Evaluation of global climate models for Indian monsoon climatology

Evan Kodra\textsuperscript{1}, Subimal Ghosh\textsuperscript{2} and Auroop R Ganguly\textsuperscript{1}

\textsuperscript{1} Civil and Environmental Engineering, Northeastern University, Boston, MA 02115, USA
\textsuperscript{2} Civil Engineering Department, Indian Institute of Technology Bombay, Mumbai 400076, India

E-mail: a.ganguly@neu.edu

Received 27 November 2011
Accepted for publication 10 January 2012
Published 1 February 2012
Online at stacks.iop.org/ERL/7/014012

Abstract
The viability of global climate models for forecasting the Indian monsoon is explored. Evaluation and intercomparison of model skills are employed to assess the reliability of individual models and to guide model selection strategies. Two dominant and unique patterns of Indian monsoon climatology are trends in maximum temperature and periodicity in total rainfall observed after 30 yr averaging over India. An examination of seven models and their ensembles reveals that no single model or model selection strategy outperforms the rest. The single-best model for the periodicity of Indian monsoon rainfall is the only model that captures a low-frequency natural climate oscillator thought to dictate the periodicity. The trend in maximum temperature, which most models are thought to handle relatively better, is best captured through a multimodel average compared to individual models. The results suggest a need to carefully evaluate individual models and model combinations, in addition to physical drivers where possible, for regional projections from global climate models.

Keywords: regional and decadal prediction, Indian monsoon, Indian temperature trends, climate model evaluation

Online supplementary data available from stacks.iop.org/ERL/7/014012/mmedia

1. Introduction

The effects of climate change on India are critical to multiple stakeholders, including resource managers and adaptation researchers, owing to a growing and vulnerable population along with changes in urbanization and land use (Garg et al. 2009, O’Brien 2004), as well as to policymakers because of emissions negotiations in light of India’s rapidly emerging economy (Pachauri 2009). Indian monsoon rainfall is critical to food and water security (Auffhammer et al. 2011, Gupta et al. 2010), while temperature has been identified as a predictor of monsoon rainfall (Parthasarathy et al. 1989) and also has important impacts on agriculture and public health (Dash and Mamgain 2011). Global climate models (GCMs) generate 21st century predictions for both temperature and precipitation; this allows for a discussion of two related hypotheses: (1) model evaluation of historical (20th century) GCM runs and their combinations (e.g. multimodel averages, or MMAs) through statistical error metrics offer predictive insights relevant for future regional projections (e.g. Indian monsoon climatology), and (2) predictive skills of GCMs, which may be shown to relate to their ability to capture key physical processes rather than improve statistical metrics alone, thus possessing additional credibility for regional projections in the future under non-stationary conditions.

2. Data and methodology

Rainfall and temperature are the two variables for which detailed observations exist, which have been extensively studied in the context of Indian monsoon (e.g. Kothyari and Singh 1996, Goswami et al. 2006, Ghosh et al. 2011) and which are known to have significant impacts as discussed previously. Trends and patterns of Indian monsoon mean and
The observed mean pattern is known to be dominated by low-frequency variability (Goswami et al. 2006, Ghosh et al. 2011). While mean temperature trends in India have been reported to be similar to those seen around the globe, the region’s maximum and minimum temperature patterns have been anomalous: maximum temperature has shown a significant increasing trend and has contributed overwhelmingly to the long-term upward trend in mean temperature, while no significant long-term trend has been found for minimum temperature (Kumar et al. 1994). On the other hand, accelerated warming has been reported (Kothawale et al. 2010) in minimum, mean and maximum temperatures more recently (1971–2007). The minimum temperature patterns in the pre-monsoon period are also known to be predictors of the total monsoon rainfall.

Based on the above considerations, the monsoon rainfall (June–July–August–September), as well as maximum and minimum temperatures (both March–April–May), are obtained over India from observations and model-simulations. The observed data were obtained from the Indian Institute for Tropical Meteorology in Pune, India. We select only the seven publicly available GCMs from the third phase of the World Climate Research Programme’s (WCRP) Coupled Model Intercomparison Projects (CMIP3) for which all three variables are publicly available. The data are available from the Program for Climate Model Diagnostics and Intercomparison (PCMDI) website maintained by the Lawrence Livermore National Laboratory of the United States Department of Energy. Since multiple initial condition runs are not available for all models, one initial condition run is selected for each (in the supplementary information (S1) available at stacks.iop.org/ERL/7/014012/mmedia), alternative initial condition runs are explored. The observational and model datasets are summarized in tables 1 and 2, respectively. Spatially averaged time series are obtained from each dataset. Then, anomalies (i.e. z-scores) of the variables rainfall, maximum temperature and minimum temperature are computed by subtracting their respective means and then dividing by their respective standard deviations from a sufficiently long baseline period (1900–99); a 30 yr moving average filter is subsequently applied to each variable to suppress high-frequency variability. Time series plots for rainfall are de-trended with ordinary least squares regression. These preprocessed variables are labeled AIMR (All India Monsoon Rainfall), TMIN (minimum temperature) and TMAX (maximum temperature). AIMR and TMIN share strikingly similar periodicity; TMAX shows a significant increasing trend, while TMIN does not, confirming past research (Kumar et al. 1994); (supplemental figure S1 available at stacks.iop.org/ERL/7/014012/mmedia). Thus, we appropriate AIMR and TMAX as the primary targets of interest, while TMIN is considered a predictor for AIMR. To evaluate the predictive ability of individual GCMs and MMAs, all possible equally weighted time series combinations of GCMs are compared to the observed record for AIMR and TMAX using, median absolute per cent error (MAPE). Additional skill metrics, including mean squared error (MSE), mean absolute deviation (MAD), and the probabilistic global search lausanne (PGSL) are used as another method for evaluating GCMs and compared to observations and are discussed in the S1 (figures S2 and S3 available at stacks.iop.org/ERL/7/014012/mmedia). In the main letter, we refer to a ‘best’ individual GCM as the GCM which performs best with respect to the skill metric of choice, MAPE. Moving average anomalies for AIMR and TMAX have been provided as supplementary files to facilitate reproducibility of the results.

### Table 1. Availability of observed data (Parthasarathy et al 1995).

| Variable          | Spatial resolution | Availability |
|-------------------|--------------------|--------------|
| Rainfall          | All India          | 1813–2006    |
|                   | Meteorologically homogeneous subdivisions | 1871–2008 |
| Maximum temperature | All India          | 1901–2003    |
|                   | Meteorologically homogeneous subdivisions | 1901–2003 |
| Minimum temperature | All India          | 1901–2003    |
|                   | Meteorologically homogeneous subdivisions | 1901–2003 |

### Table 2. Availability of GCM outputs (rainfall, maximum temperature and minimum temperature).

| Model name          | Center                                | Spatial resolution | Availability |
|---------------------|---------------------------------------|--------------------|--------------|
| GISS AOM (AOM)      | Goddard Institute for Space Studies (GISS), USA | 3.0° × 4.0°         | 1850–2000    |
| BCCR BCM2.0 (BCCR)  | Bjerknes Centre for Climate Research (BCCR), Norway | 2.8° × 2.8°        | 1850–1999    |
| MIROC3.2 HiRes (MIROC-HI) | National Institute for Environmental Studies (NIES), Japan | 1.1° × 1.1° | 1900–2000    |
| MIROC3.2 MedRes (MIROC-MED) | National Institute for Environmental Studies (NIES), Japan | 2.8° × 2.8°        | 1850–2000    |
| INM CM3.0 (INM)     | Institute for Numerical Mathematics (INM), Russia | 4.0° × 5.0°        | 1871–2000    |
| NCAR PCM (PCM)      | National Centre for Atmospheric Research (NCAR), USA | 2.8° × 2.8°        | 1890–1999    |
| CSIRO Mk3.0 (CSIRO) | Commonwealth Scientific and Industrial Research Organization (CSIRO), Australia | 1.8° × 1.8°        | 1871–2000    |
Figure 1. Evaluation of a seven model ensemble reveals the predictive skills of GCMs and their combinations. The skills in reproducing a moving average of anomalies, or 30 yr moving average of the $z$-scores, for de-trended AIMR ((a)–(c)) and TMAX ((d)–(f)) are measured as the MAPE. For de-trended AIMR periodicity, climate model skills are poor in general, but BCCR does significantly better than the rest, possibly owing to the ability to handle the Atlantic multi-decadal oscillation (AMO) and corresponding influence on AIMR periodicity. The addition of less skillful models in succession (a) causes the error to continuously increase, which is also visually apparent from the time series plots (b) and box-plots (c) of all model combinations. The box-plots show the skills of all model combinations, from all possible (127) model combinations. While the best individual GCM BCCR is ranked number 1, the 7-GCM MMA ranks 60. (Note: figure S5 (available at stacks.iop.org/ERL/7/014012/mmedia) displays results for AIMR when alternative initial condition runs are explored in the same manner as this figure, suggesting that some models and hence results are sensitive to choice of initial conditions.) For TMAX trends, climate model skills are known to be better compared to other variables hence multiple models and their combinations do relatively better ((d)–(f)). The best individual GCM INM, ranked 39 of 127, outperforms the MMA narrowly, which is ranked 48. Performance metrics other than MAPE are shown in the SI (figure S3 available at stacks.iop.org/ERL/7/014012/mmedia).

3. Results

3.1. Statistical evaluation of individual GCMs and MMAs

BCCR is clearly identified as the best GCM, statistically, for AIMR (although model-specific results are different when using alternate initial condition runs for several models, see figures S4 and S5 available at stacks.iop.org/ERL/7/014012/mmedia). The addition of any GCM after BCCR degrades error performance; past work (Kripalani et al 2007) has also shown that BCCR may be a viable candidate for simulating the Indian monsoon. The full 7-GCM MMA performs significantly worse than the best individual GCM for AIMR, which is apparent from visual inspection (figure 2). Figures 1(a)–(c) show the consistently higher relative ranking of the best individual GCM versus the MMA as well as the successive degradation of model skills as suboptimal models are cumulatively added to the better performing models (this same particular insight is also obtained with alternative initial condition runs, see figure S5 available at stacks.iop.org/ERL/7/014012/mmedia). Here, the greater contribution of the less optimal models predominates and drives the skill reduction. The inclusion or exclusion of the best individual GCM (figure 2(a)) within the MMA does not appear to change the relative results significantly, showing that the inclusion of the other six GCMs results in a reduction from the statistical skills of the best individual GCM.

INM is the best individual GCM for TMAX, but by a narrow margin. The full seven-model MMA outperforms the best individual GCM. As more models are added to an MMA, the MMA appears to asymptotically approach a stable and lower error metric than is achieved with most individual models (figures 1(e) and (f)). The inclusion of INM does not appear to affect the MMA substantially (figure 2(b)).

Overall, these results suggest that evaluation of the historical GCM runs may reveal insights about their credibility and inform the model selection process, and that in the case of AIMR an a priori choice of the MMA of the full multimodel ensemble may be inappropriate, while it may be acceptable for TMAX. The same overall conclusions are established with other common statistical metrics (see figure S3 and tables S1–S3 available at stacks.iop.org/ERL/7/014012/mmedia).

3.2. Exploring explanations for model performance

Next, we form hypotheses for the broad physical intuition for the relative performance of GCMs with corresponding empirical tests as applicable, both for AIMR (figure 3) and TMAX (figure 4).

3.2.1. AIMR. Current understanding of the aggregate physical mechanisms relating North Atlantic sea surface temperature anomalies and AIMR (Goswami et al 2006) is outlined
in figure 3(a). The Atlantic multi-decadal oscillation (AMO), with a periodicity of 65–70 yr, influences (Schlesinger and Ramankutty 1994, Delworth and Mann 2000) the meridional gradient of tropospheric temperature in the region, which in turn drives the Eurasian temperature with a periodicity of about 70 yr and influences the AIMR periodicity, which we find to be about 67 yr through a best-fit sinusoidal on observations (figure 3(c); figure S2 available at stacks.iop.org/ERL/7/014012/mmedia). The empirical determination of TMIN with a periodicity of about 60 yr as a predictor of AIMR may be derived from this relationship between Eurasian temperature and AIMR. The correspondence of AIMR with TMIN and the best-fit sinusoidal on observations (figure 3(c); figure S2 available at stacks.iop.org/ERL/7/014012/mmedia) is shown in figures 3(b) and (c). The ability of the best individual GCM, BCCR, to capture both the AIMR periodicity and the TMIN periodicity is shown in figures 3(d) and (e).

Confidence in the improved performance of BCCR alone over MMAs (figures 1(a)–(c)) could be reinforced by the dual facts that the multimodel IPCC-AR4 ensemble without considering BCCR has been shown to fail to capture the AMO (Knight 2009), while the BCCR has been shown to reproduce the AMO relatively well, although this performance is found via an ensemble of initial condition perturbations (Ottera et al 2010), while only one initial condition run for BCCR is publicly available and can be used in this work. On the other hand, in one recent climate model evaluation work (Stoner et al 2009), BCCR does not appear to reproduce temporal and spatial aspects of the AMO very realistically relative to other models. Thus, while the connection presented here between TMIN, AMO and AIMR may hold some promise, it must be interpreted carefully. Further caution is suggested by the comparison of alternate initial condition runs for GCMs that have more than one run publicly archived, where BCCR no longer stands out clearly as the most skillful GCM (see figures S4 and S5 available at stacks.iop.org/ERL/7/014012/mmedia).

3.2.2. TMAX. While INM is the best individual GCM for TMAX trends (figure 4(b)), the seven-member MMA performance is very similar. We suggest two (not necessarily mutually exclusive) hypotheses for explaining model performance results related to TMAX versus AIMR. First, climate models typically handle aggregate-scale temperature processes better than rainfall (figure 4(a); IPCC 2007), although the model skills may be relatively low in the tropics (Lin 2007) and for finer scale processes which dictate regional climate (IPCC 2007). Natural climate variability, as well as global man-made change other than greenhouse gas emissions-induced warming, may add to uncertainties at regional scales. Second, model-averaging may be better suited to the near monotonic and linear trend of TMAX. As a comparison, for AIMR the observed periodic pattern may be harder for model ensembles to capture owing to a possible dampening of periodicity upon averaging (i.e. upon averaging competing periodic signals, variability or periodicity may be reduced or flattened).

While scientific explanation of GCM skills may increase our confidence in projections of Indian monsoon climatology, credibility of the insights presented here may benefit from a
Figure 3. Physical basis for low-frequency AIMR periodicity and comparison of the best individual GCM BCCR with six other GCMs. The Atlantic multi-decadal oscillation (AMO), with a periodicity of 65–70 yr, leads to a meridional gradient of tropospheric temperature, which in turn causes a temperature anomaly in Eurasia, subsequently leading to the periodicity in AIMR of about 67 yr. The minimum temperature or TMIN (for March–April–May) over all India, which has been empirically shown to have predictive value over AIMR (June–July–August–September rainfall), may be in turn caused by the Eurasian temperature anomaly. The observed TMIN shows a periodicity of about 60 yr. The overall physical basis for AIMR periodicity is shown schematically (a). AIMR and TMIN clearly show similar periodicity (b) and the slight lag effect. The best-fit sinusoidal ((c), red and blue lines), which captures the primary low-frequency oscillation, shows a 67 yr periodicity for both TMIN and AIMR. The BCCR model is the only one of seven GCM which appears to capture the periodicity, amplitude and phase of both AIMR and TMIN ((d)–(e)) relatively well.

more comprehensive evaluation and exploration of the bases for model performance. We acknowledge that the exploration of bases in this work is limited to the two variables at hand and at present has been conducted over all of India, leaving more fine-grained analysis (or generalizations to other regions across the works) for future work.

4. Conclusion

We present an interpretable approach for evaluating climate model skill in the context of Indian monsoon climatology. Results suggest that BCCR may be capturing physical drivers of AIMR, and thus its skill may actually be founded on aggregate physical processes and not merely on chance. The use of equally weighted multimodel averages (MMAs) has been empirically justified as a robust measure in regional climate research literature (e.g. Pierce et al 2009, Santer et al 2009) and is sometimes used by default in assessment reports to policymakers (e.g. Karl et al 2009). However, there is substantial debate surrounding its use and appropriateness (Knutti 2010, Knutti et al 2010, Perkins et al 2009) especially for regional climate projections and assessments. In this work, the full seven-model MMA appears suboptimal compared to the best individual model for AIMR but more appropriate in the case of TMAX. The differences in the insights and performance of models (or their combinations)
suggest caution in the *a priori* use of multimodel averages when forming regional projections, specifically for Indian monsoon climatology. The importance of process-based evaluation of all climate models within an ensemble, when possible and appropriate, is emphasized. While apparent physical support of GCM skills could potentially increase our confidence in the projections, several important caveats mandate caution in their interpretation. Outputs of several GCMs may be quite sensitive to initial conditions at the regional scale at which their performance is evaluated (see figures S4 and S5 available at stacks.iop.org/ERL/7/014012/mmedia), increasing the complexity of producing credible regional projections. Evaluations of individual and multimodel performance and their generalization to regional projections or assessments deserve careful thought, particularly in view of variability from structural differences across models, differences owing to initial conditions, and intrinsic model variability. A distinction may need to be made between what could be considered a true skill of an individual model or a model combination versus statistical chance, especially since routine operations like averaging and de-trending could potentially introduce artifacts that seem to resemble signals where none may exist. Given non-stationarity under climate change and the importance of regional projections for adaptation and mitigation, a comprehensive study of model or multimodel skills along with a comprehensive assessment of the physical drivers may be appropriate. For Indian monsoon climatology, a thorough evaluation of the performance of the models and their combinations at multiple sub-regions over India based on mechanistic understanding may be appropriate.

**Acknowledgments**

This research was partially sponsored by the US Department of Energy Office of Science (FWP ERKP782) and the bulk of the work was performed at Oak Ridge National Laboratory (ORNL). ORNL is managed by UT-Battelle, LLC, for the US Department of Energy under contract DE-AC05-00OR22725. The work was completed through funding provided by the National Science Foundation Expedition Grant No. 1029711. Subimal Ghosh was supported at ORNL by the Department of Science and Technology of the Government of India through the Better Opportunities for Young Scientists in Chosen Areas of Science and Technology (BOYSCAST) fellowship.

**References**

Auffhammer M, Ramanathan V and Vincent J R 2011 Clim. Change at press (doi:10.1007/s10584-011-0208-4)
Dash S K and Mamgain A 2011 J. Appl. Meteorol. Climatol. **50** 1842–58
Delworth T L and Mann M E 2000 Clim. Dyn. **16** 661–76
Garg A, Dhiman R C, Bhattacharya S and Shukla P R 2009 Environ. Manage. 43 779–89
Ghosh S, Das D, Kao S-C and Ganguly A R 2011 Nature Clim. Change doi:10.1038/nclimate1327
Goswami B N, Madhusoodanan M S, Neema C P and Sengupta D 2006 Geophys. Res. Lett. 33 L02706
Gupta P K, Panigrathy S and Parihar J S 2010 NNRMS Bull. 35 57–62
IPCC (International Panel on Climate Change) 2007 Climate models and their evaluation Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change ed S Solomon et al (Cambridge: Cambridge University Press)
Karl T R, Melillo J M and Peterson T C (ed) 2009 Global Climate Change Impacts in the United States (Cambridge: Cambridge University Press)
Knight J R 2009 J. Clim. 22 1610–25
Knutti R 2010 Clim. Change 102 395–404
Knutti R et al 2010 J. Clim. 23 2739–58
Kothawale D R, Munot A A and Kumar K K 2010 Clim. Res. 42 89–104
Kothiyari U C and Singh V P 1996 Hydrol. Process. 10 357–72
Kripalani R H, Oh J H, Kulkarni A, Sabade S S and Chaudhari H S 2007 Theor. Appl. Climatol. 90 133–59
Kumar K R, Kumar K K and Pant G B 1994 Geophys. Res. Lett. 21 677–80
Lin J 2007 J. Clim. 20 4497–525
O’Brien K 2004 Glob. Environ. Change 14 303–13
Ottera O H, Bentsen M, Oranje H and Suo L 2010 Nature Geosci. 3 688–94
Pachauri R 2009 Nature 461 1054
Parthasarathy B, Munot A A and Kotawale D R 1995 Monthly and seasonal rainfall series for All-India homogeneous regions and meteorological subdivisions, 1871–1994 Research Report 65 (Pune: Indian Institute of Tropical Meteorology)
Parthasarathy B, Rupa Kumar K and Sontakke N A 1989 Theor. Appl. Climatol. 42 93–110
Perkins S E, Pitman A J and Sisson S A 2009 Geophys. Res. Lett. 36 L06710
Pierce D W, Barnett T P, Santer B D and Gleckler P J 2009 Proc. Natl Acad. Sci. USA 106 8441–6
Santer B D et al 2009 Proc. Natl Acad. Sci. USA 106 14778–83
Schlesinger M E and Ramankutty N 1994 Nature 367 723–6
Stoner A M K et al 2009 J. Clim. 22 4348–72