Can Regional Climate Models Represent the Indian Monsoon?

PHILIPPE LUCAS-PICHER,* JENS H. CHRISTENSEN,† FAHAD SAEED,† PANKAJ KUMAR,‡ SHAKEEL ASHARAF,§ BODO AHRENS,‖ ANDREW J. WILTSHIRE,‖ AND STEFAN HAGEMANN#

* Centre National de Recherches Météorologiques (CNRM-GAME), Météo-France, Toulouse, France, and Danish Meteorological Institute, Copenhagen, Denmark
† Danish Meteorological Institute, Copenhagen, Denmark
§ Max Planck Institute for Meteorology, Hamburg, Germany
‖ Institute for Atmospheric and Environmental Sciences, Goethe University Frankfurt, Frankfurt am Main, Germany
# Met Office Hadley Centre for Climate Prediction and Research, Exeter, United Kingdom

(Manuscript received 11 June 2010, in final form 28 February 2011)

ABSTRACT

The ability of four regional climate models (RCMs) to represent the Indian monsoon was verified in a consistent framework for the period 1981–2000 using the 45-yr European Centre for Medium-Range Weather Forecasts (ECMWF) Re-Analysis (ERA-40) as lateral boundary forcing data. During the monsoon period, the RCMs are able to capture the spatial distribution of precipitation with a maximum over the central and west coast of India, but with important biases at the regional scale on the east coast of India in Bangladesh and Myanmar. Most models are too warm in the north of India compared to the observations. This has an impact on the simulated mean sea level pressure from the RCMs, being in general too low compared to ERA-40. Those biases perturb the land–sea temperature and pressure contrasts that drive the monsoon dynamics and, as a consequence, lead to an overestimation of wind speed, especially over the sea. The timing of the monsoon onset of the RCMs is in good agreement with the one obtained from observationally based gridded datasets, while the monsoon withdrawal is less well simulated. A Hovmöller diagram representation of the mean annual cycle of precipitation reveals that the meridional motion of the precipitation simulated by the RCMs is comparable to the one observed, but the precipitation amounts and the regional distribution differ substantially between the four RCMs. In summary, the spread at the regional scale between the RCMs indicates that important feedbacks and processes are poorly, or not, taken into account in the state-of-the-art regional climate models.

1. Introduction

South Asian summer is dominated by the Indian monsoon, which spans four months from June to September and provides the major input of water for a large fraction of the world total population living in India, Bangladesh, Myanmar, and Nepal (Goswami 2005). Most global climate models (GCMs) simulate the general migration of the seasonal tropical rain (e.g., Christensen et al. 2007). However, the observed maximum rainfalls during the monsoon season along the west coast of India, the north Bay of Bengal, and northeast India are poorly simulated by most GCMs (Christensen et al. 2007; Kripalani et al. 2007). This is likely caused by the coarse resolutions of the GCMs, which are not able to correctly represent the regional forcings such as the steep topography of the Himalayas and the Western Ghats (Rupa Kumar et al. 2006).

The computer power currently available constrains GCMs to perform long global climate simulations on a regular grid at a horizontal resolution of ~200 km. Nevertheless, Rajendran and Kitoh (2008) performed two 10-yr time slices with a global general circulation model at 20-km horizontal resolution, but this design requires very large computational resources that only
very few climate modeling groups can afford. Their use of high horizontal resolution significantly improved the spatial distribution of the precipitation over South Asia, especially on the edge of the steep mountains. For more than 20 years, an alternative approach to high-resolution climate simulations has consisted of the use of a regional climate model (RCM) to dynamically downscale a GCM simulation or a reanalysis (e.g., Giorgi 2006). By using a domain covering a certain region of the globe, the RCMs are able to efficiently perform a climate simulation at a horizontal resolution of typically 50 km or less. While being controlled by the large-scale boundary conditions taken from a GCM or a reanalysis, the RCMs take advantage of their higher resolution to improve the description of the regional forcings, such as the mountain orography, land cover, and the land–sea contrasts.

Many studies have been carried out to verify the ability of RCMs to simulate the Indian monsoon. Initial studies performed short simulations using GCMs as lateral boundary conditions (LBCs), focusing on a few months or years to verify the validity of the approach (Bhaskaran et al. 1996; Jacob and Podzun 1997; Ji and Vernekar 1997; Bhaskaran et al. 1998; Vernekar and Ji 1999). Then, several studies used RCMs driven with reanalysis data at their boundaries (Park and Hong 2004; Venkata Ratnam and Krishna Kumar 2005; Venkata Ratnam and Cox 2006; Dash et al. 2006; Saeed et al. 2009; Dobler and Ahrens 2010; Saeed et al. 2011). In their sensitivity study, Dash et al. (2006) found that the amount and distribution of rainfall simulated by RegCM3 using the Grell parameterization is closer to the observations than using the Kuo parameterization of convection. Saeed et al. (2009) showed that the implementation of an irrigation scheme reduces the warm bias of the Max Planck Institute Regional Model (REMO) for the north of India and improves the precipitation distribution. Recently, Dobler and Ahrens (2010) performed RCM simulations with the Consortium for Small-Scale Modelling (COSMO) Climate Local Model (CLM) RCM over South Asia, using the ECHAM5–Max Planck Institute Ocean Model (MPIOM) GCM and the 45-yr European Centre for Medium-Range Weather Forecasts (ECMWF) Re-Analysis (ERA-40; Uppala et al. 2005) as driving fields. They found that the mean and temporal variability of most monsoon indices are improved with the CLM RCM compared to its driving field with ECHAM5–MPIOM, but not with ERA-40.

Although the above-mentioned studies are useful to determine the behavior of the RCMs, none of them analyzed the performance of multiple RCMs using a consistent framework over a long period of time with reanalysis as LBCs. It has been mentioned in earlier studies that multiyear simulations must be incorporated to provide meaningful climate statistics and to identify significant model errors (Fu et al. 2005). Moreover, the RCM ability to maintain the realistic large-scale circulation by the use of reanalysis LBCs allows the isolation of the regional feedbacks (Park and Hong 2004). The Regional Climate Model Intercomparison Project (RMIP) seeks to improve the RCM simulations of the East Asian climate by evaluating its strengths and weaknesses in a common framework (Fu et al. 2005). Analyzing simulations from this project, Feng and Fu (2006) showed that RCMs have the capacity to reproduce the basic spatial patterns of precipitation, but there are distinctions in the location and the intensity. Unfortunately, in the latter study, the Indian region is treated as a whole and located very close to the lateral boundary.

Given the importance of the Indian monsoon and its associate precipitation for the water availability over the Indian subcontinent, three European Union (EU) projects [Water and Global Change (WATCH), Twinning European and South Asian River Basins (BRAHMATWINN), and Himalayan Glacier Retreat and Changing Monsoon Pattern (HighNoon)] focused recently on the water cycle of the Indian monsoon and on potential future changes. The RCMs play an important role in these projects to downscale the coarse global climate simulations from GCMs and to provide regional climate information that takes into account regional forcings, feedbacks, and processes. Accordingly, there is a strong demand to evaluate the potential of RCMs to simulate the regional distribution of water over South Asia. Here, we meet this need by verifying the ability of four RCMs to simulate the Indian monsoon in a common multiyear framework using reanalysis LBCs that represent the observed state of the past weather. We explore the potentials and limitations of the current state-of-the-art RCMs by analyzing the simulated spatial and temporal distributions of precipitation over South Asia. The experimental setup of each model and the observationally based datasets are described in section 2. Section 3 presents the analysis of the RCMs to represent the Indian monsoon. The analysis is divided in five subsections: spatial distribution, temporal distribution, monsoon onset and withdrawal, Hovmöller diagram, and investigation of added value. A general summary and conclusions are given in section 4.

2. Experimental setup

For the three EU projects mentioned in section 1, four RCMs were applied over the South Asian region: HIRHAM5 (Christensen et al. 2006) from the Danish Meteorological Institute in WATCH, REMO (Jacob
et al. 2007) from the Max Planck Institute for Meteorology (MPI-M) in WATCH and HighNoon, Met Office Hadley Centre’s regional climate model version 3 (HadRM3; Buonomo et al. 2007) in HighNoon, and COSMO CLM (Dobler and Ahrens 2010) from the University of Frankfurt in BRAHMATWINN. These RCMs generated simulations over South Asia with the same reanalysis driving field. The four models are shortly described in Table 1. The four RCMs used in this study are developed in Europe and they are adjusted for the midlatitude’s atmospheric dynamics. Although a RCM domain is portable around the globe, South Asia remains a nonnative domain for those RCMs where their parameterizations, including particularly their convection schemes, are not adapted to the tropical climate. In their study, using the RCM CLM with the same configuration over seven domains over the globe, Rockel and Geyer (2008) found that the quality of the simulations for temperate and continental climate was similar to the one over Europe. However, they uncovered a systematic land–sea contrast for the tropical climate with overestimation of precipitation over the warm oceans, stressing the major role of the convection scheme. They concluded that, in order to get optimal results, one standard model configuration is not appropriate for all climate zones.

The orography of South Asia and the delimitations of the four RCM domains, without their buffer zones, are shown in Fig. 1. The domain locations were chosen to cover the whole of India and to avoid the lateral boundaries crossing the Himalayas. The HIRHAM5 and REMO models simulated the climate on an almost identical domain with a spatial resolution of 0.5° (~55 km). The HadRM3 uses a smaller domain with a resolution of 0.44° (~49 km). Finally, CLM uses a comparable domain size to REMO and HIRHAM5 but with a resolution of 0.44° (~49 km) and a slight northward shift. It should be noted that the position of the southern boundary, close to the equator, may have an impact on the large-scale atmospheric circulation because of the intertropical convergence zone and the ascending branch of the Hadley cell, which cannot be fully developed in the buffer zone of the RCM domains.

The model with the smallest domain is expected to be more constrained by the driving field and should have less freedom than the other models, as explained in Lucas-Picher et al. (2008). In a sensitivity study using an RCM with different domain sizes over South Asia, Bhaskaran et al. (1996) found that the simulations are insensitive to the domain size for the mean bias and the variability. This finding is in contrast with a corresponding study over Europe (Jones et al. 1995), which revealed a strong dependence of the RCM results on domain size. Bhaskaran et al. (1996) suggested that the synoptic disturbances generated within a midlatitude RCM domain are larger in amplitude than those generated in a tropical domain and, to a greater extent, capable of interacting with the long-wave circulation inherited from the driving model. To extend those studies, the potential impact of the domain size and location on the simulations will be considered throughout the analysis.

For each model, a continuous simulation was realized for the 1958–2002 period with ERA-40 (Uppala et al. 2005) LBCs at T159 horizontal resolution from the ECMWF modeling system. The sea surface temperature taken from ERA-40 is prescribed to the models. The “u” and “v” components of the winds, atmospheric temperature, specific humidity, and surface pressure are transmitted to the RCM every 6 h for each atmospheric level of the respective model.
Throughout the paper, the models are validated using different observationally based gridded datasets, which are described in Table 2. It is important to mention the following:

- The number of stations per grid cell is available for the Asian Precipitation–Highly Resolved Observational Data Integration towards Evaluation of the Water Resources (APHRODITE) project and the Global Precipitation Climatology Centre (GPCC). This information can be very useful to determine to what extent the gridded precipitation is determined from station data or derived using interpolation between the stations.

- Although the Climate Prediction Center (CPC) Merged Analysis of Precipitation (CMAP) and Global Precipitation Climatology Project (GPCP) datasets are comparable, they use different satellite measurements and algorithms. Yin et al. (2004) compared GPCP and CMAP for the period 1979–2001. Their analysis shows that both products are close over land and that GPCP is more reasonable over the ocean than CMAP.

Finally, the ERA-40 reanalysis fields are presented throughout the paper to determine the information about variables generated and provided by the driving field. The wind and mean sea level pressure variables, derived from the ERA-40 reanalysis, provide useful estimates where no observationally based gridded dataset is available. ERA-40 is a comprehensive reanalysis of the state of the atmosphere using measurements from satellites, weather balloons, and ground stations. Variables such as temperature and mean sea level pressure can be considered as quasi-observed datasets because they are directly impacted by the measurements owing to data assimilation. However, precipitation and wind speed are generated by the ERA-40 numerical weather prediction model, which is driven by the assimilated measurements.

3. Results

The analysis is performed on the common domain of the four RCMs corresponding to the entire domain of HadRM3, without its buffer zone. For most of the analysis, the fields of each model are shown on their respective grids and are not interpolated to a common grid. This approach preserves the spatial structures and prevents artificial smoothing because of the interpolation required to change the spatial projection. Finally, the analysis focuses on the monsoon season from June to September 1981–2000 because of the availability ofobservationally based datasets for this period. Since the simulations are started in 1958, it is expected that each model had enough time for spinup and that all simulations are in equilibrium during the analysis period 1981–2000.

a. 1981–2000 climatology (spatial distribution)

To get a first insight of the monsoon simulated by the RCMs, Fig. 2 presents the June–September (JJAS)
1981–2000 precipitation climatology for the four RCMs (CLM, HadRM3, HIRHAM5, and REMO) and five observationally based gridded datasets [APHRODITE, Climatic Research Unit (CRU), GPCP, GPCC, and Willmott and Matsuura (WIL_MAT)] and ERA-40. In general, the RCMs have similar spatial distributions of precipitation than the observationally based datasets. Among the precipitation patterns well captured by the models, we notice the precipitation band on the west coast of India associated with the orographically induced uplift of the humid low-level westerlies from the Arabian Sea due to the presence of the Western Ghats. This precipitation band produced by the RCMs is closer to the observationally based datasets than the ERA-40 precipitation. This is due to the higher horizontal resolution of the RCMs, which allows a better definition of the orography compared to ERA-40. Other spatial features well captured by the RCMs are the relative high amount of precipitation in the central part of India and on the west coast of Myanmar.

However, close examination of the regional distribution of precipitation reveals significant differences between the RCMs but also between the observationally based datasets. Some of those differences are located in the north of India, in Bangladesh, and in Nepal. Compared to the observationally based datasets, the CLM model is generating too much precipitation on the west coast of India and not enough in the center of India and in the Himalayas. This feature was also observed by Rockel and Geyer (2008), who ran the CLM model over South Asia. They suggested that the overestimation of precipitation over the mountain ranges (Western Ghats) and warm oceans seems to remove moisture from the atmosphere and to generate weak precipitation over land (as observed in central India). The HIRHAM5 model has a dry bias in Bangladesh and Myanmar and exaggerates precipitations on the east coast of India. As observed for other simulations over Africa and Europe, the HIRHAM5 model seems to generate too much precipitation on the coast and has difficulty transporting moisture over the land.

The spatial distributions of the observationally based gridded datasets at 0.25°, 0.5°, and 2.5° (Figs. 2e–i) are generally consistent, except in Myanmar and the north-east of India, where observation stations are scarce according to Figs. 2k,l, which show the number of stations per grid cell for APHRODITE and GPCP. In these regions, the precipitation distribution depends crucially on the interpolation procedure required to provide

---

**Table 2. Summary of the six observationally based gridded datasets used in the present study.**

<table>
<thead>
<tr>
<th>Variables</th>
<th>APHRODITE</th>
<th>CRU 2.1</th>
<th>GPCC</th>
<th>WIL_MAT 2.01</th>
<th>CMAP</th>
<th>GPCP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time resolution</td>
<td>Daily</td>
<td>Monthly</td>
<td>Monthly</td>
<td>Monthly</td>
<td>Pentad (five-day mean)</td>
<td>Pentad (five-day mean)</td>
</tr>
<tr>
<td>Spatial resolution</td>
<td>0.25° × 0.25°</td>
<td>0.5° × 0.5°</td>
<td>0.5° × 0.5°</td>
<td>0.5° × 0.5°</td>
<td>2.5° × 2.5°</td>
<td>2.5° × 2.5°</td>
</tr>
<tr>
<td>Source</td>
<td>Observation network</td>
<td>Observation network</td>
<td>Observation network</td>
<td>Observation network</td>
<td>Observation network and satellite</td>
<td>Observation network and satellite</td>
</tr>
<tr>
<td>Land–ocean coverage</td>
<td>Land</td>
<td>Land</td>
<td>Land</td>
<td>Land and ocean</td>
<td>Land and ocean</td>
<td>Land and ocean</td>
</tr>
<tr>
<td>Spatial coverage</td>
<td>Monsoon Asia, Russia, and Middle East</td>
<td>Global</td>
<td>Global</td>
<td>Global</td>
<td>Global</td>
<td>Global</td>
</tr>
</tbody>
</table>

a APHRODITE data are available online at http://www.chikyu.ac.jp/precip/.
b CRU data are available online at http://www.cru.uea.ac.uk/cru/data/hrg/cru_ts_2.10/.
c GPCC data are available online at ftp://ftp-anon.dwd.de/pub/data/gpcc/html/fulldata_download.html.
d WIL_MAT data are available online at http://climate.geog.udel.edu/~climate/html_pages/download.html.
e CMAP data are available online at http://www.esrl.noaa.gov/psd/data/gridded/data.cmap.html.
f GPCP data are available online at http://lwf.ncdc.noaa.gov/oa/wmo/wdcamet-ncdc.html.
a continuous gridded coverage despite the lack of stations. Moreover, the stations are probably in the mountain valleys where the cities are located. Consequently, the precipitation amounts are not well observed and probably underestimated in the Himalayas, Ghats, and the Arakan Yoma in the northwest of Myanmar. In general, the ERA-40 reanalysis closely reproduces the observationally based datasets.

Figure 3 shows the JJAS 1981–2000 2-m temperature bias for the four RCMs—CLM, HadRM3, HIRHAM5, and REMO—and five observationally based datasets: APHRODITE, CRU, GPCP, GPCC, WIL_MAT, and ERA-40. The figures showing precipitation (mm day$^{-1}$) have the legend at the top right. The number of stations per grid cell is shown for APHRODITE and GPCC with legend at bottom right.
which, with a grid spacing of 1.125°, are not well resolved in ERA-40.

Recent studies showed that irrigation, agricultural intensification, and land-use changes can have a major impact on the climate of South Asia (Douglas et al. 2009, Saeed et al. 2009, and Niyogi et al. 2010). While addressing a similar warm bias in REMO, Saeed et al. (2009) argued that northern India and Pakistan (Indus and Ganges basins) are the most intensely irrigated regions of the world and nonrepresentation of irrigation may lead to this warm bias. In a sensitivity experiment with the representation of irrigation in REMO, they

---

**FIG. 3.** JJAS 1981–2000 2-m temperature bias for the four RCMs—(a) CLM, (b) HadRM3, (c) HIRHAM5, and (d) REMO—and (e) the observationally based dataset WIL_MAT and (f) ERA-40 with regard to (g) the observationally based dataset CRU (°C). The colors of the bias maps are associated with the top right legend and the colors of the absolute map of CRU are associated with the bottom right legend.
showed a cooling potential of up to 5°C, which was attributed to the local recycling of moisture. The appearance of a similar warm bias in HIRHAM5 and CLM also points to the lack of representation of irrigation for that region. In contrast, HadRM3 does not show a warm bias but rather a small cold bias over that region. This is mainly related to the overestimation of precipitation, which leads to an overly wet surface state and enhanced evaporation (not shown). Additionally, the size of the HadRM3 domain may also contribute to this compensation. Because of its small size, the HadRM3 has probably a stronger control from the driving field and less freedom to drift away from the LBCs compared to the three other models having a larger domain.

Figure 4 shows the JJAS 1981–2000 mean sea level pressure for the four RCMs and ERA-40. Most models underestimate the mean sea level pressure compared to ERA-40. This behavior is linked to the warm bias of the RCMs, which creates a heat low, especially in northwestern India and Pakistan. The warm bias and the underestimation of the mean sea level pressure modify the differential heating over the land and the ocean and perturb the atmospheric dynamics of the system, which affects the wind at 850 hPa (Fig. 5). HIRHAM5, CLM, and HadRM3 overestimate the wind compared to ERA-40, especially in the Arabian Sea and the Bay of Bengal, while REMO does the opposite. It is also interesting to note the difference in wind direction and intensity over northern India. CLM and HIRHAM5, which have a dry bias in northern India (Figs. 2a,c), have slow winds in the north of India (Figs. 5a,c) compared to ERA-40, REMO, and HadRM3. As pointed out by Ji and Vernekar (1997), the strength and the position of the Somali jet, which is generally too weak in GCM simulations, are crucial in determining the precipitation pattern over the Indian peninsula. The strong winds of CLM and HIRHAM5 in the Arabian Sea may also explain why the precipitation on the west coast of India, which is generated by the uplift

FIG. 4. JJAS 1981–2000 average mean sea level pressure (hPa) for the four RCMs—(a) CLM, (b) HadRM3, (c) HIRHAM5, and (d) REMO—and for (e) ERA-40.
of the air meeting the Western Ghats, is larger compared to the observationally based datasets.

b. 1981–2000 mean annual cycle over selected large basins (temporal distribution)

The second part of the analysis focuses on the mean annual cycle of precipitation and temperature. We chose three basins (Indus, Brahmaputra, and Ganges) from the 5-min large watershed delineation datasets of Graham et al. (1999) and two basins (Krishna and Godavari) from the 0.5° major river basin of Oki and Sud (1998). Figure 6 shows the five basins interpolated to the HIRHAM5 model grid. The Ganges and Brahmaputra basins have similar characteristics, with strong rainfall at the edge of the Himalayas. The Godavari and Krishna basins of southern India have a warm spring and a relatively cold summer after the monsoon precipitation starts. Finally, we considered the Indus basin in Pakistan and the northwest of India, where the conditions are mainly warm and dry in the plains and much colder in the Himalayas. These five divisions provide a good representation of different types of climate in South Asia. The climate of those divisions, compared with the observations, will indicate the performance of the models for different types of climate all affected by the Indian monsoon. Few grid cells in the north of the Indus basin are missing from the HadRM3 domain, but this has a minor impact on the analysis because of the dry conditions.

Figure 7 shows the 1981–2000 mean annual cycle precipitation simulated by the four RCMs and ERA-40 for the five basins. The precipitation from four observationally based datasets (APHRODITE, CRU, GPCC, and WIL_MAT) is shown altogether with a gray shade to get an estimate of the uncertainty from the observations. For Indus, Ganges, and Brahmaputra (Figs. 7a–c) located in northern India, we see a large spread of precipitation among the RCMs. HadRM3 overestimates the precipitation compared to the observationally based datasets, while CLM and REMO underestimate the precipitation. The situation is different for the Godavari and Krishna (Figs. 7d,e) located in southern India, where HIRHAM5 overestimates the precipitation and REMO has the opposite trend. In general, the observationally based datasets agree well with a small spread (indicated by the gray shade in Fig. 7). In this case, the CMAP and GPCP were not used because they were underestimating the precipitation close to the Himalayans, probably owing to their coarse horizontal resolution of 2.5°. ERA-40 is generally in good agreement with the observationally based datasets, except for the Indus basin, where ERA-40 is overestimating the precipitation. While the amounts of precipitation are different between the models, the temporal distribution of the precipitation is well captured by the RCMs, as shown by the high correlation coefficients (indicated on the Fig. 7) between the models and the observations.

With respect to the 1981–2000 mean annual cycle of temperature, HIRHAM5 and REMO have a warm bias for the Indus and Ganges basins (Figs. 8a,b) in northern India. For Godavari and Krishna in the south of India, a warm spring is clearly observed (Figs. 8d,e), followed by colder conditions associated with the monsoon precipitation (Figs. 7d,e), which cools the surface through the evaporation process. In October and November, the warming over Godavari and Krishna is simulated differently by each model because of the land surface scheme, which evaporates the water accumulated during the summer monsoon period (Figs. 8a,b). For the Brahmaputra basin, it is interesting that HIRHAM5 and REMO have a warm bias (Fig. 8c) compared to the observations, despite the fact that the amount of water was reasonably well simulated (Fig. 7c). HadRM3 has a cold bias for the Indus basin (Fig. 8a), which is probably linked to the overestimation of precipitation (Fig. 7a). For all basins, the observationally based datasets CRU and WIL_MAT are in good agreement with a spread of less than 2°C all throughout the years. The 2-m temperature of ERA-40 follows closely the observationally based datasets. In summary, the temperatures simulated by the RCMs have a good temporal distribution, as shown by the high correlation coefficients between the RCMs and the observations, but are generally too warm, particularly for REMO.

c. Monsoon onset and withdrawal

Since the monsoon onset and circulation are mainly driven by the differential heating of the Indian Ocean and the adjacent land areas (Krishnamurti and Ramanathan 1982), the warm bias of most RCMs can affect the strength and onset of the Indian monsoon circulation. To determine if the timing of the monsoon is well captured by these models, the monsoon onset and withdrawal were computed for the RCMs, the observationally based datasets, and ERA-40 using the normalized pentad precipitation index (NPPI) presented by Kitoh and Uchiyama (2006). The NPPI is defined as

\[
NPPI = \frac{P - P_{\text{min}}}{P_{\text{max}} - P_{\text{min}}},
\]

where \( P \) is the pentad precipitation climatology, and \( P_{\text{max}} \) and \( P_{\text{min}} \) are the annual maximum and minimum \( P \) at each grid point, respectively. Compared to Kitoh and Uchiyama (2006), we did not smooth the pentad precipitation climatology. The onset date is then defined as
FIG. 5. As in Fig. 4, but showing wind direction (arrows) and wind speed (colors) (m s$^{-1}$) at 850 hPa.
the Julian pentad when NPPI exceeds threshold value (0.618) for the first time. The withdrawal date is defined as the pentad when NPPI drops below this threshold for the last time of the annual cycle. The threshold value of 0.618, corresponding to the golden ratio, was chosen in Zeng and Lu (2004), where they obtained a reasonable global picture of the onset and withdrawal. Other methods can be used to determine the beginning and the end of the monsoon season, but the goal of this analysis is to verify the agreement between the models and the observationally based datasets, regardless of the method used to determine the timing. The computation is based on the 1981–2000 pentad climatology. Consequently, the onset and withdrawal identified with the NPPI do not correspond to the mean beginning and end of the monsoon, but rather to an index computed from the precipitation climatology. This index removes the bias of the model computation for the onset and withdrawal by normalizing all the values of the pentad to generate a value between 0 and 1. The value 1 of the NPPI is reached once during the annual cycle and corresponds to the strongest pentad from the climatology.

Figure 9 shows the monsoon onset as computed with the NPPI for the four RCMs, three observationally based datasets, and ERA-40. The monsoon onset is at pentad 32 or 34 on the east coast of India and then it increases, going inland in a northwest direction. The spatial distribution of the onset is slightly different from one model to another, while it is consistent between the observationally based datasets. The onset date of CLM is later than that of the other models and may be associated with the dry bias of this model. Also, the onset for CLM occurs later than pentad 40 on the east coast of India. This is an artifact of the NPPI, as large precipitation events happening late during the mean annual cycle influence its computation (see also below). Late monsoon onset also due to this artifact can also be observed in the south of India for all the observationally based datasets, but not on the east coast. The monsoon onset is generally well captured by REMO, but in the north and south of India, there are few places where the onset is later than pentad 40. This behavior is also due to excessive precipitation events simulated later in the year. The RCMs differ more significantly in monsoon withdrawal than in onset simulation (see Fig. 10). It appears that the withdrawal is difficult to capture and is largely model dependent. In contrast, the CLM withdrawal simulates well the observationally based datasets, while the withdrawals of the other RCMs occur in general too late. As for the onset, there is still a good consistency between the withdrawal periods in the observationally based datasets.

d. Hovmöller diagrams for precipitation (temporal and spatial analyses)

The Hovmöller diagram integrates the spatial and temporal analyses, which allows us to visualize the motion of the intertropical convergence zone and the monsoon. In the following analysis, a Hovmöller diagram is used by zonally averaging the data in a box between 10°N to 25°N, and 78°E to 90°E, indicated in black in Fig. 6. This box was selected to follow the propagation of the Indian monsoon inland, toward the north of India, without taking into account the precipitation band on the west coast of India. The analysis is done over land to exclude the coastal sea precipitation and to take advantage of the high-resolution APHRODITE dataset only available over land.

Figure 11 shows the Hovmöller diagrams for the 1981–2000 mean annual cycle of precipitation of the selected box for the four RCMs, three observationally based datasets, and ERA-40. For all plots, there is a distinct northward motion of the monsoon precipitation from May to August, and then a southward motion from August to November. The CLM (Fig. 11a) underestimates the precipitation, which is in agreement with the dry bias of this model. Conversely, HadRM3 (Fig. 11b) overestimates the precipitation during the monsoon period in the north of India. The strong bias of HIRHAM5 on the east coast of India is clearly visible between 15° and 20°N in July, August, and September. Finally, the migration of the monsoon for REMO (Fig. 11d) is close to the observations, but shows overestimation of the precipitation for a few events depicted by the red stripes through the monsoon period. The observationally
based datasets (Figs. 11e–g) agree well with each other, but the values are higher and with a sharper time resolution for APHRODITE, owing to the daily values compared to pentad values for GPCP and CMAP. ERA-40 has a low bias in precipitation compared to APHRODITE, which is in agreement with the previous results (see Fig. 2). Overall, the latitudinal motion of the Indian monsoon is well captured by the RCMs, but the intensity of precipitation varies between the models.

e. Investigation of added value using spatial filtering

The last part of the analysis focuses on the identification of added value and mesoscale details generated by the RCMs. To investigate the added value of each RCM, a smooth field of the JJAS 1981–2000 precipitation is generated by computing a 5 × 5 grid point running spatial average (Figs. 12a–d) for each model. With this method, the spatial resolution of a field is similar to the one of a GCM (~250 km), allowing us to observe the large-scale
signal. Then, to extract the small-scale features and identify the mesoscale signal (Figs. 12f–i), the original precipitation field (Figs. 2a–d) of each RCM is subtracted from its smoothed field (Figs. 12a–d). It is worth mentioning that this analysis does not identify the “real added value,” which is only obtained by comparing a GCM and an RCM. However, this analysis approximates the added value without handling the bias of two different models (RCM and its driving field) by comparing a model variable to itself using a spatial filter. This computation is also done with the observationally based dataset APHRODITE (originally at 0.25°) using a 10 × 10 grid point running spatial average over land only (Fig. 12e) to obtain a field at a similar resolution as the smooth RCM fields of ~250 km. Then, the mesoscale details from APHRODITE (Fig. 12j) are obtained by subtracting the original fields (Fig. 2e) from the smoothed field (Fig. 12e). The smoothed and residual fields from APHRODITE are not directly comparable to the RCM fields, because only land values are available for APHRODITE, but this

FIG. 8. As in Fig. 7, but for 2-m temperature (°C) and two observationally based datasets (CRU and WIL_MAT).
still gives us an idea of the small-scale details in the observationally based dataset.

In Figs. 12a–d, the smoothed fields of precipitation exhibit the large-scale features by filtering the small-scale details like the precipitation band on the west coast of India. The smoothed fields are closer to each other than the original fields, but they are still different. The residual fields (Figs. 12f–i), which show the mesoscale signal, have similar patterns between each model with a precipitation band on the west coast of India in the Himalayas and on the coast of Myanmar. Those mesoscale patterns are also visible in the residual field of APHRODITE (Fig. 12j). From the residuals of HIRHAM5 (Fig. 12h), we observe that the precipitation bands on the coast are shifted compared to the other models. This indicates that the precipitation is treated differently in

Fig. 9. Monsoon onset (pentads) computed with the normalized pentad precipitation index from 1981–2000 for the four RCMs—(a) CLM, (b) HadRM3, (c) HIRHAM5, and (d) REMO—for three observationally based datasets—(e) APHRODITE, (f) CMAP, and (g) GPCP—and for (h) ERA-40.
HIRHAM5. HIRHAM5 and HadRM3 (Figs. 12g,h) show a noisy field of the residuals compared to CLM and REMO (Figs. 12f,i). It appears that HIRHAM5 and HadRM3 are more sensitive to the orographic forcing and generate small-scale features that are not present in REMO and CLM.

4. Summary and conclusions

In this study, we verified the ability of four RCMs (CLM, HadRM3, HIRHAM5, and REMO) to represent the Indian monsoon characteristics for the period 1981–2000, using LBCs from ERA-40. The precipitation distribution simulated on the west coast of India, in central India, and on the coast of Myanmar is well captured by the models. However, some RCMs (CLM and HIRHAM5) show precipitation biases, especially on the east coast of India, in Bangladesh, and in the Himalayas, which could be due to missing or poor representation of regional processes and feedbacks. Most models (CLM, HIRHAM5, and REMO) are too warm compared to the observations in northern India. It seems that land-use changes
FIG. 11. Hovmöller diagrams for the 1981–2000 average precipitation (mm day$^{-1}$) between 78° and 90°E (see black rectangle in Fig. 6) on land only for the four RCMs—(a) CLM, (b) HadRM3, (c) HIRHAM5, and (d) REMO—for three observationally based datasets—(e) APHRODITE, (f) CMAP, and (g) GPCP—and for (h) ERA-40.
FIG. 12. JJAS 1981–2000 average precipitation (mm day$^{-1}$) smoothed with a $5 \times 5$ grid point running spatial average for the four RCMs: (a) CLM, (b) HadRM3, (c) HIRHAM5, and (d) REMO. (e) Respective field for the APHRODITE dataset using a $10 \times 10$ grid point running spatial average. JJAS 1981–2000 average precipitation (mm day$^{-1}$) residuals computed by subtracting the original fields (Figs. 2a–e) with the smoothed field (Figs. 12a–e) for the four RCMs—(f) CLM, (g) HadRM3, (h) HIRHAM5, and (i) REMO—and for the (j) APHRODITE dataset.
and irrigation, not taken into account by any of the RCMs, may serve to cool the surface with increased evaporation (Douglas et al. 2009; Saeed et al. 2009). As a consequence of the warm bias, the RCMs’ mean sea level pressure in northern India is lower than in the ERA-40 reanalysis. This further perturbs the land–sea temperature contrast, which is one of the monsoon forcings, and also influences the large-scale dynamics as indicated by the wind speed overestimation of three RCMs (CLM, HadRM3, and HIRHAM5) compared to ERA-40 over the Bay of Bengal and the Arabian Sea.

Over the five basins documented, the mean annual cycles of precipitation and 2-m temperature underline the biases identified in the spatial distributions. There is a spread between the precipitation produced by the RCMs over the Ganges and Brahmaputra basins located in the northeast of India and Bangladesh, at the edge of the Himalayas. For the Godavari and Krishna basins located in the south of India, HIRHAM5 is too wet. For the Indus basin in the northwest of India, HadRM3 overestimates the precipitation. Most RCMs show a warm bias over the basins, except for the HadRM3, which has a cold bias for the Indus basin, which is probably linked to its wet bias. The monsoon onset, analyzed using the normalized pentad precipitation index (NPPI; Kitoh and Uchiyama 2006), is in general well captured by the RCMs. However, the monsoon withdrawal differs between models and observations. As depicted by the Hovmöller diagrams, the models reproduce adequately the meridional motion of the monsoon, but the amounts are different. Finally, the “added value” of the RCMs was investigated using a 5 × 5 grid point running spatial average to smooth the JJAS precipitation. The mesoscale signals identified with the residual field are similar in nature among the RCMs, with high values on the west coast of India, the foothill of the Himalayas, and the coast of Myanmar.

We found that the RCMs produce large-scale monsoon features for precipitation and temperature that compare reasonably with observationally based datasets. However, the simulations are less in agreement with observations in some regions. In certain cases, RCMs add details at the regional scale, which were not available from the forcing data, but the large spread of solutions between the RCMs reveal that the added regional details are not yet as robust in certain areas as one would hope for. One possible way to circumvent the regional disparities between RCMs and the observationally based datasets is to apply bias correction (Dobler and Ahrens 2008; Piani et al. 2010). However, this requires us to define any climate change information within a framework of uncertainty, which remains difficult. In addition, the added value of a RCM may become more obvious if GCM data are used for climate change studies, as these do not contain observational data like ERA-40.

The variation between the spatial distributions of precipitation generated by the different RCMs exposes the great sensitivity of some components of the RCMs, such as the convection and land surface schemes to simulate the monsoon dynamics. In this regard, more work is required to find suitable convection schemes and their tuning to generate reliable climate model simulations. Moreover, according to the spread between RCMs and their disagreement with observations, it is obvious that some feedbacks or processes are poorly, or not, taken into account in the state-of-the-art RCMs. There is a pressing need to include new components, which simulate the missing processes, in the climate models because those processes might play an important role at the regional scale. Some initial studies have been done in this way by looking at the influence of irrigation (Douglas et al. 2009; Saeed et al. 2009). Other land-use changes, which were found to have an impact on the monsoon, are urbanization (Kishtawal et al. 2010) and soil moisture (Chang et al. 2009). Also, some studies are verifying the impact of the aerosol on the monsoon (Meehl et al. 2008; Wang et al. 2009). Finally, recent work implies that a climate simulation is substantially more realistic, with respect to the spatial and temporal distribution of monsoon rainfall, when using a regional coupled atmosphere–ocean model compared to an uncoupled atmosphere-only model (Venkata Ratnam et al. 2009).

A recent study exposed the differences between simulations using the same RCM but different reanalyses as LBCs (Wang and Yang 2008). It would be interesting to explore the sensitivity of one RCM driven at the boundaries with different reanalyses over South Asia. Moreover, a systematic study is requested to determine the optimal size and location of the limited area domain used by a RCM. Finally, whether the RCM produces wet and dry spells, as the one identified by Singh and Ranade (2010) using observationally based data, remains to be determined. CORDEX (Giorgi et al. 2009), which is a worldwide effort to conduct RCM simulations in a common framework, has the potential to provide valuable information about the ability of RCMs to simulate the monsoon, given the fact that one of the domains chosen is covering South Asia. The Regional Climate Model Intercomparison Project for Asia (RMIP; Fu et al. 2005) will also provide insightful information on a domain covering eastern Asia and encompassing India.

Acknowledgments. This study was supported by funding from the European Union within the WATCH.
REFERENCES


Lucas-Picher, P., D. Caya, S. Biner, and R. Laprise, 2008: Quantification of the lateral boundary forcing of a regional project by the EU projects HighNoon (Contract 227087) and BRAHMATWINN (Contract 036592). Also, Andrew Wiltshire was financially supported by the Joint DECC–Defra Met Office Hadley Centre Climate Program (Contract GA01101). B. Ahrens acknowledges funding from the Hessian initiative for the development of scientific and economic excellence (LOEWE) at the Biodiversity and Climate Research Centre (BiK-F), Frankfurt/Main. We thank Dr. Maryse Picher for editing the manuscript.


