Monsoon prediction: are dynamical models getting better than statistical models?

Sulochana Gadgil and J. Srinivasan

The Indian summer monsoon rainfall (ISMR) has remained remarkably stable during the last 140 years. The ISMR has varied between 70% and 120% of the long-term average of about 85 cm. Monsoon seasons with ISMR of less than 90% of the average are considered to be droughts, whereas those with more than 110% rainfall are considered as excess rainfall seasons. Although the variation is not large (Figure 1), it has a substantive impact on our agriculture and gross domestic product (GDP)[1]. We, therefore, can never take the monsoon for granted and every April/May with the summer heat scorching the countryside, we start worrying about the upcoming monsoon. In April 2012, a paper on monsoon prediction by Delsole and Shukla[2] generated considerable interest[3]. It was shown that the skill of predicting ISMR with coupled atmosphere–ocean models initialized in May is statistically significant and is much higher than empirical prediction from sea surface temperatures (SSTs) of May. Does this imply that dynamical models are getting better than statistical models? We consider here the implications of this paper and recent developments in monsoon prediction.

Monsoon is predicted either by statistical models based on analysis of historical data to determine the relationship of ISMR, to a variety of atmospheric and oceanic variables over different parts of the world prior to the summer monsoon season, or by dynamical models based on the laws of physics. During 1932–2004, the operational statistical models evolved with changes in the predictors and finally use of power regression instead of linear regression, but the forecast skill did not improve much[4]. However, nonlinear models developed recently have shown very high skill in the prediction of ISMR[5,6].

With the rapid development of complex models of the atmosphere and coupled atmosphere–ocean system, and enhanced availability of computer power, in the last decade, it became possible to generate predictions of the monsoon with these dynamical models. These predictions are generated by integrating the governing equations from an initial state prior to the season. For many of us, dynamical models, based on laws of physics, are more appealing than statistical models which are based on the past behaviour of the system. However, until 2005, the skill of monsoon prediction by the dynamical models was too poor[4,7] for them to be useful. Since then, there has been considerable improvement in the skill of the dynamical models[2,8].

As the dynamical models are based on laws of physics, understanding why there are many models and how the models can be improved is important. The resolution of dynamical models is coarse and hence they do not model clouds from first principles. The relationship between clouds and other atmospheric parameters (called convective parameterization) is different in different models. In addition, parameterization of the energy fluxes between the interface of the atmosphere and ocean/land and land surface processes also vary. These parameterizations are empirical (with several tuneable parameters) and can be made more accurate when there are more field observations. During the last 40 years, extensive observations about clouds have been obtained from satellites and hence convective parameterizations have improved. When important facets of the phenomenon of interest are elucidated and theoretical studies lead to more insight into the underlying mechanisms, efforts are made to tune the models to represent the features/mechanisms realistically. In this way, the models can be developed to a level at which they yield better simulation of the phenomenon and higher skill in prediction.

The first breakthrough in seasonal forecasting over the tropics was the development of models that could simulate and predict El Niño Southern Oscillation (ENSO). This was achieved because of the elucidation of the nature of ENSO with data from a new observational system of buoys over the equatorial Pacific and satellite data, and unravelling of the underlying mechanisms[9–11]. As there is a strong link between the Indian monsoon and ENSO with a high propensity of droughts during the warm phase (El Niño)[12,13], it was expected that skillful predictions of ISMR would be generated by dynamical models. However, until recently, the skill of prediction by the atmospheric or coupled general circulation models (AGCMs and CGCMs respectively) was poor. For example, almost all AGCMs in an international Atmospheric Model Inter-comparison project (AMIP) simulated deficit ISMR for the excess monsoon season of 1994 (Figure 2 a). Yet, almost all AGCMs could simulate at least the correct sign of the ISMR anomaly (whether deficit or excess) for the droughts or excess monsoon seasons associated with ENSO (e.g. 1988, Figure 2 b). In addition to variations in ENSO, variations in SST of the Indian Ocean and the winds and rainfall over this ocean
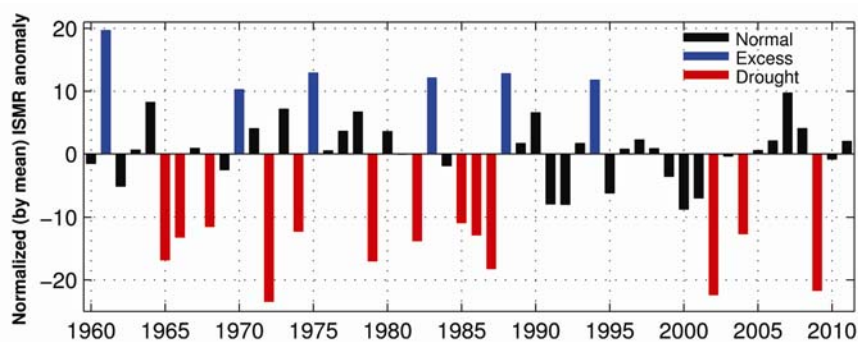
Figure 1. Anomalies (difference between the actual and the average) of the observed Indian summer monsoon rainfall (ISMR) expressed as a percentage of the mean ISMR, during 1960–2011.
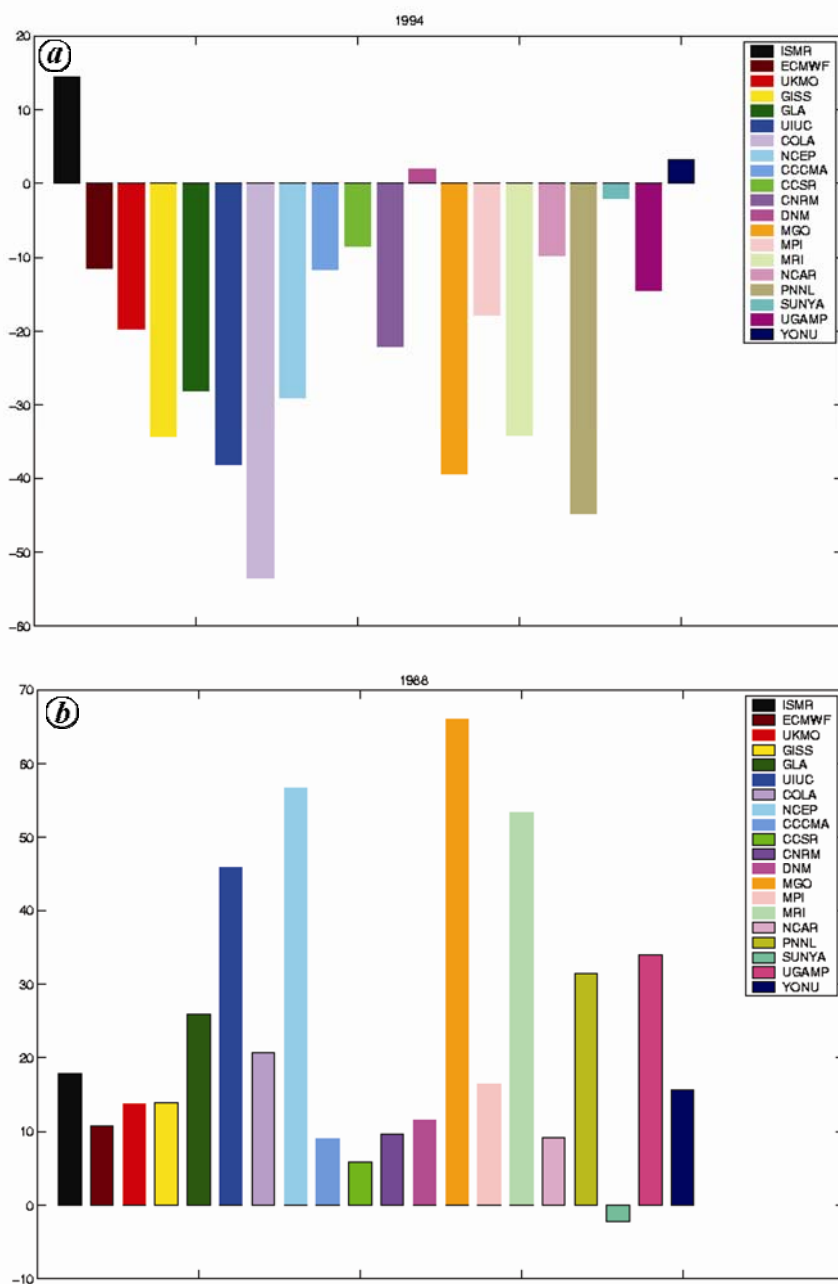
Figure 2. Observed and simulated ISMR anomalies by the AGCMs in AMIP for 1994 (**a**) and 1988 (**b**) respectively.

influence ISMR. The east–west, cold–warm SST anomaly that develop over the equatorial Indian ocean is called Indian Ocean Dipole (IOD)[14]. A strong favourable phase of IOD played an important role in producing excess rainfall in 1994 and in counterbalancing the adverse impact of ENSO in 1997 (ref. 15). Hence, it was suggested that the poor skill of the simulation of ISMR by AGCMs could be attributed to the poor skill in simulating the monsoon–IOD relationship[16].

In order to estimate the skill of coupled atmosphere–ocean models, several models have been run for 1960–2005 under the European Union project ENSEMBLES, with initial conditions specified from the observations for each year to generate 'retrospective' predictions, following a similar exercise conducted earlier under the DEMETER project for 1960–2001. Analysis of DEMETER predictions by Preethi et al.[7] showed that the skill in prediction of

ISMR as well as IOD was rather poor. The analysis of predictions of the ENSEMBLES project by Rajeevan et al.[8] showed that the correlation between the ISMR predicted by the multi-model ensemble and the observed ISMR was 0.45 and much larger than the correlation coefficient of only 0.28 for 1960–2001 for the multi-model ensemble of DEMETER. It is important to note that although there has been improvement in the skill of the climate models, the correlation of the predictions of the multi-model ensemble with observation of 0.45 implies that only about 20% of the observed variance is explained. It was found that most of the coupled models capture the monsoon–ENSO interaction, but not the monsoon–IOD interaction. Hence, most models simulate drought condition in association with the strong El Niño of 1997 when the ISMR was normal.

Delsole and Shukla[2] have also analysed retrospective predictions from the ENSEMBLES project and shown that the skill of the models initialized in May, in predicting ISMR, is statistically significant. They have further shown that the correlation of the predictions of the multi-model ensemble with observations for 1960–2005 is much higher than the correlation with observations (0.15–0.21) of statistical models derived from the SSTs of May for this period. A major conclusion of the study is the attribution of the superior skill of the dynamical models relative to statistical models based on May SSTs, to more skillful predictions of June–September SSTs by the climate models. However, it is not clear that this superior skill in prediction of the monsoon season SSTs is the most important factor leading to the relatively high skill of the coupled models for monsoon prediction. If that were so, then one would have expected atmospheric models forced with observed SSTs to have a very high skill in simulating the interannual variation of ISMR, which is contrary to the poor skill revealed in the assessment of such simulations in a large number of studies[17–19].

Dynamical and statistical models

It should be noted that Delsole and Shukla's[2] conclusion regarding superiority of dynamical models is based on comparison with one specific set of the statistical models, viz. those based only

on the May SSTs. Rajeevan et al.[5] have developed new statistical models based on the ensemble multiple linear regression and projection pursuit regression techniques, which used new methods of predictor selection and model development. The model performance was evaluated for the period from 1981 to 2004, by sliding the model training period with a window length of 23 years. The correlation of ISMR with the predictions of these models is very high, ranging from 0.78 to 0.88 (explaining about 60% of the variance), which is much higher than that with the prediction by dynamical models (which is around 0.46 or less for the different models/ensembles of models, i.e. explaining about 21% of the variance or less). The performance of dynamical models in predicting the recent droughts has not been satisfactory. Thus, none of the dynamical models predicted deficit rainfall for the severe droughts of 2002 (ref. 4) and 2009 (ref. 20). However, a statistical model based on SSTs developed by Sahai et al.[21] could predict deficit in 2002 and the multi-model ensemble of statistical models developed by Rajeevan et al.[5] predicted a drought for 2009. A new statistical model based only on the past rainfall data was recently proposed by Iyengar and Raghukanth[6]. This model was trained using ISMR data for the period 1872–1990 and tested for the period 1991–2002. They showed that the correlation for the forecast period 1991–2002 was 0.91. It appears, therefore, that operational forecasts will have to be based primarily on statistical model predictions until the skill of dynamical models improves to a level which is higher than that of the best statistical models.

For the summer monsoon season of 2012, the India Meteorological Department has predicted ISMR to be 96% of the average[22] on the basis of statistical models, whereas the Indian Institute of Tropical Meteorology has predicted 104%

on the basis of a state-of-the-art coupled ocean-atmosphere model[23]. The performance of the monsoon in the remaining half of the season, will determine which of the two predictions turns out to be closer to the observed.

We do expect the skill of dynamical models in simulating the interannual variation of the Indian summer monsoon to improve in the near future. For this, a focused effort on improving the skill of the models in simulating the important links of the monsoon to events over the Indian and Pacific Ocean is essential. For improving the parameterizations, data from satellites, special observational studies of the Indian Climate Research Programme, new platforms over the Indian seas such as buoys and ARGO floats have to be used. For example, the observations made during the Bay of Bengal experiment (BOBMEX)[24] revealed that the relationship of the evaporation over the Bay of Bengal to the surface wind was different from that over other tropical oceans such as the West Pacific. This is yet to be incorporated in the parameterization. The national monsoon mission provides a unique opportunity for achieving this in a time-bound manner.

1. Gadgil, Sulochana and Gadgil, Siddhartha, *Econ. Pol. Wkly*, 2006, **XLI**, 4887–4895.
2. Delsole, T. and Shukla, J., *Geophys. Res. Lett.*, 2012; http://dx.doi.org/10.1029/2012GL051279
3. Predicting the Indian Monsoon, in Research Highlights. *Nature*, 2012, **484**, 291.
4. Gadgil, Sulochana, Rajeevan, M. and Nanjundiah, R., *Curr. Sci.*, 2005, **88**, 1389–1400.
5. Rajeevan, M., Pai, D. S., Anil Kumar, R. and Lal, B., *Clim. Dyn.*, 2007, **28**, 813–828; doi: 10.1007/s00382-006-0197-6.
6. Iyengar, R. N. and Raghukanth, S. T. G., *Meteorol. Atmos. Phys.*, 2004, **90**, 17–36.
7. Preethi, B, Kripalani, R. H. and Krishna Kumar, K., *Clim. Dyn.*, 2010, **35**, 1521–1539.
8. Rajeevan, M., Unnikrishnan, C. K. and Preethi, B., *Clim. Dyn.*, 2012, **38**, 2257–2274; doi:10.1007/s00 382–011–1061–x.
9. Rasmussen, E. M. and Carpenter, T. H., *Mon. Weather Rev.*, 1982, **110**, 354–384.
10. Cane, M. A., Zebiak, S. E. and Dolan, S. C., *Nature*, 1986, **321**, 827–832.
11. Philander, S. G. H., *El Nino, La Nina and the Southern Oscillation*, Academic Press, New York, 1990, p. 293.
12. Sikka, D. R., *Proc. Indian Acad. Sci.* (*Earth Planet. Sci.*), 1980, **89**, 179–195.
13. Rasmusson, E. M. and Carpenter, T. H., *Mon. Weather Rev.*, 1983, **111**, 517–528.
14. Saji, N. H., Goswami, B. N., Vinayachandran, P. N. and Yamagata, T., *Nature*, 1999, **401**, 360–363.
15. Ashok, K., Guan, Z. and Yamagata, T., *Geophys. Res. Lett.*, 2001, **28**, 4499–4502.
16. Gadgil, Sulochana, Rajeevan, M. and Francis, P. A., *Curr. Sci.*, 2007, **93**, 182–194.
17. Gadgil, Sulochana and Sajani, S., *Climate Dyn.*, 1998, **14**, 659–689.
18. Sperber, K. R. and Palmer, T. N., *J. Climate*, 1996, **9**, 2727–2750.
19. Wang, B., Ding, Q., Fu, X., Kang, I.-S., Jin, K., Shukla, J. and Doblas-Reyes, F., *Geophys. Res. Lett.*, 2005, **32**, L15711; DOI:10.1029/2005Gl022734.
20. Nanjundiah, R. S., CAOS Report No. 2009 AS 02, 2009.
21. Sahai, A. K., Grimm, A. M., Satyan, V. and Pant, G. B., *Clim. Dyn.*, 2003, **20**, 855–863.
22. Press release on Long Range Forecast, June 2012; http://www.imd.gov.in/section/nhac/dynamic/
23. Goswami, B. N., *Yojana*, July 2012, vol. 56, pp. 32–34.
24. Bhat et al., *Bull. Am. Meteorol. Soc.*, 2001, **10**, 2217–2243.

Sulochana Gadgil* and J. Srinivasan[†] are in the Centre for Atmospheric and Oceanic Sciences, [†]Also in the Divecha Centre for Climate Change, Indian Institute of Science, Bangalore 560 012, India. *e-mail: sulugadgil@gmail.com