Assessment of CMIP5 Model Simulations of the North American Monsoon System

KERRIE L. GEIL, YOLANDE L. SERRA, AND XUBIN ZENG

Department of Atmospheric Sciences, University of Arizona, Tucson, Arizona

(Manuscript received 18 January 2013, in final form 15 May 2013)

ABSTRACT

Precipitation, geopotential height, and wind fields from 21 models from phase 5 of the Coupled Model Intercomparison Project (CMIP5) are examined to determine how well this generation of general circulation models represents the North American monsoon system (NAMS). Results show no improvement since CMIP3 in the magnitude (root-mean-square error and bias) of the mean annual cycle of monthly precipitation over a core monsoon domain, but improvement in the phasing of the seasonal cycle in precipitation is notable. Monsoon onset is early for most models but is clearly visible in daily climatological precipitation, whereas monsoon retreat is highly variable and unclear in daily climatological precipitation. Models that best capture large-scale circulation patterns at a low level usually have realistic representations of the NAMS, but even the best models poorly represent monsoon retreat. Difficulty in reproducing monsoon retreat results from an inaccurate representation of gradients in low-level geopotential height across the larger region, which causes an unrealistic flux of low-level moisture from the tropics into the NAMS region that extends well into the post-monsoon season. Composites of the models with the best and worst representations of the NAMS indicate that adequate representation of the monsoon during the early to midseason can be achieved even with a large-scale circulation pattern bias, as long as the bias is spatially consistent over the larger region influencing monsoon development; in other words, as with monsoon retreat, it is the inaccuracy of the spatial gradients in geopotential height across the larger region that prevents some models from realistic representation of the early and midseason monsoon system.

1. Introduction

The evolution of the North American monsoon system (NAMS) can be described as having development, mature, and decay stages similar to but less intense than its larger Asian counterpart. During the development stage (May–June), the extratropical jet weakens and migrates to the north resulting in decreased frequency of synoptic-scale transient activity from the midlatitudes over northern Mexico and the southwestern United States (Higgins et al. 1997). A thermal surface low forms in the desert regions (Rowson and Colucci 1992) and a pronounced anticyclone at jet stream level develops over southwestern Mexico (Barlow et al. 1998), analogous to the Tibetan high over Asia (Tang and Reiter 1984). Mid- to upper-level flow shifts from westerly in May–June to easterly and southeasterly around the west side of the anticyclone by July (Douglas et al. 1993; Higgins et al. 1997). Low-level flow into the monsoon region is strongly influenced by the evolution of the North Atlantic subtropical high (NASH) and North Pacific subtropical high (NPSH). As the subtropical highs build and move northward, northwesterly flow from the NPSH is reduced over the northern Gulf of California and the westward extension of the NASH brings southerly flow into eastern Mexico and the U.S. Great Plains (Schmitz and Mullen 1996; Higgins et al. 1997; Barlow et al. 1998). Southerly winds flow over the Gulf of California (Badan-Dangon 1991; Douglas et al. 1993) and convective precipitation quickly spreads to the northwest along the western slopes and foothills of the Sierra Madre Occidental (SMO; Douglas et al. 1993). The mature stage (July–August) brings the precipitation maximum over the SMO and increased precipitation coincides with increased vertical moisture transport (Douglas et al. 1993; Schmitz and Mullen 1996). The decay stage (September–October) is conceptually the reverse of the development stage, but is
more gradual (Higgins et al. 1997; Barlow et al. 1998). The main source of monsoon moisture has been debated over recent decades, but it is most likely that low-level moisture is advected mainly from the Gulf of California and the eastern Pacific, while the Gulf of Mexico contributes to upper-level moisture (Schmitz and Mullen 1996; Berbery 2001; Higgins et al. 2003). NAMS region continental moisture sources are also important, as precipitation recycling contributes to monsoon season rainfall (Bosilovich et al. 2003; Dominguez et al. 2008).

Global and limited-area model simulations have been conducted in the past to evaluate the representation of the NAMS and the results show a wide range of model ability. Arritt et al. (2000) demonstrated that the Met Office (UKMO) HadCM2 global model could simulate generally realistic NAMS circulation and precipitation, whereas Yang et al. (2001) showed that the National Center for Atmospheric Research (NCAR) CCM3 global model was unable to simulate these NAMS features. Liang et al. (2008) found that only 1 of 17 CMIP3 global models was able to realistically reproduce the NAMS precipitation annual cycle, interannual variability in precipitation, and key circulation patterns such as the monsoon high and the westward extension of the NASH with the associated low-level southerly flow. Stensrud et al. (1995) reproduced monsoon mesoscale circulation and the general features of deep convection with the Fourth-generation Pennsylvania State University–NCAR Mesoscale Model (MM4) limited-area model, and Berbery (2001) showed that NCEP'sEta limited-area model could additionally reproduce the diurnal cycle of moisture flux. Gao et al. (2007) used the MM5 limited-area model to demonstrate improvement over global models in representing spatial and temporal precipitation patterns but found model deficiencies in representing the evolution of the diurnal cycle. Castro et al. (2007a) used the Regional Atmospheric Modeling System (RAMS) limited-area model driven with global reanalysis data and found that the model's enhanced representation of the surface boundary produced an acceptable diurnal cycle of summer precipitation in the monsoon region that was not captured by the driving reanalysis. A recent study by the same group using the Weather Research and Forecasting Model (WRF; Castro et al. 2012) showed the potential for limited-area models to improve seasonal NAMS forecasts. The use of higher resolution limited-area models that are able to capture the diurnal cycle of convection, as opposed to coarser general circulation models that do not have this capability, is stressed by Castro et al. (2007a,b, 2012) for drawing conclusions with respect to regional climate variability and prediction.

The present study is comprised of a series of analyses aimed at assessing how well the phase 5 of the Coupled Model Intercomparison Project (CMIP5) suite of coupled general circulation models (CGCMs) is able to represent the NAMS. The two analysis regions include a smaller core domain and a larger extended domain (Fig. 1). Our core domain is smaller but similar to that used by the North American Monsoon Experiment (NAME; Higgins et al. 2006) and related studies (e.g., Higgins and Gochis 2007; Gutzler et al. 2009), while our extended domain includes the larger NAMS region. The uniformity of the annual cycle of precipitation across all grid points within our core domain has been visually verified, as in Higgins et al. (1999), to ensure we have selected an area with a consistent monsoon precipitation signal.

Fig. 1. Core NAMS domain (23.875°–28.875°N, 108.875°–104.875°W) and extended NAMS domain (15.125°–34.875°N, 119.875°–90.125°W), shown over regional topography.
The observational data and CMIP5 model simulations used in this study are described in section 2. The annual precipitation cycle and comparison to CMIP3 is presented in section 3, while section 4 evaluates the spatial progression of the monsoon using spatial gridpoint correlation. Daily precipitation is utilized to calculate monsoon onset and retreat dates in section 5, composites of the best and worst performing models are discussed in section 6, and a summary is presented in section 7.

2. Observational data and model simulations

a. Observational data and reanalysis

Monthly precipitation observations are obtained from the recently developed National Oceanic and Atmospheric Administration (NOAA) 0.5° × 0.5° gridded precipitation dataset (P-NOAA) provided by Drs. Russ Vose and Ed Cook and is described by Castro et al. (2012). This dataset was created from station data and considers the dependence of precipitation on elevation, similar to the Parameter-Elevation Regressions on Independent Slopes Model (PRISM) dataset that covers only the United States (Daly et al. 1994). For our daily time resolution analysis, we use the Tropical Rainfall Measuring Mission (TRMM) 3B42v6 daily precipitation estimates (Huffman et al. 2007), which are provided on a 0.25° × 0.25° spatial grid. The dataset is created using several types of satellite measurements and also incorporates monthly station observations to improve accuracy. We have chosen to use the daily TRMM satellite dataset as opposed to a lower spatial resolution in situ daily dataset based on the importance of higher spatial resolution over variable terrain (Gochis et al. 2004). Also, in a study comparing different satellite-based precipitation estimates to 2004 NAME gauge data, Gochis et al. (2009) showed that the TRMM 3B42v6 product performs well over the monsoon region.

Monthly geopotential height and wind are provided by the European Centre for Medium-Range Weather Forecasts (ECMWF) Interim Re-Analysis (ERA-Interim; Dee et al. 2011) and are obtained from the Research Data Archive (RDA; http://rda.ucar.edu: dataset ds627.0). ERA-Interim is produced at spectral T255 horizontal resolution with 60 vertical layers and is provided at 6-hourly intervals on a 0.75° × 0.75° grid with 37 vertical pressure levels.

b. Coupled general circulation model simulations

The source of CGCM climate simulations is the CMIP5 multimodel ensemble archive, made available online by the Program for Climate Model Diagnosis and Intercomparison (PCMDI; http://pcmdi3.llnl.gov/esg/cet). The historical experiment is chosen for this analysis, which imposes changing atmospheric and land surface conditions consistent with past observations, including changes in atmospheric composition due to anthropogenic and volcanic influences, solar forcing, concentrations of short-lived species and aerosols from both natural and anthropogenic sources, and land use (Taylor et al. 2009). For details regarding CMIP5 experimental design, the reader is referred to Taylor et al. (2009, 2012). Table 1 provides information on the 21 CGCMs used for this study, which have atmospheric components ranging in horizontal grid resolution from 0.56° × 0.56° in longitude by latitude to 3.75° × 2.47° and oceanic horizontal grids ranging from 0.28° × 0.2° to 1.98° × 1.2° resolution (Gent et al. 2011; Volodin et al. 2010; http://data.giss.nasa.gov/modelE/ar5). We recognize that this range of resolution is still relatively coarse for the representation of detailed topography and the resultant small-scale atmospheric (e.g., convective) processes. Model composite statistics of high versus low horizontal and vertical resolutions (not shown) for each of the analyses in this study did not reveal major differences in model performance, implying that even the highest resolution model examined is still too coarse to capture small-scale topographically influenced processes.

All observations, reanalysis, and simulations are regrided to the TRMM 0.25° × 0.25° master grid using bilinear spatial interpolation, which facilitates direct comparison. The reference period for this study is 1979–2005 for all model simulations, reanalysis, and observations, except for TRMM daily precipitation, which only includes the years 1998–2010. Testing using the ERA-Interim precipitation indicates that the precipitation metrics presented in this study are insensitive to the difference in reference periods (between 1979–2005 and 1998–2010).

3. Annual cycle of precipitation

The annual cycle of precipitation within the core NAMS domain is characterized by relatively dry winter months followed by an early spring minimum, a sharp rise during late spring leading to a summertime peak, and a return to a secondary minimum in the fall (see Fig. 3). The wettest months are July, August, and September, when the bulk of the annual precipitation occurs, whereas the driest months are March, April, and May. Following the methods of Liang et al. (2008), Fig. 2 compares the modeled and observed precipitation annual cycle using three metrics. For each model, the root-mean-square (rms) error of monthly mean rainfall is shown in Fig. 2a, the percent bias in annual rainfall
<table>
<thead>
<tr>
<th>Model</th>
<th>Model expansion</th>
<th>Code</th>
<th>Modeling group</th>
<th>Country</th>
<th>Ensemble member</th>
<th>AGCM grid resolution</th>
<th>OGCM grid resolution</th>
</tr>
</thead>
<tbody>
<tr>
<td>BCC-CSM1.1</td>
<td>Beijing Climate Center, Climate System Model, version 1.1</td>
<td>BCC</td>
<td>Beijing Climate Center, China Meteorological Administration</td>
<td>China</td>
<td>r1i1p1</td>
<td>128 × 64L26</td>
<td>360 × 232L40</td>
</tr>
<tr>
<td>CanESM2</td>
<td>Second Generation Canadian Earth System Model</td>
<td>CAN</td>
<td>Canadian Centre for Climate Modelling and Analysis</td>
<td>Canada</td>
<td>r1i1p1</td>
<td>128 × 64L35</td>
<td>256 × 192L40</td>
</tr>
<tr>
<td>CCSM4</td>
<td>Community Climate System Model, version 4</td>
<td>CCS</td>
<td>National Center for Atmospheric Research</td>
<td>United States</td>
<td>r1i1p1</td>
<td>288 × 200L26</td>
<td>384 × 320L60</td>
</tr>
<tr>
<td>CNRM-CM5</td>
<td>Centre National de Recherches Météorologiques Coupled Global Climate Model, version 5</td>
<td>CNR</td>
<td>Centre National de Recherches Météorologiques/Centre Européen de Recherche et Formation Avancées en Calcul Scientifique</td>
<td>France</td>
<td>r1i1p1</td>
<td>256 × 128L31</td>
<td>362 × 292L42</td>
</tr>
<tr>
<td>CSIRO Mk3.6.0</td>
<td>Commonwealth Scientific and Industrial Research Organisation Mark, version 3.6.0</td>
<td>CSI</td>
<td>Commonwealth Scientific and Industrial Research Organization, Queensland Climate Change Centre of Excellence</td>
<td>Australia</td>
<td>r1i1p1</td>
<td>192 × 96L18</td>
<td>192 × 189L31</td>
</tr>
<tr>
<td>GFDL CM3</td>
<td>Geophysical Fluid Dynamics Laboratory Climate Model, version 3</td>
<td>GF3</td>
<td>NOAA/Geophysical Fluid Dynamics Laboratory</td>
<td>United States</td>
<td>r1i1p1</td>
<td>144 × 90L14</td>
<td>360 × 200L50</td>
</tr>
<tr>
<td>GFDL-ESM2G</td>
<td>Geophysical Fluid Dynamics Laboratory Earth System Model with Generalized Ocean Layer Dynamics (GOLD) component</td>
<td>GFG</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GFDL-ESM2M</td>
<td>Geophysical Fluid Dynamics Laboratory Earth System Model with Modular Ocean Model 4 (MOM4) component</td>
<td>GFM</td>
<td></td>
<td></td>
<td>r1i1p1</td>
<td>144 × 90L24</td>
<td>360 × 210L63</td>
</tr>
<tr>
<td>GISS-E2-R</td>
<td>Goddard Institute for Space Studies Model E, coupled with the Russell ocean model</td>
<td>GIS</td>
<td>National Aeronautics and Space Administration (NASA) Goddard Institute for Space Studies</td>
<td>United States</td>
<td>r6i1p1</td>
<td>144 × 90L4</td>
<td>288 × 180L32</td>
</tr>
<tr>
<td>HadCM3</td>
<td>Hadley Centre Coupled Model, version 3</td>
<td>HA3</td>
<td>Met Office Hadley Centre</td>
<td>United Kingdom</td>
<td>r1i1p1</td>
<td>96 × 73L19</td>
<td>288 × 144L20</td>
</tr>
<tr>
<td>HadGEM2-CC</td>
<td>Hadley Centre Global Environment Model, version 2– Carbon Cycle</td>
<td>HGC</td>
<td></td>
<td></td>
<td>r1i1p1</td>
<td>192 × 145L60</td>
<td>360 × 216L40</td>
</tr>
<tr>
<td>HadGEM2-ES</td>
<td>Hadley Centre Global Environment Model, version 2–Earth System</td>
<td>HGE</td>
<td></td>
<td></td>
<td>r1i1p1</td>
<td>192 × 145L38</td>
<td>360 × 216L40</td>
</tr>
<tr>
<td>INM-CM4.0</td>
<td>Institute of Numerical Mathematics Coupled Model, version 4.0</td>
<td>INM</td>
<td>Institute of Numerical Mathematics</td>
<td>Russia</td>
<td>r1i1p1</td>
<td>180 × 120L21</td>
<td>360 × 340L40</td>
</tr>
<tr>
<td>IPSL-CM5A-LR</td>
<td>L’Institut Pierre-Simon Laplace Coupled Model, version 5A, coupled with Nucleus for European Modelling of the Ocean (NEMO), low resolution</td>
<td>IPL</td>
<td>L’Institut Pierre-Simon Laplace</td>
<td>France</td>
<td>r1i1p1</td>
<td>96 × 96L39</td>
<td>182 × 149L31</td>
</tr>
<tr>
<td>IPSL-CM5A-MR</td>
<td>L’Institut Pierre-Simon Laplace Coupled Model, version 5A, coupled with NEMO, mid resolution</td>
<td>IPM</td>
<td></td>
<td></td>
<td>r1i1p1</td>
<td>143 × 144L39</td>
<td>182 × 149L31</td>
</tr>
</tbody>
</table>
based on the mean monthly climatological observed rainfall value (1.66 mm day$^{-1}$) is shown in Fig. 2b, and the phase lag in number of months is shown in Fig. 2c. All three metrics are calculated for the core NAMS domain using 12 monthly climatological precipitation values. A correlation is computed between the model annual cycle and the observed cycle at each (monthly interval) time lag and the phase lag is defined as the time lag with the highest correlation. The range of rms error in annual rainfall totals is 0.76–2.74 mm day$^{-1}$ and the average model error is 1.47 mm day$^{-1}$. The large majority of models are biased wet, with an average bias of 51.3% and a range from −42% to 136%. Although there is no bias calculation in Liang et al. (2008) for comparison, similarity between the range of rms error (0.46–2.23 mm day$^{-1}$) in their study of CMIP3 models and that of the CMIP5 models in this analysis indicates that there has been no improvement in the magnitude of the simulated annual cycle of monthly precipitation and, in fact, the lowest and highest rms values have increased slightly since the previous generation of CGCMs. On the other hand, there does seem to be improvement in the timing of seasonal precipitation shifts, with 13 out of 21 (62%) CMIP5 models having a phase lag of zero months as compared to 6 out of 17 (35%) CMIP3 models in Liang et al. (2008). Figure 3 shows the precipitation annual cycle for all models separated into three groups by phase lag value. Small, moderate, and large phase lag models are defined as those with zero, 1-month, and greater than 1-month phase lags, respectively. The only model that captures all characteristic features of the annual cycle with the proper timing is the Met Office Hadley Centre (MOHC) HGE model; however, this model is too wet for 11 out of 12 months. Of the small and moderate phase lag models, a common problem is the difficulty in ending the monsoon season, as reflected by the insufficient fall minimum and high precipitation in the fall and winter seasons. This problem is also seen in the multimodel mean.

4. Seasonal spatial correlation of monsoon variables

To verify the appropriate spatial progression of NAMS onset and retreat, we assess the spatial pattern of monsoon variables within the extended NAMS analysis region. This is accomplished using a simple point-to-point spatial correlation of climatological monthly model values of precipitation, geopotential height at 500 and 850 hPa, and meridional and zonal wind at 500 and 850 hPa, to the corresponding TRMM satellite data or ERA-Interim data. Pressure level correlation indices at 500 and 850 hPa are created by averaging the

<table>
<thead>
<tr>
<th>Model</th>
<th>Country</th>
</tr>
</thead>
<tbody>
<tr>
<td>MIROC-ESM</td>
<td>Japan</td>
</tr>
<tr>
<td>MIROC-H</td>
<td>Japan</td>
</tr>
<tr>
<td>MIROC2-H</td>
<td>Japan</td>
</tr>
<tr>
<td>MIROC3</td>
<td>Japan</td>
</tr>
<tr>
<td>MPI-ESM-LR</td>
<td>Germany</td>
</tr>
<tr>
<td>MRI-CGCM3</td>
<td>Japan</td>
</tr>
<tr>
<td>NorESM1-M</td>
<td>Norway</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Code</th>
<th>Country</th>
<th>Modeling group</th>
</tr>
</thead>
<tbody>
<tr>
<td>MIE</td>
<td>Japan Agency for Marine-Earth Science and Technology, Atmosphere and Ocean Research Institute (The University of Tokyo), and National Institute for Environmental Studies</td>
<td>Japan</td>
</tr>
<tr>
<td>MI4</td>
<td>MIE</td>
<td>Japan</td>
</tr>
<tr>
<td>M5</td>
<td>MIE</td>
<td>Japan</td>
</tr>
<tr>
<td>MI5</td>
<td>MIE</td>
<td>Japan</td>
</tr>
<tr>
<td>MPI</td>
<td>MPI Max Planck Institute for Meteorology</td>
<td>Germany</td>
</tr>
<tr>
<td>MRI</td>
<td>MRI Meteorological Research Institute</td>
<td>Japan</td>
</tr>
<tr>
<td>NOR</td>
<td>NOR Norwegian Climate Centre</td>
<td>Norway</td>
</tr>
</tbody>
</table>

Table 1. (Continued)
corresponding correlations of geopotential height and winds. It is important to note that Poisson grid filling (using standard routines included with the NCAR Command Language V6.0.0; NCAR 2012) is used to interpolate the 850-hPa model output below ground to obtain continuous fields at this pressure level. Correlations are calculated for each month during the May through October monsoon season and the average correlations of the May–June (MJ) early season, the July–August (JA) midseason, the September–October (SO) late season, and the full May–October (MJJASO) season are used to rank the models.

Table 2 shows the results of the correlation analysis for the five highest and lowest ranked models. The range of highest to lowest ranked correlation is usually smallest for the 500-hPa index (e.g., seasonal range of \( r = 0.64–0.91 \)) and largest for precipitation (e.g., seasonal range of \( r = 0.23–0.79 \)). The IPL model even has a negative precipitation correlation \( (r = -0.05) \) during the decay stage. The highest correlations occur earlier in the monsoon season, while the lowest correlations occur later in the season. Seasonally, 16 out of 21 models perform better at 500 hPa than at 850 hPa (not explicitly shown in Table 2), partly because of less small-scale variability and hence fewer spatial degrees of freedom at 500 hPa. It is interesting that three of the five models that do not perform better at 500 hPa than at 850 hPa (CNR, HGC, and HGE) consistently have the highest precipitation correlations. Relatively good or bad correlation at the 500-hPa level is not predictive of 850-hPa level or precipitation correlation; however, better correlation at the 850-hPa level usually corresponds to better correlation of precipitation during the monsoon season.

These results and the results of the annual cycle of precipitation analysis are used to choose the models appropriate for the daily time resolution onset and retreat analysis that follows. Models with a large phase lag (>1 month) and models in the bottom quintile of precipitation or 850-hPa index seasonal correlations have the poorest representations of the NAMS and

![Fig. 2. Annual precipitation cycle (a) rms error (mm day$^{-1}$), (b) bias (%), and (c) phase lag yielding maximum correlation (months) for all models with respect to monthly P-NOAA observations over the core NAMS domain. Multimodel median and mean values are also shown in each panel. All metrics are calculated using 12 monthly climatological values for the period 1979–2005. Phase lag of +1 means the highest correlation of all 12 months is between monthly observations ordered January through December and monthly model precipitation ordered February through January.](image-url)
are therefore eliminated from the daily analysis. This eliminates seven models: namely, the GIS, INM, IPL, IPM, MIE, MI4, and NOR models.

Note that neither the correlation nor the correlation difference between models needs to be statistically significant for the ranking in Table 2. Actually, it is not easy to address the statistical significance because of spatial autocorrelation in both the meridional and zonal directions. To reduce the spatial autocorrelation, we have thinned the data at 10, 20, 30, and 40 gridbox increments (i.e., with 2.5°, 5°, 7.5°, and 10° distances between adjacent data points). All correlations over these coarse grids for the top five ranked models in each category and for all other models are found to be statistically

FIG. 3. Annual precipitation cycle over the core NAMS domain for all models grouped by (a) small phase lag (lag = 0 months), (b) moderate phase lag (lag = 1 month), and (c) large phase lag (lag > 1 month). The multimodel mean (gray dashed line) for each category is shown in (a)–(c). (d) The all-model mean (dashed line) and spread (shading). Colors represent different modeling centers and solid vs dashed lines of the same color differentiate models from a common center.
significant at the 0.001 confidence level based on a Student’s t test. Other methods of computing statistical significance with a more robust technique for accounting for spatial autocorrelation were not pursued because of the apparent small value added for the type of analyses presented in this study.

5. Monsoon onset and retreat

Interannual variability of the NAMS impacts local ecosystems, agriculture, and the general public; therefore, there is considerable interest in determining the NAMS annual onset and retreat dates (e.g., Higgins et al. 1997; Zeng and Lu 2004; Ellis et al. 2004; Arias et al. 2012). Identification of model annual onset and retreat dates using daily precipitation provides a higher time resolution evaluation of monsoon season phasing than the monthly phase lag analysis in section 3. We employ the method of Higgins et al. (1997) for identifying the onset of monsoon rains. A precipitation index (PI) is created by time averaging TRMM daily precipitation observations at each grid point in the core domain and subsequently averaging all grid points together. The PI time series, shown in Fig. 4, is used to set the precipitation magnitude and duration threshold criteria for defining monsoon onset at $1.3 \text{ mm day}^{-1}$ and 3 days, respectively. These criteria are tested using individual years of daily TRMM observations to ensure they yield a set of reasonable onset dates. For each year, monsoon onset occurs the first time the threshold criteria are satisfied after 1 May. Since precipitation is much more variable following than preceding the monsoon season, the daily precipitation time series for each year of data are examined to set the monsoon retreat criteria at $1.3 \text{ mm day}^{-1}$ and 7 days. For each year, monsoon retreat occurs when the threshold criteria are first satisfied after 1 September. Table 3 shows the observed monsoon onset and retreat dates for each year based on these criteria. The average calendar date of onset for the core domain is 18 June, which agrees with...
the results of Higgins et al. (1999) (see their Fig. 12). The median date of onset is 19 June and the median date of retreat is 28 September.

Table 4 shows model median onset and retreat dates together with the corresponding lag from observations (in number of days) using the observation-defined threshold criteria, herein referred to as the absolute criteria. The CSI model onset is closest to the observed median onset date by a wide margin with a +3-day onset lag (Table 4) and also has a reasonable onset standard deviation (Table 4) and visual clarity of the onset in the daily climatology (Fig. 5). Most models have a standard deviation of onset dates that is within ±3 days of that of the observations (15.2 days) and daily climatological precipitation that displays a clear monsoon onset, although generally onset is early. These properties are reflected in the multimodel average standard deviation of onset (14.9 days; Table 4); daily climatological precipitation that displays a clear monsoon onset, although generally onset is early. The HGE model retreat is closest to the observed median retreat date with a +4-day retreat lag, but the standard deviation of retreat dates almost doubles that of the observational data and retreat is not as visually clear in the daily climatology as it is in the observations. The standard deviation of retreat dates for most models is amplified by at least a factor of 2 using the absolute criteria, as compared to the 8.3-day standard deviation of the observations. Most models also extend the monsoon past the observed retreat date. Additionally, most models lack a visually clear retreat in the daily climatological precipitation, with the end of the monsoon season blending into excessive fall and winter precipitation. These properties are also seen in the multimodel average standard deviation of retreat (18.1 days; Table 4); daily climatological precipitation (Fig. 5); and late 6 October retreat date (Table 4), which is 8 days later than the observed date.

The use of absolute criteria may not be suitable for many models due to errors in annual precipitation cycle amplitude; therefore, we have additionally computed model-relative threshold criteria as an alternate method of defining monsoon onset and retreat. Our method is constructed around that of Zeng and Lu (2004), who used a normalized precipitable water index to define globally unified summer monsoon onset and retreat dates. We use precipitation instead of precipitable water to define a normalized precipitation index (NPI) for the TRMM observations,

\[
NPI = \frac{P_{\text{threshold}} - P_{\text{min}}}{P_{\text{max}} - P_{\text{min}}},
\]

where \(P_{\text{threshold}}\) is the daily precipitation magnitude threshold of 1.3 mm day\(^{-1}\) and \(P_{\text{min}}\) and \(P_{\text{max}}\) are the 13-yr climatological values of the area-averaged minimum and maximum monthly TRMM precipitation in

![Fig. 4. Climatological (1998–2010) area-averaged daily (thin line) TRMM precipitation index and 5-day running mean (heavy line) in millimeters per day for the core NAMS domain.](image-url)
because the observed median onset retreat is 4 days early on 24 September and the standard deviation of retreat (Table 5) is much larger than that of the observational data. The model-relative criteria act to adjust the multimodel onset and retreat dates 10 and 4 days closer to the observed dates, respectively.

Interestingly, high monthly and seasonal precipitation correlations do not necessarily imply monsoon onset and retreat dates that are close to observations. The best example of this is demonstrated with the results from the CNR model, which has the highest 850-hPa correlations and ranks in the top four models for precipitation correlations throughout the entire monsoon season (Table 2) but has a median absolute onset date that is 39 days earlier (Table 4) and a model-relative onset that is 29 days earlier (Table 5) than the observed date. A visual inspection of the daily precipitation field for each year (not shown) reveals consistent small precipitation events in early May that satisfy the definition of monsoon onset even though a clearer onset signal is visible later in the daily climatological precipitation. Adjustment of the monsoon onset definition was not able to remedy the problem. The disparity between the daily and monthly resolution analysis results for the CNR model is a good example of how important insight can be gained with higher time resolution model output and how it is possible for lower time resolution model output to be misleading. Still, the daily precipitation analysis could also be deceptive on its own, as an investigation of the six models deemed inappropriate for the daily analysis (not shown) reveals consistent small precipitation events in May that satisfy the definition of monsoon onset as early as 1 May and 1 September, respectively. Model median onset and retreat dates with the corresponding lag from observations (in number of days) using the model-relative threshold criteria are shown in Table 5, and Fig. 5 depicts the difference between absolute and model-relative thresholds, shown over the daily climatological precipitation for each model and for the multimodel mean. The HGE and BCC models have onset dates that are closest to the observed median onset with a 2-day lag (Table 5), reasonable onset standard deviations (Table 5), and visual clarity of the onset in the daily climatologies (Fig. 5). The CSI model-relative retreat is the closest to the observed median retreat date with a +2-day lag, but the standard deviation of the retreat dates is about 1.5 times that of the observational data. Furthermore, although a retreat is visible in the daily climatology of CSI, the fall minimum is weak and quickly leads to exaggerated fall and winter precipitation. The multimodel average median onset is 12 days early on 7 June and the standard deviation of onset (Table 5) is similar to that of the observational data, while the average median onset retreat is 4 days early on 24 September and the standard deviation of retreat (Table 5) is much larger than that of the observational data. The model-relative criteria act to adjust the multimodel onset and retreat dates 10 and 4 days closer to the observed dates, respectively.

6. Composites

Our final analysis visually elucidates the previously demonstrated wide range in model ability to reproduce key spatial and temporal features of the NAMS. Composites of the monthly fields of precipitation, 850-hPa geopotential height, and 850-hPa winds are constructed to illustrate the major differences between models with the best and worst representations of the NAMS based on the measures discussed below.

Models with large phase error at monthly time resolution (GIS, INM, and MIE) are not considered for compositing because of the unrecognizable or very poor representation of the annual cycle of NAMS region precipitation. Also, the monthly fields are examined for each individual model to rule out the presence of any spurious model output. During this process, anomalous large-scale waves were discovered in the geopotential height field from the CSI model and consequently this
Fig. 5. Daily precipitation climatology over the core NAMS domain for (top left) observations (1998–2010), (top right) multimodel mean (1979–2005), and 14 models (1979–2005). Short dashed lines depict the absolute threshold (1.3 mm day$^{-1}$), and long dashed lines depict model-relative thresholds for defining yearly monsoon onset and retreat dates.
model is not considered for compositing either. The best models must rank within the top five of the remaining models for seasonal precipitation and 850-hPa correlations and must have a daily absolute onset lag of less than 30 days, resulting in a composite of the HA3, HGE, and HGC models. The worst models must rank within the bottom five of the remaining models for seasonal precipitation and 850-hPa correlations, resulting in a composite of the IPL, IPM, and NOR models. We also note that all models satisfying the requirements for the best composite have been previously categorized as having small phase error at monthly time resolution, whereas all models satisfying the requirements for the worst composite have been previously categorized as having moderate phase error at monthly time resolution (see Figs. 3a,b).

Figure 6 shows precipitation, 850-hPa geopotential height, and 850-hPa winds for the observational data, the best composite, and the worst composite for the development stage month of June, the mature stage month of August, and the decay stage month of October. Circulation features important to the proper representation of the NAMS are clearly seen in the observations (Fig. 6, top). During the development and mature stages, the NASH, as seen in the geopotential height field, strengthens and extends toward the west. The westward extension of the NASH brings easterly flow and moisture into the eastern portion of the study region from the Gulf of Mexico and the northward movement of the NPSH allows for moisture transport into the western portion of the region from the eastern Pacific and the Gulf of California, enhancing convergence and precipitation over land. During this period, a clear northward progression of precipitation along the Sierra Madre Occidental from southern Mexico is visible. The observations also show (Fig. 6, top) cross-equatorial flow, southwesternly winds in the ITCZ that create regions of enhanced convergence, and generally weak winds over the eastern North Pacific. Flow along the western coast of Mexico brings moisture from the ITCZ toward the Mexican mainland, whereas this flow weakens and turns offshore during the decay stage, which effectively severs the connection between the monsoon region and ITCZ moisture. Monsoon decay also brings the eastward retreat of the NASH; the weakening of the NASH, NPSH, and ITCZ; and the southeasterly retreat of precipitation along the Sierra Madre due to reduced convergence and moisture availability.

The composite of the three best models (Fig. 6, middle) illustrates most of the circulation features seen in the observations during the development and mature stages and demonstrates that current CGCM’s are capable of realistically representing the NAMS, even at horizontal resolutions coarser than 1°. That being said, some significant issues are also apparent. For example, the 850-hPa geopotential heights are biased low, precipitation is biased high, and the connection to tropical moisture extends well into the fall, resulting in a poor representation of monsoon retreat.

The composite of the three worst models (Fig. 6, bottom) fails to adequately reproduce most of the circulation features seen in the observations. Although the composite does show strengthening and weakening of the NASH and NPSH, the NASH is overextended toward the west and the NPSH is far too weak, resulting in an anomalously strong north–south gradient of geopotential height over the eastern North Pacific. The gradient in the tropics produces strong zonal winds within and to the north of the ITCZ, preventing the proper development of the ITCZ and cutting off ITCZ moisture to the monsoon region. Reduced convergence over Latin America results in extremely poor representation of the tropical wet season over this area and prevents the observed progression of monsoon precipitation from the southeast toward the northwest along the mountainous west coast of Mexico. During the decay stage, the retreat of the subtropical highs shifts the unrealistic tropical gradient in geopotential height toward the north, which steers winds and moisture over the eastern North Pacific to the northwest along the Mexican coast, resulting in exaggerated precipitation over the core NAMS domain during the fall.

7. Summary

A total of 21 CMIP5 coupled general circulation models are examined to determine how well the models
Fig. 6. Monthly climatological precipitation (color shading), 850-mb geopotential height (contours), and 850-mb winds (vectors) for (top) the data, (middle) the best model composite, and (bottom) the worst model composite for (left) the development stage month of June, (center) the mature stage month of August, and (right) the decay stage month of October.
represent the North American monsoon system. Analyses include a comparison of the annual cycle of area-averaged monthly precipitation over a core monsoon domain to CMIP3 results of Liang et al. (2008), a spatial correlation of monthly model precipitation and pressure level variables to data, an assessment of monsoon onset and retreat dates as determined from daily precipitation, and a model composite analysis of the best versus the worst representations of the NAMS.

There has been no improvement in the magnitude (rms error and bias) of the mean annual cycle of monthly precipitation over the core NAMS region since CMIP3, but the timing of seasonal changes in precipitation has improved with 27% more CMIP5 than CMIP3 models having zero phase lag. Despite this, a few models do not have a recognizable monsoon signal at all. Also, the multimodel mean annual cycle is biased wet and exhibits the common problem of late monsoon termination.

Monsoon season correlations of monthly model output to observational data establish that most models have the highest correlation at the 500-hPa level and the lowest correlations for precipitation; however, relatively good or bad performance at the 500-hPa level is not predictive of 850-hPa level or precipitation performance.

The multimodel mean onset and retreat dates are 23 days early and 9 days late, respectively, using the absolute criteria for defining monsoon onset and retreat. Yearly model onset variability is comparable to that of the observational data, but yearly model retreat variability is much greater than what is seen in the observations. On average, the model-relative onset and retreat dates are an improvement over the absolute dates because of the prevailing wet bias in model precipitation.

The 850-hPa composite of best models reproduces the development and mature stages of the NAMS, but the composite of worst models fails to adequately illustrate most of the precipitation and circulation features seen in the observations. The large-scale circulation pattern bias seen in the best model composite is spatially consistent over the larger region influencing monsoon development and thus still allows for a successful representation of the NAMS during the development and mature stages. In contrast, the spatial inconsistency of large-scale circulation pattern bias in the worst models prevents a realistic representation of the NAMS during the same period. Neither the composite of best models nor the composite of worst models realistically captures the retreat of the NAMS because of an extended connection to tropical moisture that causes excessive fall and winter precipitation. Models that best capture the relevant large-scale circulation patterns at low levels usually have a realistic representation of the NAMS, while performance at midlevels does not appear to be a major factor.

We have shown the importance of large-scale features to the representation of the NAMS in a suite of CMIP5 models that are still relatively coarse for capturing the detailed regional topography and resultant small-scale NAMS processes. Model composites of high versus low horizontal and vertical resolutions (not shown) did not reveal major differences in model performance with respect to NAMS representation, implying that even the highest resolution model examined is still too coarse to capture small-scale topographically influenced processes. There is room for improvement in the representation of the NAMS for many models by way of more accurate representation of low-level large-scale circulation features, but improvement in the representation of the NAMS in the best models is likely limited until increased model resolution allows for the capture of small-scale NAMS processes. Finally, we encourage subsequent CMIP collaborations to output more daily model fields, which were not available for all models and variables examined in this study, thereby limiting most analyses to monthly time resolution.

Acknowledgments. This work was supported by the National Oceanic and Atmospheric Administration MAPP Grant GC10-398. We would like to thank Drs. Russ Vose and Ed Cook for providing the P-NOAA dataset. We also thank Dave Gochis and two anonymous reviewers for their constructive suggestions. We acknowledge the Computational and Information Systems Laboratory (CISL) at the National Center for Atmospheric Research (NCAR) for maintaining the RDA and the National Science Foundation for their sponsorship of NCAR. We additionally acknowledge the World Climate Research Programme’s (WCRP) Working Group on Coupled Modelling (WGCM), which is responsible for CMIP, and we thank the climate modeling groups (listed in Table 1 of this paper) for producing and making available their model output. For CMIP, the U.S. Department of Energy’s Program for Climate Model Diagnosis and Intercomparison provides coordinating support and led development of software infrastructure in partnership with the Global Organization for Earth System Science Portals.

REFERENCES


