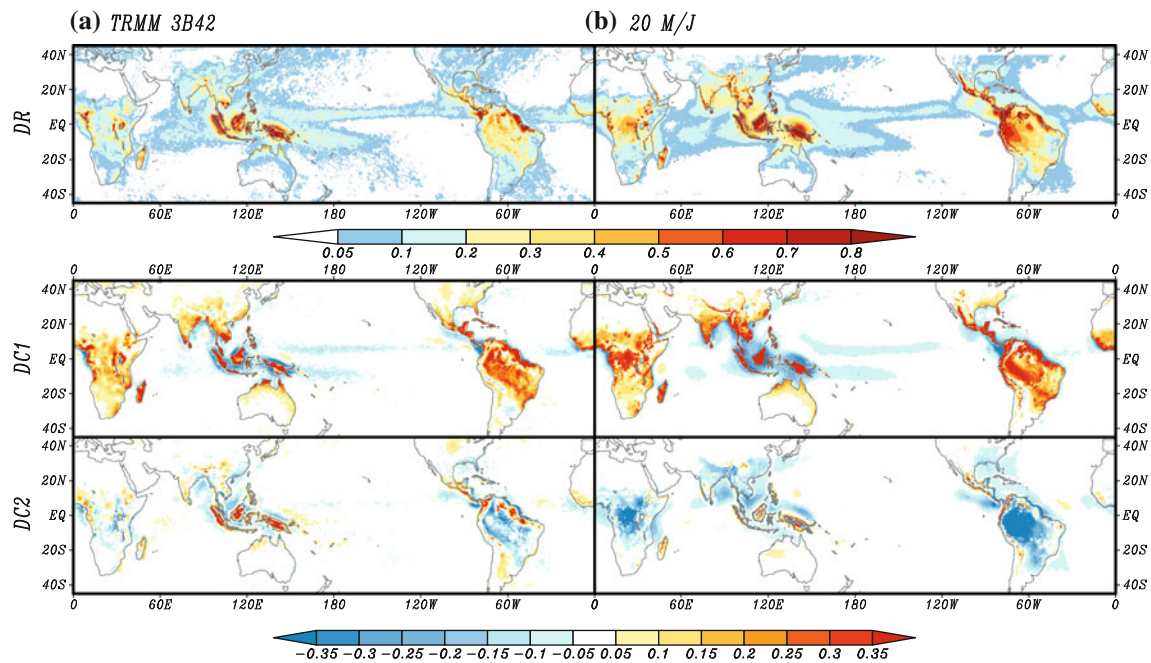
Fig. 8 **a** Observed diurnal range of precipitation rate (*top*), the land–sea contrast mode (DC1) as depicted by afternoon (*15Z*) minus early morning (*6Z*) precipitation rate (*middle*), and the transitional mode (DC2) as described by the late evening (*21Z*) minus noon (*12Z*) precipitation rate (*bottom*). **b** The model counterparts obtained from 20-km-mesh MRI/JMA GCM forced by the observed SST. The units are mm/h. The TRMM 3B42 data are derived from the period of 1998–2007 and the simulated in the 20-km MRI/JMA AGCM are derived from the period of 1979–2005. The horizontal resolution is set to $0.25° \times 0.25°$ in line with the original resolution of TRMM 3B42 data. Note that 3 h time shift is made to 3B42 data

and spatial patterns. The amplitude of the land–sea contrast mode in the model, however, is over-estimated especially over the South America. This is consistent with the overestimated diurnal amplitude.

The diurnal precipitation exhibits complex propagation patterns in coastal areas, thus, detailed diurnal propagations in the coastal regions are further evaluated in terms of local peak time in precipitation (Fig. 9). Observation shows two propagation features, often concurring along the same land–sea boundary. The seaside propagation is characterized by offshore phase propagation, with peaks occurring from late evening to noon of the next day (2100–1200 LST). The offshore propagation is partially due to land surface cooling-induced offshore convergence over nearby oceanic regions. The landside propagation has inland phase propagation with peaks occurring from noon to evening (1200–2100 LST). Although the model tends to produce earlier and stronger development of diurnal precipitation, the propagation features are well reproduced in each major monsoon region. For instance, the off-shore movement over water and the in-land movement over land can be found in the Maritime Continent and the coastal areas of the Bay of Bengal (Fig. 9a), the west Africa and Madagascar (Fig. 9b), and Central America and northeast Brazil (Fig. 9c). Therefore, the model seems to capture

well the fundamental physical processes responsible for the propagation in the coastal areas. The major deficiency of the model is the excessively large amplitude, which can be seen in most regions. Another discrepancy is seen over the central area of the Maritime Continent where the maximum precipitation takes place in the afternoon (14–17 LST) in the model, a few hours earlier than observation (18–23 LST). In comparison with other 6 lower resolution AMIP II models, the 20 km MRI/JMA model is able to simulate most realistic diurnal cycle in terms of diurnal range and the two leading diurnal cycle (DC1 and DC2) modes (Fig. 10). Results in Fig. 10 indicate that all the MRI/JMA simulations are better than the AMIP II models individually, which would argue that the physics of this MRI model might be better. Its vertical resolution is also better. These physical processes include heat balance at the land surface and associated boundary layer changes of thermal structure and stability, differential heating/cooling along terrain slopes, diurnal contrasts in longwave radiative cooling of clouds, diurnal variation of sea surface temperature, and excitation of gravity waves which carry away diurnal signals of convection from the source region (Wallace 1975; Mapes et al. 2003; Yang and Smith 2006; Johnson 2010; Ploshay and Lau 2010).
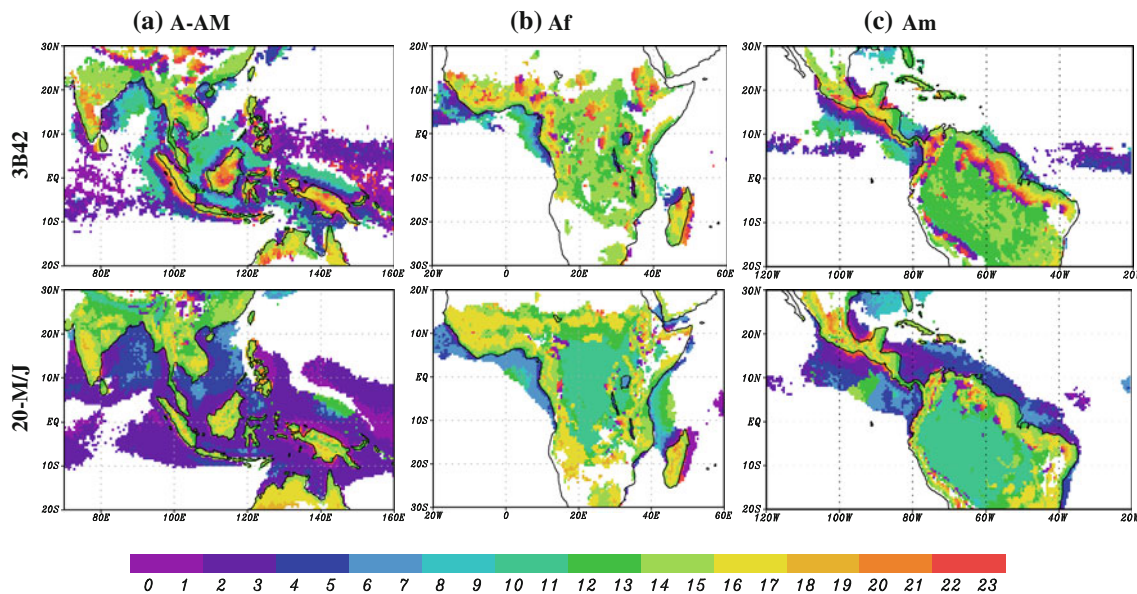
Fig. 9 Evolution of the diurnal cycle in the coastal areas in terms of peak local time in precipitation for **a** Asian–Australian, **b** African, and **c** American sectors. For each sector the *upper panels* are observed (TRMM 3B42) and the lower panels are the simulated with the 20 km MRI/JMA model. The peak local time is shown by *shades* with 2 h increment. Note that the local peak time in 3B42 precipitation is estimated with hourly data constructed using the Fourier transform and is shifted forward by 3 h to correct its biases (see Sect. 2 for more details). Insignificant diurnal cycles with DR less than 2.4 mm/day are masked. The observed and simulated climatology is derived for 10 years (1998–2007) and 25 years (1979–2003) data, respectively
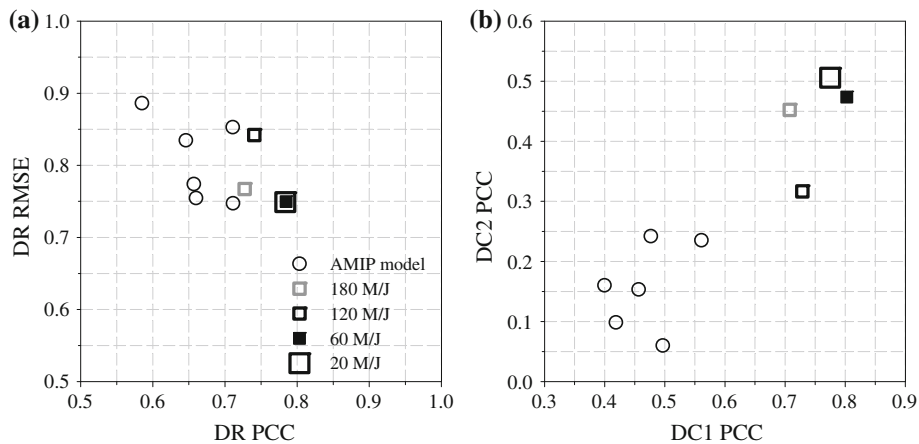

Fig. 10 Objective assessment of climatological diurnal cycle simulated in various AGCMs. The diurnal cycle is evaluated in terms of **a** diurnal range (DR) and **b** land–sea contrast (DC1) and transition (DC2) modes. In **a**, the *abscissa* is pattern correlation coefficient (PCC) between TRMM 3B42 and each model and *ordinate* is root mean square error (RMSE) of each model. In **b**, the *abscissa* and *ordinate* denote PCC of DC1 and DC2 between TRMM 3B42 and each model, respectively. The result is based on the horizontal resolution of $2° \times 2°$

7 Impact of model resolution on simulation of the annual and diurnal cycles

How does the 20-km MRI/JMA AGCM perform compared to its lower-resolution (60/120/180-km) versions? For convenience of comparison, all models' outputs were interpolated onto the observational grid systems; thereby high-resolution features simulated in the high resolution models are lost. It should be kept in mind this disadvantage of the comparison method to high resolution model. Table 2 compares the performance of the same MRI/JMA model with four different resolutions in simulation of the annual and diurnal cycles. For all metrics fields except the monsoon domain, which is assessed by the threat score, the pattern correlation coefficient is used to measure the degree of success. Use of the normalized root mean square error as a measure yields the same conclusion. Obviously, for both the annual and diurnal cycles the performance of the

Table 2 Comparison of the pattern correlation coefficients between the observed and the simulated annual cycle and diurnal cycle calculated from the MRI/JMA GCM with four different resolutions

	Annual cycle				Diurnal cycle		
	AM	AC	MPI	MPD	DR	DC1	DC2
180 M/J	0.887	0.828	0.827	0.585	0.727	0.708	0.452
120 M/J	0.889	0.839	0.818	0.601	0.741	0.729	0.317
60 M/J	0.885	0.839	0.841	0.605	0.785	0.802	0.474
20 M/J	0.900	0.873	0.854	0.619	0.784	0.775	0.506

The metrics for annual cycle includes the fields of annual mean (AM), annual cycle (AC, the combined two annual cycle modes), global monsoon precipitation intensity (MPI) and domain (MPD). The metrics for diurnal cycle includes fields of diurnal range (DR) and the land–sea contrast mode (DC1) and transitional mode (DC2) of the diurnal cycle

20 km resolution is generally superior to its lower-resolution counterparts. However, for the annual cycle, the skills of the three lower-resolution models (60, 120, and 180 km) seem to be comparable. This result suggests that increase of resolution from high (60 km) to very high (20 km) yields significant improvement on the annual cycle simulation. On the other hand, increase of resolution from medium (180 km) to high (60 km) resolution does not show significant improvement. In the recent study, Kitoh and Kusunoki (2008) compared 180 and 20 km resolution of MRI model and found that 20-km mesh simulates better orographic rainfall (both location and intensity) and Baiu front (structure and rainfall). For the diurnal cycle, the 60 km version yields a comparable performance with the 20 km version, and they are better than the 180 and 120 km version.

8 Summary and discussion

To objectively represent the forced response of AGCMs to solar forcing on the diurnal and annual time scales in a compact and effective way, we used EOF analysis to guide the means of extracting the dominant modes of the annual and diurnal cycles. the advantages of using EOF modes as metrics is that it reduces dimensions of the metrics, preserves amplitudes of the variation (see Sect. 1 for details) and facilitates objective quantification of performance.

We have shown that the simulation of the annual cycle by the 20-km MRI/JMA model outperforms those simulated by 12 individual lower-resolution AMIP II AGCMs and has comparable performance to the 12 models' ensemble simulation in terms of the proposed metrics. The 20 km MRI/JMA model is also able to simulate realistic diurnal precipitation in terms of the diurnal range, spatial patterns of the two leading diurnal modes, and the complex

in-land and off-shore propagations in the coastal areas of the major monsoon regions. However, the model exaggerates the amplitudes of the diurnal range over the land. This caveat is likely attributed to the problems in the model's representation of the land surface process.

Comparison of four model versions with different resolution (180, 120, 60, and 20 km) reveals that the 20 km MRI/JMA model simulates the most realistic annual and diurnal cycles. However, the improved performance is not a linear function of the increasing resolution. For the annual cycle, increase of resolution from high (60 km) to very high (20 km) leads to marked improvement. On the other hand, for the diurnal cycle increase of resolution from medium (180 km) to high (60 km) resolution shows significant improvement. Note also that increase of resolution is one way but a more important way is to improve model physical parameterization. The use of high resolution model with better represented physical parameterizations that were adequately tuned to the increased resolution may lead an improved atmospheric response to the external forcing.

By comparing multi-models, we have a better chance to understand what model processes need improvement. One of the common weaknesses is the poor simulation of the monsoon domains in East Asia-western North Pacific and North America-eastern North Pacific. The deficiency suggests that the model physics cannot replicate correctly the annual migration of the subtropical highs and associated monsoon troughs in the western North Pacific and North Atlantic Oceans where the east-west continent–ocean thermal contrast is prominent. This is one of the major challenges in climate models' simulation of the annual cycle. Another common weakness is associated with the simulation of the equinoctial asymmetric mode of the annual cycle. The lack of reality in simulating spring-fall asymmetry is primarily related to the errors in response to seasonal variations of the equatorial cold tongues in the western hemisphere and to the seasonal transition of land–ocean thermal contrast in the eastern hemisphere. In terms of diurnal cycle, the land–atmosphere interaction process is suggested to be improved. Further examination will be left to future studies.

Although the 20 km MRI/JMA GCM can reproduce very realistic diurnal cycle, it simulates Madden-Julian oscillation poorly (Liu et al. 2009b). Therefore, the notion that correct representation of diurnal cycle in a GCM can lead to a better representation of longer time scale phenomena such as MJO (Slingo et al. 2003) seems to be not the case in the MRI/JMA model.

References

Adler RF, Huffman GJ, Chang A, Ferraro R, Xie P, Janowiak J, Rudolf B, Schneider U, Curtis S, Bolvin D, Gruber A, Susskind J, Arkin P, Nelkin E (2003) The version-2 global precipitation climatology project (GPCP) monthly precipitation analysis (1979-present). J Hydrometeor 4:1147–1167

Collier JC, Bowman KP (2004) Diurnal cycle of tropical precipitation in a general circulation model. J Geophys Res 109:D17105. doi:10.1029/2004JD004818

Dai A (2001) Global precipitation and thunderstorm frequencies. Part II: diurnal variations. J Clim 14:1112–1128

Dai A, Trenberth KE (2004) The diurnal cycle and its depiction in the Community Climate System Model. J Climate 17:930–951

Dai A, Wang J (1999) Diurnal and semidiurnal tides in global surface pressure fields. J Atmos Sci 56:3874–3891

Dai A, Lin X, Hsu K-L (2007) The frequency, intensity, and diurnal cycle of precipitation in surface and satellite observations over low- and mid-latitudes. Clim Dyn 29:727–744

Hastenrath SL (1968) Fourier analysis of Central American rainfall. Arch Meteor Geophys Bioklimatol B16:81–94

Horn LH, Bryson RA (1960) Harmonic analysis of the annual March of precipitation over the United States. Ann Assoc Am Geogr 50:157–171

Hsu C-P, Wallace JM (1976) The global distribution of the annual and semiannual cycles in precipitation. Mon Weather Rev 104:1093–1101

Huffman GJ, Adler RF, Bolvin DT, Gu GJ, Nelkin EJ, Bowman KP, Hong Y, Stocker EF, Wolff DB (2007) The TRMM multi-satellite precipitation analysis: quasi-global multi-year, combined-sensor precipitation estimates at fine scale. J Hydrometeor 8:38–55

JMA (2002) Outline of the operational numerical weather prediction at the Japan Meteorological Agnecy (Appendix to WMO numerical weather prediction progress report). Japan Meteorological Agency, 157 pp. http://www.jma.go.jp/JMA_HP/jma/jma-eng/jma-center/nwp/outline-nwp/index.htm

Johnson RH (2010) The diurnal cycle of monsoon convection. In: Chang CP (ed) The global monsoon system: research and forecast, World Scientific (in press)

Kang IS, Jin K, Wang B, Lau KM, Shukla J, Krishnamurthy V, Schubert SD, Wailser DE, Stern WF, Kitoh A, Meehl GA, Kanamitsu M, Galin VY, Satyan V, Park CK, Liu Y (2002) Intercomparison of the climatological variations of Asian summer monsoon precipitation simulated by 10 GCMs. Clim Dyn 19:383–395

Kikuchi K, Wang B (2008) Diurnal precipitation regimes in the global tropics. J Clim 21:2680–2696

Kim HJ, Wang B, Ding Q (2008) The global monsoon variability simulated by CMIP3 coupled climate models. J Clim 21:5271–5294

Kitoh A, Kusunoki S (2008) East Asian summer monsoon simulation by a 20-km mesh AGCM. Clim Dyn 31:389–401

Kusunoki S, Yoshimura J, Yoshimura H, Noda A, Oouchi K, Mizuta R (2006) Change of baiu rain band in global warming projection by an atmospheric general circulation model with a 20-km grid size. J Meteorol Soc Japan 84:581–611

Lee J-Y et al (2010) How are seasonal prediction skills related to models' performance on mean state and annual cycle? Clim Dyn. doi:10.1007/s00382-010-0857-4

Liu J, Wang B, Ding Q, Kuang X, Soon W, Zorita E (2009a) Centennial variations of the global monsoon precipitation in the last millennium: results from ECHO-G model. J Clim 22:2356–2371

Liu P, Kajikawa Y, Wang B, Kitoh A, Yasunari T, Li T, Annamalai H, Fu X, Kikuchi K, Mizuta R, Rajendran K, Waliser DE, Kim D (2009b) Tropical intraseasonal variability in the MRI-20km60L AGCM. J Clim 22:2006–2022

Mapes BE, Warner TT, Xu M (2003) Diurnal patterns of rainfall in northwestern South America. Part III: diurnal gravity waves and nocturnal convection offshore. Mon Weather Rev 131:830–844

Mizuta R, Oouchi K, Yoshimura H, Noda A, Katayama K, Yukimoto S, Hosaka M, Kusunoki S, Kawai H, Nakagawa M (2006) 20-km-Mesh global climate simulations using JMA-GSM model—mean climate states. J Meteor Soc Jpn 84:165–185

Mokhov II (1985) Method of amplitude-phase characteristics for analyzing climate dynamics. Meteor Gidrol 5:80–89

Murakami H, Matsumura T (2007) Development of an effective non-linear normal-mode initialization method for a high-resolution global model. J Meteorol Soc Jpn 85:187–208

Oouchi K, Yoshimura J, Yoshimura H, Mizuta R, Kusunoki S, Noda A (2006) Tropical cyclone climatology in a global-warming climate as simulated in a 20 km-mesh global atmospheric model: frequency and wind intensity analyses. J Meteorol Soc Jpn 84:259–276

Ploshay JJ, Lau NC (2010) Simulation of diurnal cycle in precipitation and circulation during boreal summer with a high-resolution GCM. J Clim (Submitted)

Rayner NA, Parker DE, Horton EB, Folland CK, Alexander LV, Rowell DP, Kent EC, Kaplan A (2003) Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. J Geophys Res 108(D14):4407. doi:10.1029/2002JD002670

Slingo J, Inness P, Neale R, Woolnough S, Yang GY (2003) Scale interactions on diurnal to seasonal timescales and their relevance to model systematic errors. Ann Geophys 46:139–155

Sperber KR, Palmer TN (1996) Interannual tropical rainfall variability in general circulation model simulations associated with the atmospheric model intercomparison project. J Clim 9:2727–2750

Tao S, Chen L (1987) A review of recent research on the East Asian summer monsoon in China. In: Chang CP, Krishnamurti TN (eds) Monsoon meteorology. Oxford University Press, Oxford, pp 60–92

Waliser D et al (2009) MJO simulation diagnostics. J Clim 22:3006–3030

Wallace JM (1975) Diurnal variations in precipitation and thunderstorm frequency over the conterminous United States. Mon Weather Rev 103:406–419

Wang B (1994) Climatic regimes of tropical convection and rainfall. J Clim 7:1109–1118

Wang B, Ding Q (2006) Changes in global monsoon precipitation over the past 56 years. Geophys Res Lett 33:L06711. doi:10.1029/2005GL025347

Wang B, Ding Q (2008) Global monsoon: dominant mode of annual variation in the tropics. Dyn Atmos Ocean 44:165–183

Wang B, Wu R, Lau KM (2001) Interannual variability of the Asian summer monsoon: contrasts between the Indian and the western North Pacific-east Asian monsoons. J Clim 14:4073–4090

Wang B, Bao Q, Hoskins B, Wu G, Liu Y (2008) Tibetan Plateau warming and precipitation change in East Asia. Geophys Res Lett 35:L14702. doi:10.1029/2008GL034330

Wilks DS (1995) Statistical method in Atmospheric Sciences, International geophysics Series, vol 59. Academic Press, London

Wu G-X, Liu Yimin (2003) Summertime quadruplet heating pattern in the subtropics and the associated atmospheric circulation. Geophys Res Lett 30(5):1201. doi:10.1029/2002GL016209,5_1-4

Wu G-X, Liu Y, Zhu X, Li W, Ren R, Duan A, Liang X (2009) Multi-scale forcing and the formation of subtropical desert and monsoon. Ann Geophys 27:3631–3644

Xie P, Arkin PA (1997) Global precipitation: a 17-year monthly analysis based on gauge observations, satellite estimates, and numerical model outputs. Bull Am Meteor Soc 78:2539–2558

Yang S, Smith EA (2006) Mechanisms for diurnal variability of global tropical rainfall observed from TRMM. J Clim 19: 5190–5226