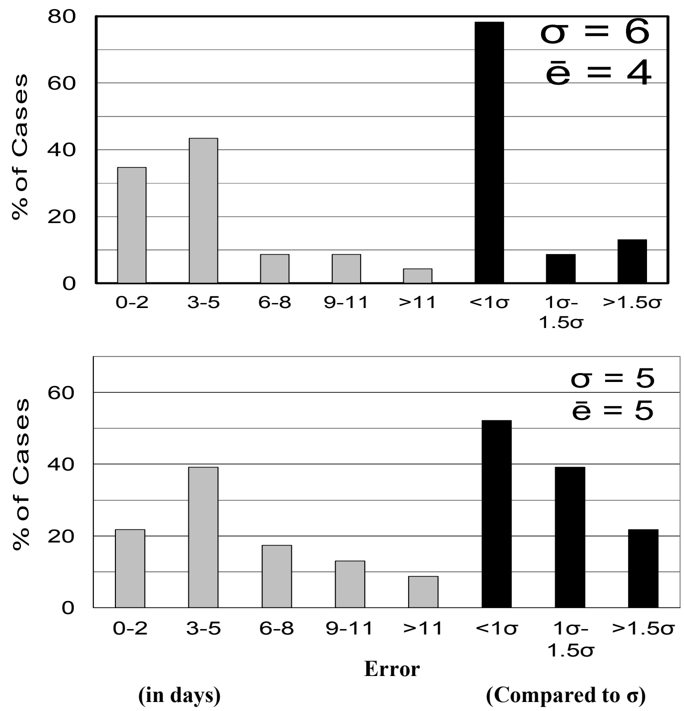
C-MMACS

Advance Forecasting of the Date of Onset of Monsoon:
Dynamical Basis, Model Configuration and Skill Evaluation

Histogram of error between predicted & observed date of onset



P Goswami and K C Gouda

CSIR Centre for Mathematical Modelling and Computer Simulation
Wind Tunnel Road, Bangalore-560 037, India

Research Report RR CM 0716
October 2007

Publication Status: Communicated
Distribution: Unrestricted

CSIR Centre for Mathematical Modelling and Computer Simulation
Wind Tunnel Road, Bangalore-560 037, India
Tel No: +91 80 25224667 Fax +91 80 25220392
Website: http://www.cmmacs.ernet.in
E-mail: goswami@cmmacs.ernet.in

Advance Forecasting of the Date of Onset of Monsoon Rainfall over India

P Goswami and K C Gouda

Abstract

The onset of Indian summer monsoon (ISM) represents one of the most dramatic transitions in the regional circulation pattern. The onset also marks the beginning of the main rainy season for India; advance and accurate forecast of the day of the onset of monsoon (DOM) thus has application in many sectors. Advance dynamical forecasting of DOM, however, is rarely attempted due to the poor skill of most GCM in predicting ISM rainfall. A primary cause for poor skill in forecasting parameters like rainfall appears to be the loss of predictability due to noise introduced by local synoptic processes. However, sharp transitions in the regional circulation pattern and associated rainfall, which are likely to be less affected by synoptic noise, may have higher predictability, somewhat similar to the way that monthly mean parameters are more predictable. We explore this premise for advance forecasting of onset of ISM over Kerala and show that significant skill is possible in advance forecasting of DOM. We use a global circulation model (GCM) with a special feature, variable resolution, to meet the special requirements of forecasting DOM. Based on a set of objective and validated criteria, hindcasts of DOM are generated in complete operational setting from a 5-member ensemble for each year for the period 1980 to 2003. The hindcasts are evaluated in terms of a number of parameters; the mean error with respect to observations in the 24 ensemble average forecasts is 4 days, which implies high skill. Implications of the results for predicting certain weather and climate processes are discussed.

Introduction

The inherent limits to predictability of atmospheric processes due to chaotic nature of the underlying dynamics are well known (Lorenz 1965; Webster et al. 1998; Sperber and Palmer 1996; Molteni and Palmer 1993). This is particularly true for prediction of precipitation, especially at range longer than a few days, which still remains elusive in spite of advances in modelling, observation and computing. One reason for lack of predictability of rainfall is the synoptic noise (Moron et al. 2006). However, it is known that certain quantities, like the monthly means, have higher predictability (Shukla 1981). Another class of phenomena where synoptic variability is expected to play a relatively less dominant role is processes involving sharp, large-scale transitions which are primarily driven by regional circulation; these processes thus may have higher predictability. One such phenomenon that represents a large-scale, dramatic shift in the regional circulation system is the onset of the Indian summer monsoon (ISM). A question that needs careful analysis is whether such transitions have higher predictability than for usual (day to day) monsoon variability. It is, of course, well known that most current GCM have only limited skill in predicting monsoon rainfall (Kang et al. 2002; Sperber and Palmer 1996) beyond certain climatological features. In this backdrop, any attempt to forecast DOM more than a few days ahead may appear unreasonable. However, as mentioned above processes like sharp, large-scale transitions which are primarily driven by regional circulation may have higher predictability as the synoptic variability that reduces the predictability (Moron et al. 2006) is likely to be overwhelmed by the large signal associated with transition process.

The abrupt and large-scale shift in the regional circulation pattern associated with the onset of ISM and its socio-economic implications have been emphasized in several studies (Soman and Kumar 1993; Gadgil 2003). The onset also involves interaction of a number of processes at different scales (Lau et al. 1998) and represents significant transitions in the large-scale atmospheric and ocean circulations in the Indo-Pacific region (e.g., Rao 1976; Murakami and Nakazawa 1985; Lau et al. 1998; Hsu et al. 1999; Wu and Zhang 1998). Near India, the onset occurs initially across the peninsula's southern tip in late May to early June, progressing north-westward across most of the country in the following month. The northward progression of the monsoon onset signifies a large-scale transition of deep convection from the equatorial to continental regions (e.g., Rao 1976; Sikka and Gadgil 1980; Webster 1983; Webster et al. 1998). Similarly, several studies have emphasized the roles of large-scale interactions between surface heating and atmospheric dynamic, thermal, and hydrologic processes in monsoon transitions (Takagi et al. 2000; Hsu et al. 1999; Kumar et al. 1997; Ueda and Yasunari, 1998; Wu and Zhang 1998; Webster 1983).

Though a variety of dynamic and thermodynamic precursors have been identified (e.g., Ananthakrishnan and Soman [hereafter AS] 1991; Murakami and Nakazawa1985), one index of the large-scale transition in the regional circulation associated with the onset of monsoon is the characteristic change in the rainfall over Kerala. While there exists no unique definition, at the surface the onset is recognized as a rapid, substantial, and sustained increase in rainfall over a large scale; typically, from below 5 to over 15 mm day^{-1} during onset (AS 1988; Soman and Kumar 1993). The suddenness of rainfall fluctuations during the monsoon's transitions has been emphasized in several studies (e.g., Rao, 1976; AS 1988; Wang and Lin Ho 2002; Pearce and Mohanty 1984). Based on Kerala rainfall, the mean date of onset of monsoon (DOM) is near 1 June and varies with a standard deviation of about 7 days. Another feature that characterizes the onset is spatial coherency over a large scale, which is uncharacteristic of synoptic variability. Although the extent to which rainfall at Kerala during these transitions is determined by synoptic variability unrelated to the monsoon transitions is not well known, the large-scale and sustained component of rainfall associated with onset can be considered as a result of the large-scale processes. The prediction of DOM thus should be viewed as the prediction of such a transition rather than that of daily synoptic variability of rainfall. Besides, many earlier models didn't have the resolution to capture the dynamics over the relatively small onset domain (Kerala, about 200 Km east-west) in a global (GCM) environment. For reasons given above, it is worth exploring the forecast skill for DOM, even though skill for predicting daily rainfall in general may be low. Further, a prediction of onset only requires forecasting rainfall above a threshold (transition) rather than forecast of precise values of rainfall.

Inherent to a prediction of DOM, of course, is its (objective) definition. A number of techniques have been developed to identify monsoon onset. Some methods focus on understanding the climatological mean date of onset by region. Two methods that identify the onset date include the objective method and the more subjective declarations of the India Meteorological Department (IMD). However, large disagreements can arise in comparing objective and subjective assessments of onset. For example, in 1969 IMD's declaration of onset on 17 May disagrees with the objective classification of AS by 8 days. In 1959, the disagreements are 19 days and in 1943 and 1932 the disagreements are 17 and 19 days, respectively. Moreover, other years, such as 1979 and 1995, are associated with bogus onsets that objective methods can misdiagnose by up to three weeks (Webster et al., 1998). A challenging task is to devise a method to determine DOM that can discern the occurrence of bogus monsoon onsets. Besides, the conventional set of criteria, based on isolated station observations, are not best suited for determining DOM from (gridded) model data; an objective, calibrated set of criteria for determining DOM from model simulations is thus needed.

The objective of this work is thus two fold: we first show that a set of objective criteria can be formulated to compute the DOM from daily gridded rainfall observations prepared by IMD. As shown in section 2, these criteria encompass the conventional criteria for DOM based on rainfall, but have extended scope to avoid bogus onset and easy applicability to model forecasts. In particular, our criteria, for the first time, include post-onset persistence of rainfall to enhance reliability of the predicted DOM. The second objective of our work is to show that DOM can be predicted with a high degree of accuracy using a (variable-resolution) GCM. We begin with a description of the set of objective criteria derived from consistency between DOM announced by IMD and computed from daily gridded rainfall data in section 2. Model configuration and results on advance prediction of DOM are presented in sections 3 and 4, respectively. The last section contains our conclusions and perspectives.

Definition and Calibration of Objective Criteria: Coverage & Persistence

While there exists a large number of definitions and criteria to identify DOM (Webster 1983; Ananthakrishnan and Soman 1991; Soman and Kumar 1993; Joseph et al. 2006), a carefully formulated and calibrated set of criteria are needed for objective forecasting of DOM with a GCM. As there have been various approaches to define onset in the past (Ananthakrishnan and Soman 1991; Joseph et al. 2006), a clear and precise definition of DOM in our work is necessary. In particular, the focus of our study is monsoon over continental India, and the domain of onset is the western coast of Kerala (8-12^{o}N, 75-77^{o}E). Further. We calibrate and use a set of criteria based on a single variable to define the onset; although a number of variables have been proposed, these are dynamically related and mutually consistent. Based on both availability of long-period (gridded) observation and significance for users we choose this variable to be rainfall characterized by persistence, significance and coverage.

The biggest challenge in predicting monsoon onset, perhaps, is to avoid what are called "false" or "bogus" monsoon onsets, which are associated with propagating tropical intraseasonal disturbances unrelated to the monsoon onset (Flatau 2001, Joseph et al.1994). The disturbances are characterized by an enhancement of convection and westerly surface winds similar to the monsoon onset but occurring over a smaller scale and lasting a week or less. Often, bogus onsets are followed immediately by extended periods of weak winds and clear skies that result in heat waves and droughts in India. These droughts can cause considerable economic and agricultural damage when incorrectly predicted, as crops planted in anticipation of the monsoon are likely to fail as a result of the bogus onset. As bogus onsets can predate the actual onset by up to a several weeks, it is also important for any retrospective onset determination to be insensitive to their occurrence.

One way to avoid bogus onset is to accurately estimate the rainfall **subsequent** to the (assumed) date of onset (post-onset rainfall). This is possible through forecasts; however, the limited skill of numerical models in forecasting monsoon rainfall is well known (Sperber and Palmer 1996; Webster et al.1998), and there has not been much attempt at forecasting monsoon onset with long lead. The predictability of a system like the monsoon, however, is a complex issue, and depends on a host of processes and parameters, from model resolution to parameterization of convective processes. As mentioned earlier, it is possible that a suitably calibrated configuration will succeed to capture the process of onset which represents a very sharp change in the regional circulation pattern, and thus less sensitive to small-scale, transient variability that normally make such forecasts difficult.

The objective criteria adopted by us are based on four parameters which can be applied to model forecasts: these parameters are pre-onset persistence (PrOP), significance, spatial coverage and post-onset persistence (PoOP). It may be noted that the criteria of post-onset persistence, unlike the other three, can not be obtained from observations, and must be derived from model predictions. The PrOP and PoOP together ensure that the rainfall predicted is sustained, and not a result of a transitory system. The significance, taken in terms of rainfall above a threshold value, is once again a characteristic of the monsoon. However, the threshold for significance in observed (gridded) data and model predictions are not necessarily the same due to the model bias. Even for observed data, the threshold for isolated station observation and that for gridded rainfall may have to be different as for the latter the values at grid points will be affected by the smoothing and the interpolation processes. This parameter will thus need careful and independent calibration for gridded observation and model simulations. The spatial coverage, once again, ensures the large scale nature of the monsoon rainfall. While conventional announcement of DOM is based on station rainfall, this procedure suffers from the intrinsic and significant spatial variability of the distribution of monsoon rainfall from year to year. In particular, it is possible to have relatively high or low rainfall over a few stations without the characteristic large-scale coverage. An appropriate value for the spatial coverage to characterize large-scale nature of onset thus has to be determined. We have based the measure of large-scale nature of the onset in terms of percentage of spatial coverage of the onset domain (8-12^{o}N, 75-77^{o}E); this avoids fixed locations (stations) for determining onset and automatically incorporates the spatial variability inherent in monsoon rainfall.

The primary requirement for such an analysis with observations, a high-resolution gridded rainfall dataset, was only recently met with the availability of a 53-year (1951-2003) daily rainfall data on a $1^{o} \times 1^{o}$ degree prepared by the India Meteorological Department (IMD). The IMD dataset is based on rainfall records of 1803 stations which had a minimum 90% data availability during the analysis period (1951–2003). The station rainfall data have been projected onto a rectangular grid ($1^{o} \times 1^{o}$) for each day for the period 1951–2003. In this gridding method, the interpolated values are computed from a weighted sum of the station observations. The starting point of the grid is 6.5°N and 66.5°E. From this point, there are 35 points towards east and 32 points towards north. the methodology and analysis based on this data set are available in a number of works (Rajeevan et al. 2006; Ramesh and Goswami 2007)

The calibration of the values of the four parameters is carried out so that the statistics of the computed DOM from gridded daily rainfall data match with those of the announced DOM within margins of error inherent in the announced DOM.

Table 1 : Dates of onset of ISM computed from gridded daily rainfall from IMD for different area coverage and thresholds, average date of announced onset for period 1980-2003 is May 31 with a standard deviation of 5 days; the pre onset rainfall persistence taken is 3 days.

Post onset Persistence (Days)	Area Coverage (%)	Threshold Rain (mm/day)	Mean Onset Date	Values of σ (Days)	Mean Absolute Error (Days)
3	20	3	31	7	3
		4	32	8	3
		5	32	8	3
	30	3	33	7	3
		4	35	7	4
		5	36	8	6
	40	3	36	7	5
		4	37	7	7
		5	38	7	7
	50	3	41	7	10
		4	42	6	11
		5	43	6	12
5	20	3	32	7	3
		4	33	7	3
		5	33	7	3
	30	3	37	7	6
		4	37	8	6
		5	38	7	7
	40	3	37	7	6
		4	38	7	7
		5	39	7	8
	50	3	42	5	10
		4	42	6	11
		5	43	6	12

The calibration, based on multiple choices of each of the four parameters, presented in Table 1. The results in table 1 are based on a pre-onset persistence of daily rainfall of 3 mm/day. Although the daily rainfall over a station during the onset is generally much higher than 3 mm/day, we adopt it as the threshold value of for two reasons: first, the grid-point value of rainfall is expected to be much smaller than station observation due to the smoothing involved; secondly, 3 mm/day is the value adopted by IMD to announce as a day of significant rainfall.

It may be seen that for post-onset persistence of 3 days and spatial coverage of 30%, the DOM computed from the daily gridded data has similar statistical characteristics to the announced date of onset. Year-wise comparison of dates of onset from daily gridded data based on the above criteria with the announced dates of monsoon (Figure 1.a) shows excellent agreement, with a correlation co-efficient of 0.87 and a root mean square error of 3.8 days which is about half the value of the standard deviation in the announced dates of onset.

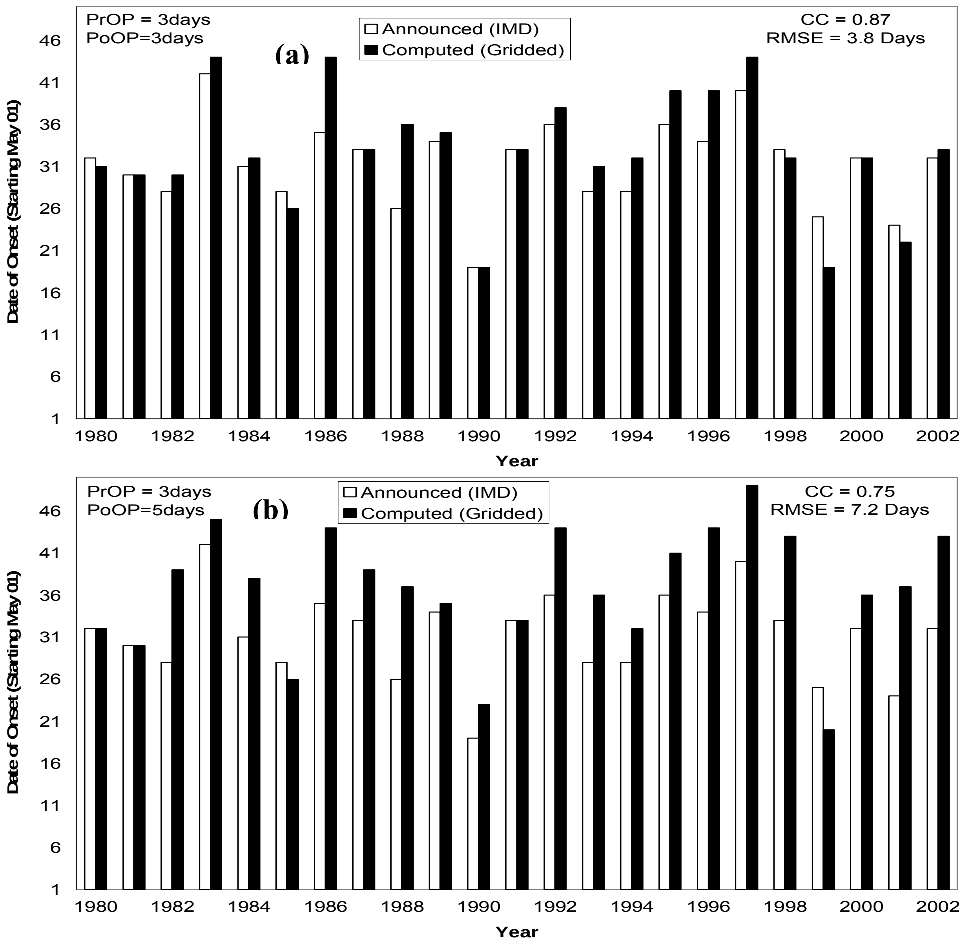


Figure 1: Comparison of dates of onset of ISM computed from gridded daily rainfall data from the India Meteorological Department with the announced dates of onset. The computed dates are based on criteria of 3 day pre-onset persistence and (a) 3day and (b) 5 day post-onset persistence of rainfall above 3 mm/day, with area coverage of 30% over the domain of onset (Kerala, 8-12°N, 75-77°E).

For our later discussions we note that a spatial coverage of 30% but a PoOP of 5 days the mean difference between the computed and the announced dates of onset is 6 days (figure 1.b) with a maximum difference of 11 days. In figure 1 the years marked with * signify years with false onset, while the years marked with open circles are El Nino years. As can be seen from the figure the errors in the years with El Nino or false onset are not particularly high. The issue of calibrating a spatial coverage characteristic of onset is further addressed by considering error (difference) between announced and computed DOM for different % coverages (figure 2) of the onset domain by daily rainfall above significant level. It may be seen that while the error does not change significantly until % coverage of 30%, there is a sharp increase beyond it. With our emphasis on the large-scale nature of onset, we shall therefore adopt an area coverage of 30% of the domain to be optimum value for characterizing onset. Although coverage of only 30% may appear small to be used as a characteristic of large-scale, it should be noted that this coverage is being considered on a daily basis, only over the continental onset domain (8-12°N, 75-77°E). Thus, although the changes associated with the onset may be observed over a wide domain, the continental coverage of onset rainfall over Kerala is about 30% of the domain. Further, the value of 30% area coverage should be considered as a calibrated

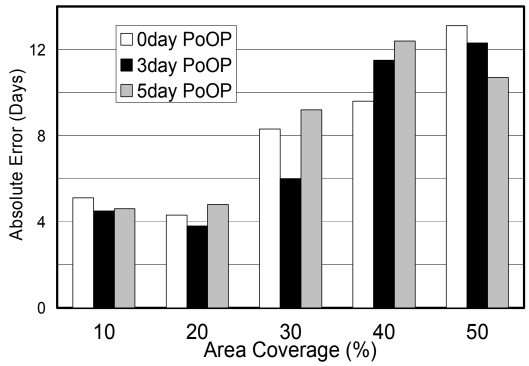


Figure 2: 53-year (1951-2003) average absolute error in computed DOM for different area coverage. The hollow, solid and shaded bars represent, respectively, computed dates based on criteria of 0 (no post-onset persistence), 3 and 5 day post-onset persistence of (threshold) rainfall above 3 mm/day. The Pre-onset persistence is 3

value for optimum agreement between the announced and computed DOM.

Forecast Configuration and Methodology

Although we expect the transition represented by the onset of ISM to have higher predictability, the methodology and the forecast configuration used to predict such transitions should, however, fulfil certain necessary requirements. It is critical to use a forecast model with sufficiently high horizontal resolution to capture the dynamics over the relatively small onset domain (Kerala, about 200 Km east-west). Several studies have shown the importance of horizontal resolution in the quality of simulation (Kentarchos et al. 2000; Boyle 1993). At the same time, the onset is a result of interaction of a number of processes at different scales (Murakami and Nakazawa, 1985; Pearce and Mohanty, 1984; Hsu et al. 1999) and represents significant transitions in the large-scale atmospheric and ocean circulations in the Indo-Pacific region. It is therefore necessary to use a forecast configuration that can allow global dynamics (large spectrum of scales) at sufficiently high resolution, especially over the onset domain. Although limited area models can support high resolution, the presence of artificial lateral boundary conditions, and the resulting absence of larger (than domain) scales in the dynamics in these models (except through lateral boundary conditions) limits their utility for investigating predictability of a process like monsoon transition. Further, although a shorter range of forecast allows less growth of numerical error, an optimum lead is necessary to allow the model dynamics to capture the evolution of circulation that results in the transition. In particular, the announced DOM covers a period from early May to late June; thus it is necessary to allow the forecast period to include a period long enough to cover this period. We have therefore considered a forecast period of May-June in our study. However, simulations with high resolution with a GCM for long period and a large number of cases to build up a meaningful statistics is an expensive proposition.

We use a GCM with stretched coordinates (variable resolution with zoom) to meet the above requirements in a computationally feasible manner. The variable resolution, discussed in a number of studies allows relatively higher resolution (zoom) over a chosen domain in a continuous and dynamically consistent manner. The basic formulation and validation of the model adopted here (version LMDZ3 developed at LMD, France), including formulation of variable resolution, have been described in detail in earlier works (Sadourny and Laval 1984; Sharma et al. 1987; Laval et al. 1996; Sabre et al. 2000). The variable resolution is prescribed as a continuous variation (stretching) in the resolution with respect to a point (zoom). The variable-resolution, however, offers more than high resolution; in a complex system like a GCM there is no clear separation between numerics, dynamics and physics and a change in the grid can affect a change of strength and distribution of convergence, and hence precipitation, over those from a uniform resolution grid. In the version used in this study, the model grid has 192 X 144 points in the horizontal direction with a zoom factor of 4 which provides a resolution of about 50 Km over the onset domain with the highest resolution near the centre of the zoom which merges smoothly to about 2° x 1.25° away from the zoom. The model physics include a diurnal cycle, a land surface module and convective parameterization scheme of Tiedtke (1989). The number of vertical levels is 19.

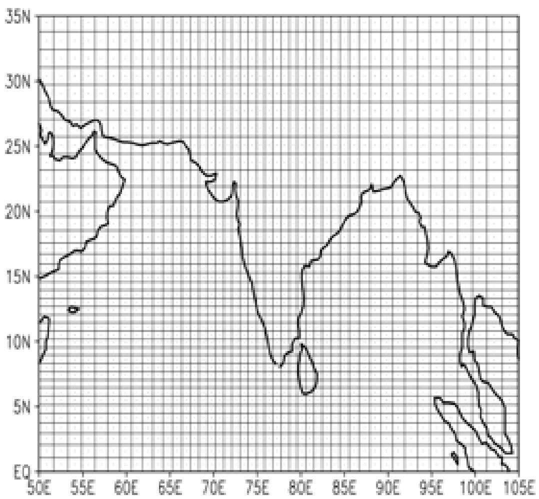


Figure 3a: Part of the structure of the onset grid with the centre of zoom at 75°E and 10°N. The total number of grid points is 192 × 144 in longitude and latitude; the highest resolution near the centre of the zoom is ≈ 50 km.(variable) which merges to a uniform 2° ×1.25° away fro zoom (stretched co-ordinate).

The initial conditions were adopted from the daily fields of NCEP Reanalysis (Kalnay et al. 1996). While dates of initial conditions were chosen to provide a lead of at least 30 days, the simulations were carried out until the end of June to allow late onset as well as to examine the model's performance in terms of stability and post-onset performance. As the simulations are carried out in a forecast setting, the SST field must be either dynamically generated or prescribed, for example from climatology. The best option in terms of dynamical consistency is to use a coupled ocean-atmosphere model. However, the current skill of almost all coupled models is well below acceptable level. Our strategy here is to adopt a (monthly) climatology of SST field from AMIP (Atmospheric Model Inter comparison Project) data set along with climatological fields of vegetation and sea ice from NCEP Reanalysis.

The variable-resolution grid was calibrated for optimum performance for simulating the process of onset. As a part of the optimization, we had considered a number of grids, including a grid of uniform resolution of 2° X 1.25°, a variable grid

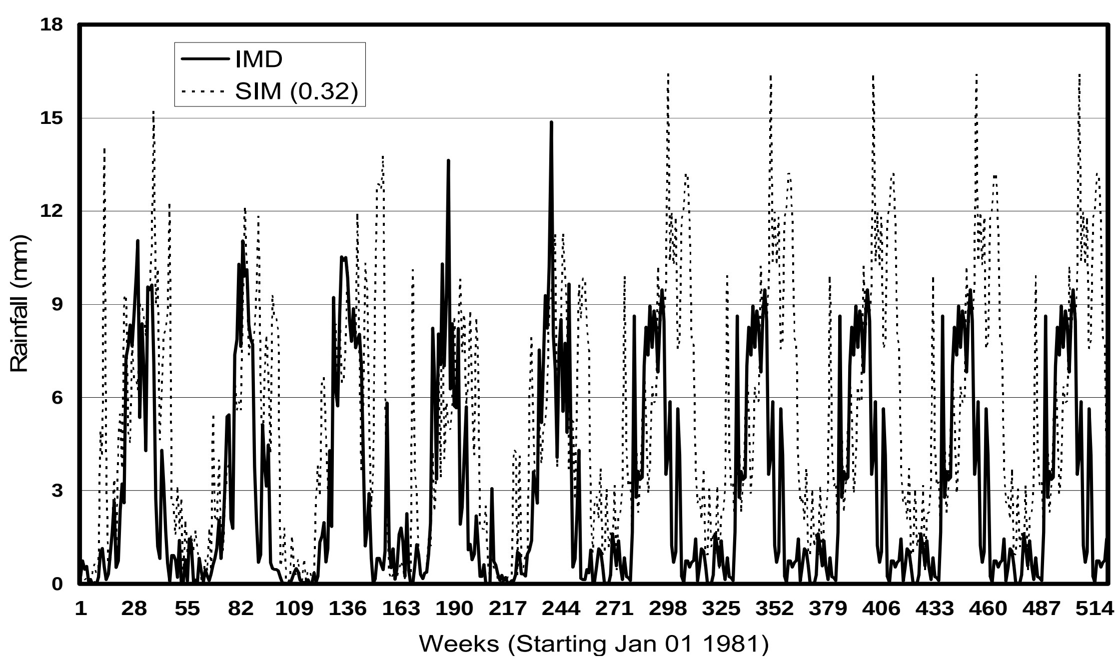


Figure 3b: Annual cycle of area-averaged weekly rainfall for 10 years (1981-1990) from IMD (thick solid line) and GCM Zoom Grid with two-member (January 1 and May 1) ensemble initial conditions (dash line).

with zoom over 15°N ,75°E and a variable grid with zoom over 75°E; 10°N (onset grid). Figure 3a shows the part of the structure of the onset grid with the centre of zoom at 10°N, 75°E. Forecasts with a given initial condition (May 01) were then generated for all the 24 years for each of the grids. The variable-resolution configuration was extensively tested for conservation properties and stability; figure 3b shows the results from a 10-year ensemble (winter + spring) simulation in terms of area averaged (8-28°N, 75-85°E) weekly rainfall. The thick line in figure 3b represents corresponding, observed rainfall from IMD; as can be seen from this figure, the model configuration reproduces the observed features well with no noticeable drift.

Hindcast skill and error statistics

All hindcast experiments, including construction of climatology, were carried out in a completely forecast setting (that is without assuming any observed information beyond the day of initial condition) for the years 1980-2003. For comparison with observed IMD data, the forecast fields were first projected to the common IMD (1° X 1°) grid.

As the onset process takes place during the late May to early June, we have examined the model's performance with the onset grid in terms of 24-year climatology (5-member ensemble average) of rainfall separately for the periods 15-31 May and 1-15 June. The spatial distribution of difference in observation and simulated climatology showed the bias to be within ±3 mm/day over most locations over the onset domain (figure 4), this bias is less than or comparable to the threshold value (3mm/day) used for significance. A comparison of 24-year climatology of daily rainfall area-averaged over (8-12°N; 75-77°E), however, showed the model predictions for each of the grids, especially for May rainfall, to have certain uniform positive bias (figure 4). This bias

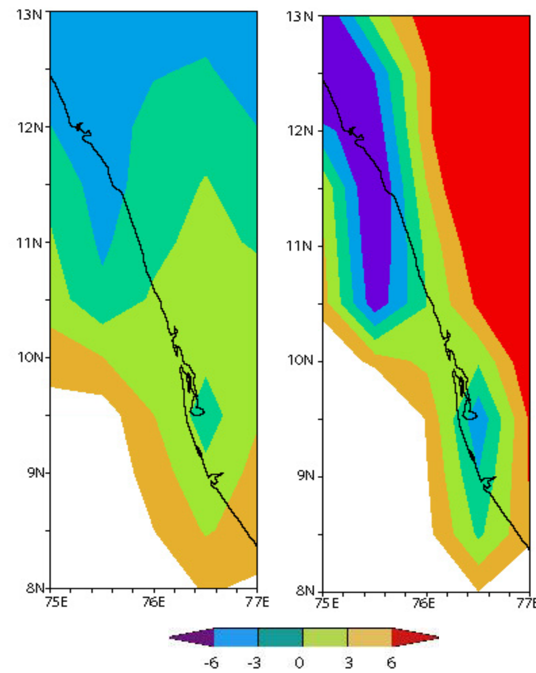


Figure 4 : 24-year average bias (with respect to corresponding climatology from IMD gridded daily rainfall) of rainfall for the periods 15-31 May (left panel) and 1-15 June (right panel). In each case, the simulated climatology represents ensemble average of 5-member simulations for 24 years with initial conditions taken between April 27 to May 1 of the respective year.

was found to be 10 mm/day for the uniform grid and 8 mm/day for the variable grids. As the prediction of DOM requires, especially for significance, daily rainfall, predictions from each grid were therefore first debiased by removing the uniform bias in the climatology. Removal of this uniform bias resulted in much better agreement between observation and the prediction (figure 5) and allowed a more objective optimization. With this removal of uniform bias, the correlation coefficients between the observed and predicted daily rainfall climatologies are 0.88, 0.73 and 0.87 for the uniform grid, zoom (15°N,75°E ,) and onset grid (10°N,75°E) respectively. In our subsequent analysis, therefore, we shall remove a uniform positive bias from model rainfall for each grid. Although in terms of climatology of daily rainfall the uniform grid performed as well as the onset grid, a comparison of average (between May 01-June 30) error

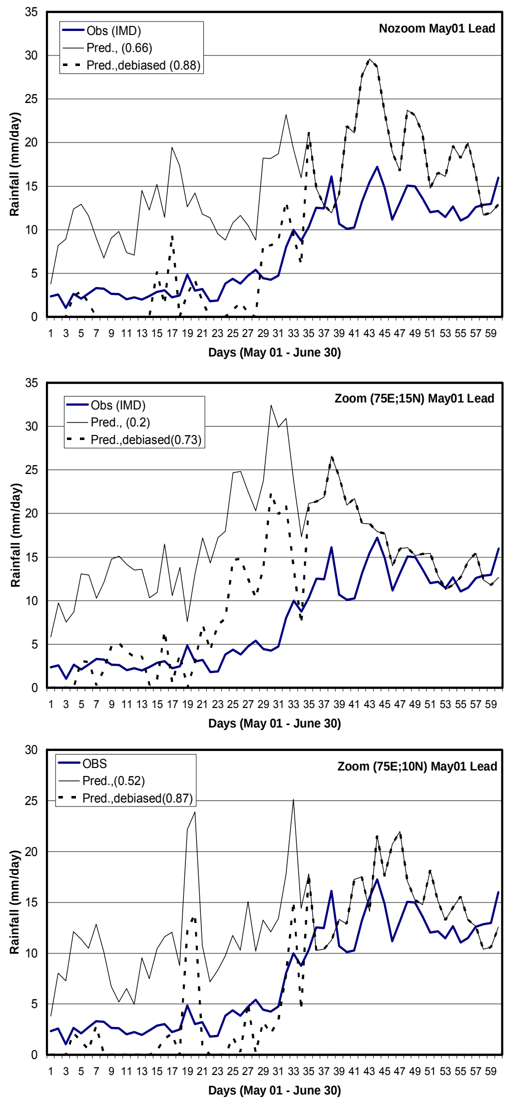


Figure 5: Climatologies (1980-2003) of area-averaged ((75-77^{o}E; 8-12^{o}N) daily rainfall computed from gridded daily rainfall (thick line) and the predicted daily rainfall from the model simulation. In prediction climatologies represent simulations with initial condition May 01. The thin solid line represents unaltered predicted data while the dashed line represents predicted data after removal of a uniform bias for the month of May. The corresponding correlation coefficients with daily observed values are given in the panel. (a) for the uniform grid (b) Variable grid with zoom centre at 75^{o}E,15^{o}N (c) Variable grid with zoom centre at 75^{o}E,10^{o}N(onset grid)

in area averaged (8-12^{O}N; 75-77^{O}E) daily rainfall during May to June period from observation and prediction for 24 years for the three grids showed (figure 6) the onset grid to have generally the lowest error.

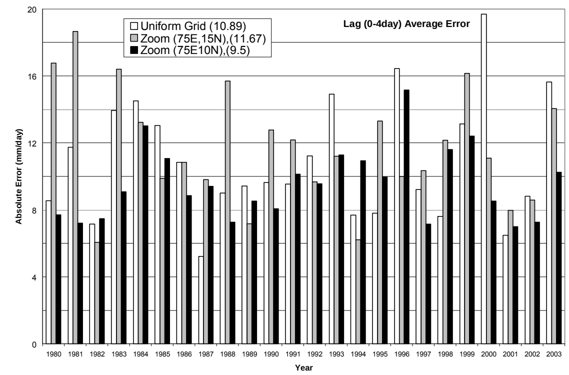


Figure 6: Average of lagged (0-4 days) absolute error between area averaged (75-77^{o}E; 8-12^{o}N) rainfall during May 01-June 30 from observation (IMD) and model forecasts for three grids: uniform grid (hollow bars), grid with zoom at 75^{o}E,15^{o}N (shaded bars) and the onset grid (zoom at 75^{o}E,10^{o}N (thick bars) for 24 years (1980-2003). The corresponding 24 year average absolute errors are given in bracket.

These errors were calculated for 5 lags (0-4 days) to allow for any shift of phase within the limit of the standard deviation (~7 days) in the observed DOM. The lagged average error for 24 years (figure 6) showed the onset grid to have the lowest error. These conclusions are later verified with analysis of forecast error. In our subsequent analysis we shall therefore present results with respect to this optimum onset grid (Figure 3a). Once again, the errors in the years with El Nino (marked with open circles) or in the years with false onset (marked with *) are not significantly or systematically higher than the mean error.

The histogram of errors between the dates of onset and the dates of onset from prediction and computed from the daily gridded IMD data shows (figure 7) the average error to be around 7 days for the ensemble members (panel a-e, figure 7); in contrast the mean error for the ensemble average is only 4 days, showing that it is an effective ensemble average (error in ensemble average less than that in individual forecast). It can be also seen that the ensemble average forecast has the highest number of cases (>75) in the error bin of 3-5 days, and has the lowest standard deviation (σ, 6 days) against 8-

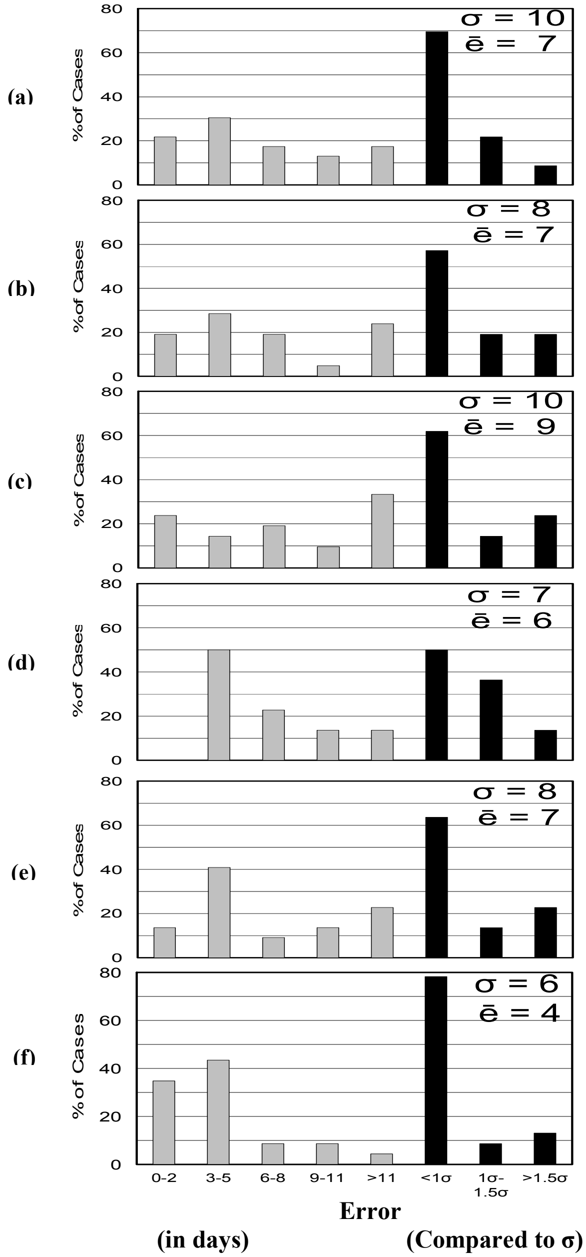


Figure 7: Histogram of errors in predicted dates of onset with respect to dates of onset computed from daily rainfall data from the India Meteorological department. Both the computed and the predicted dates are based on the criteria of significant rainfall with Pre-onset persistence of 3 days and Post-onset persistence of 5 days with area coverage of 30% of the onset domain. The significant daily rainfall is taken to be different for the two cases to allow for model bias: 3 mm (10 mm) for IMD (predicted) data. The six panels from top to bottom represent, respectively, predictions with initial conditions of (a) April 27, (b) April 28, (c) April 29, (d) April 30, (e) May 01 and (f) Average of the 5 leads. The average absolute error ē (days) and the standard deviation σ (days) for each case are given in the respective panel.

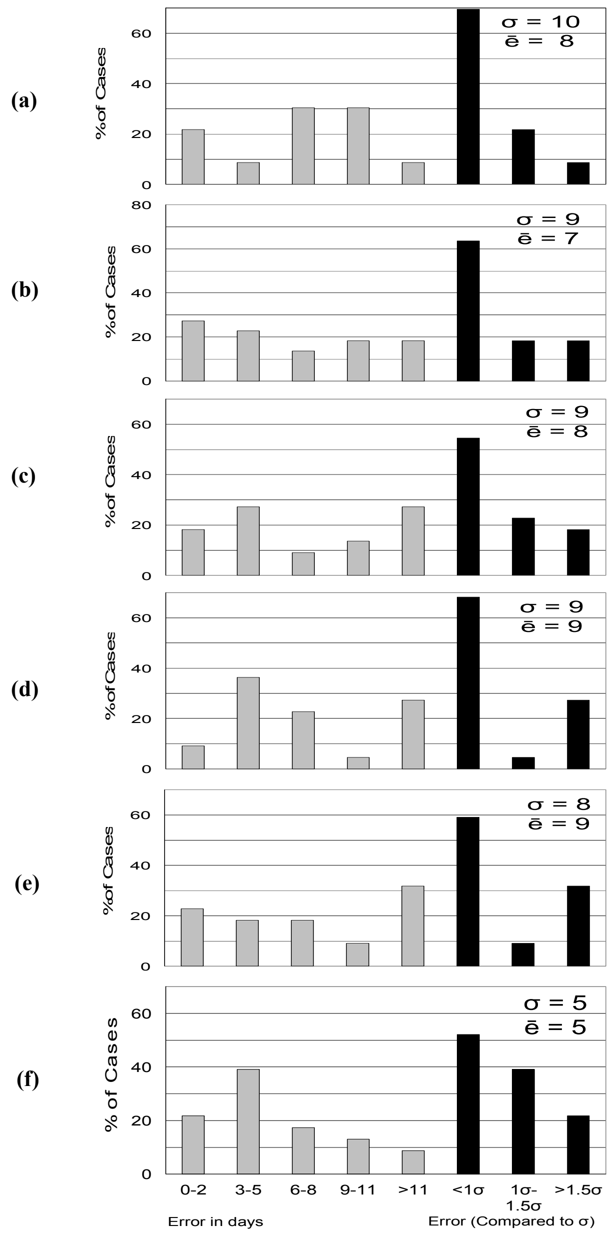


Figure 8: Histogram of errors in predicted dates of onset with respect to dates of onset computed from daily rainfall data from the India Meteorological department. Both the computed and the predicted dates are based on the criteria of significant rainfall with Pre-onset persistence of 3 days and Post-onset persistence of 3 days with area coverage of 30% of the onset domain. The significant daily rainfall is taken to be different for the two cases to allow for model bias: 3 mm (10 mm) for IMD (predicted) data. The six panels from top to bottom represent, respectively, predictions with initial conditions of (a) April 27, (b) April 28, (c) April 29, (d) April 30, (e) May 01 and (f) Average of the 5 leads. The average absolute error ē (days) and the standard deviation σ (days) for each case are given in the respective panel.

Table 2: Summary of standard deviation and absolute error computed using the criteria of 30% area coverage, and pre onset persistence period of 3days.

(a) Post-onset Persistence of Rainfall: 3 days

Lead	Onset Date	σ (Days)	ē (Days)	No of Years with error <1σ (%)
May01	June 3 (34)	8	9	59
Apr30	June 3 (34)	9	9	68
Apr29	May31 (31)	9	8	54.5
Apr28	June3 (34)	9	7	63
Apr27	June2 (33)	10	8	69.5
Ens. avg.	June 2 (33)	5	5	52
σ of onset days	1.3			

(b) Post-onset Persistence of Rainfall: 3 days

Lead	Onset Date	σ (Days)	ē (Days)	No of Years with error <1σ (%)
May01	June 7 (38)	8	7	64
Apr30	June 9 (40)	7	6	50
Apr29	June4 (35)	10	9	62
Apr28	June6 (37)	8	7	57
Apr27	June6 (37)	10	7	62
Ens. avg.	June 3 (34)	6	4	78
σ of onset days	1.8			

10 days in the individual forecasts. In terms of observed standard deviation (natural variability), about 60% of the cases have error less than 1σ, with the highest number of cases (~80%) for the ensemble average forecast. For the criterion of post-onset persistence of 5 days based on our calibration experiments (figure 7), the RMSE between the observed (computed) and the ensemble-averaged predictions DOM is 4 days, with 78% of the cases having error less than one standard deviation. A Post-Onset Persistence of 3 days shows similar results with somewhat less success 60% (Figure 8). A summary of the skill of the forecasts (table 2) for the two values of post-onset persistence for each of the initial conditions as well as the ensemble average revealed a noteworthy feature: the dispersion in the (24-year) mean onset date for the five initial conditions is less than 2 days, implying a low forecast dispersion and a high degree of reliability. We also note that for post-onset persistence of 5 days (but not for 3 days) the ensemble average forecast performs better than any of the individual forecasts.

Concluding Remarks

Though a definition of monsoon onset based on Kerala rainfall is somewhat arbitrary, it has the significance of being the date which signals the beginning of the main agricultural season for a significant fraction of the world's population. While the standard deviation in the observed DOM is only about 7 days, the separation between earliest and the latest observed dates of onset is more than 30 days. Thus an accurate and advance forecast of DOM can significantly aid agricultural planning, power and water

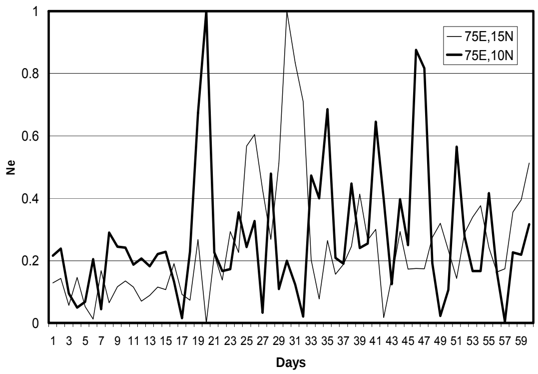


Figure 9: Normalized forecast error in daily rainfall averaged over 24 years (1980-2003), from 5 member ensemble. The error has been normalized to the respective maximum for the three grids. The forecast for each year is taken as a 5-member ensemble average from different leads (April 27 – May 01) with initial condition from daily fields from NCEP reanalysis for the respective year

management as well as sectors like tourism. The basic premise of our attempt at advance forecasting of DOM is based on three considerations. The first one is that although models have inherent limitations in predicting monsoon rainfall, very strong signal associated with processes like the monsoon onset may enhance forecast skill. The second consideration has been that prediction of DOM doesn't involve forecasting actual rainfall, but rather (sharp) transition with certain characteristics. Thirdly, we have proposed a carefully calibrated model configuration to realize this skill; in particular, a configuration capable of resolving the relatively narrow onset domain in a global environment. We have met these criteria by adopting a variable resolution GCM that can resolve the dynamics over the onset domain with sufficient resolution in a global environment. Our results show that with such a configuration significant skill in advance forecasting of DOM can be achieved. As the hindcasts were carried out in complete operational setting, the skill represented here are not just potential but actual realizable skill.

Figure 9 provides further support to our hypothesis of higher skill for predicting onset transition. The three curves in figure

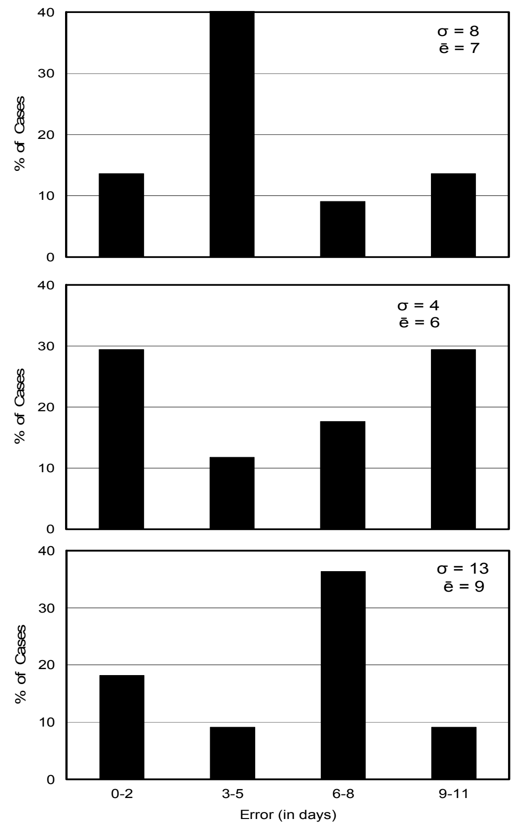


Figure 10: Histogram of errors in predicted dates of onset with respect to dates of onset computed from daily rainfall data from the India Meteorological department. Both the computed and the predicted dates are based on the criteria of significant rainfall with Pre-onset persistence of 3 days and Post-onset persistence of 5 days with area coverage of 30% of the onset domain. The significant daily rainfall is taken to be different for the two cases to allow for model bias: 3 mm (10 mm) for IMD (predicted) data. The three panels from top to bottom represent, respectively, predictions with initial conditions of May 01 for (a) Variable grid with zoom centre 75°E, 10°N (b) Variable grid with zoom centre 75°E,15°N (c) Uniform grid.. The average absolute error ē (days) and the standard deviation σ (days) for each case are given in the respective panel.

9 represent forecast error in daily rainfall normalized to the respective maximum for the three grids. As can be seen, the error is relatively low during the May 15 - June

15 period. These results once again emphasize the need for a careful calibration of the forecast configuration; although the two grids with slightly different zoom centres have very similar grid structure, the grid with zoom over 10°N and 75°E provides the lower (24 year average) error around the onset period. It is interesting to note, however, that the grid with zoom over 15°N and 75°E has generally much less errors than those from the grid with zoom over 10°N and75°E (except around onset days). This raises the possibility of combining forecasts from several grids to generate a (multi-grid) ensemble.

One possible reason for the high success rates of the predictions is the use of relatively high resolution in a global environment. This allows simulation of the spatial variability of the monsoon rainfall, without losing the large-scale organization characteristic and necessary (Goswami and Patra, 2004) for the monsoon. Consistent with our results of optimization, hindcasts with the uniform grid 2° X 1.25° resolution showed much lower skill in predicting DOM, with only about 50% of the cases having error less than 1 σ (Figure 10), although the uniform grid does capture the climatology of daily rainfall well. In addition, it is also necessary to carefully calibrate the onset grid. A change of the zoom centre from 10°N, 75°E to 15°N, 75°E, for example, was found to reduce the skill considerably (Figure 10). It is possible that enhanced resolution, with corresponding improvement in the convective parameterization schemes, will further improve the model forecasts.

The present study was carried out using climatological monthly SST to assess realizable skill; there is, however, evidence of association between interannual variability of SST and onset (Joseph et al., 2006). The significant skill of the forecasts even with climatological SST indirectly hints at a dynamical scenario in which the association between near-shore SST and the onset is a response of the former to the dynamics and forcing (precipitational cooling) due to the latter. As noted earlier, the errors in the forecasts for the El Nino years are not higher than those for the other years. This aspect needs careful examination with carefully designed sensitivity experiments.

One of the strengths of the present study is that skill evaluated represents realizable skill in an operational setting. Although potential skill (such as through use of observed SST) may turn out to be higher, such skill does not provide a true measure of realizable skill. Given the coarse resolution and the inherent errors in the initial data, uncertainties in choice of physical parameterization schemes and other model parameters, there is considerable scope to further enhance the forecast skill. Potentially this implies a very high predictability of the onset process.

References

Ananthakrishnan, R., and M. K. Soman, 1988: The onset of the southwest monsoon over Kerala 1901–1980. *J. Climatol.,* **8,** 283–296.

Ananthakrishnan, R., and M. K. Soman, 1991: The onset of the south-west monsoon in 1990. *Curr. Sci.*, **61**, 447-453.

Boyle, J.S., 1993: Sensitivity of dynamical quantities to horizontal resolution for a climate simulation using ECMWF (cycle 33) model. *J. Climate,* **6**, 796-815.

Flatau, M. K., P. J. Flatau, and D. Rudnick, 2001: The dynamics of double monsoon onsets. *J. Climate*, **14**, 4130-4146.

Gadgil, S., 2003: The Indian monsoon and its variability, *Annu. Rev. Earth Planet. Sci.*, **31**, 429-467.

Goswami B.N., Venugopal V., Sengupta D., Madhusoodanan M.S., Xavier Prince K., 2006: Increasing trend of Extreme Rain Events over India in a Warming Environment, *Science*, **314**, 5804, 1 December, 1442-1445.

Goswami, P and G.K. Patra, 2004: Characteristic Scale of Convective Organization and Monsoon Intensity, *Geophys. Res. Lett.*, **31**(24), L24109.

http://www.pcmdi.llnl.gov/amip/AMIP2EXPDSN/BCS/amipobs_dwnld.html

Hsu, H. H., C. T. Terng, and C. T. Chen, 1999: Evolution of large-scale circulation and heating during the first transition of Asian summer monsoon. *J. Climate*, **12**, 793-810.

Joseph, P. V., J. Eisheid, and R. J. Pyle, 1994: Interannual variability of the onset of the Indian summer monsoon and its association with atmospheric features, El Nino, and sea surface temperature anomalies. *J. Climate,* **7**, 81-105.

Joseph, P. V., K. P. Sooraj and C. K. Rajan, 2006 : The summer monsoon onset process over South Asia and an objective method for the date of monsoon onset over Kerala, *Int. J. Climatol.,* **26** (13), 1871-1893.

Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.,* **77**, 437-471.

Kang, I. S., and Coauthors, 2002: Intercomparison of the climatological variation of Asian summer monsoon precipitation simulated by 10 GCMs. *Climate Dyn.,* **19**,383-395.

Kentarchos, A.S., Roelofs, G.J. & Lelievetd, J., 2000: Simulation of Extratropical Synoptic-scale Stratosphere-Troposphere Exchange Using a Coupled Chemistry GCM: sensitivity to Horizontal Resolution. *J. Atmos. Sci.,* **57**, 2824-2838.

Krishnamurti, T. N., C. M. Kishtawal, Timothy E. LaRow,David R. Bachiochi, Zhan Zhang, C. Eric Williford, Sulochana Gadgil, and Sajani Surendran, 1999 : Improved weather and seasonal climate forecasts from multi-model superensemble. *Science*, **285,** 5433, 1548-1550.

Kumar, K. K., K. R. Kumar, and G. B. Pant, 1997: Pre-monsoon maximum and minimum temperatures over India in relation to the summer monsoon rainfall. Int*. J. Climatol*., **17**, 1115-1127.

Lau, K.-M, H. T. Wu, and S. Yang, 1998: Hydrologic processes associated with the first transition of the Asian summer monsoon: A pilot satellite study. *Bull. Amer. Meteor. Soc.,* 79, 1871-1882.

Laval, K., R. Raghava, J. Polcher, R. Sadourney, and M. Forichon, 1996: Simulation of 1987 and 1988 Indian monsoon using LMD GCM. *J. Climate*, **9**, 3357-3371.

Lorenz, E. N., 1965 : A study of the predictability of a 28-variable atmospheric model. *Tellus*, **17**, 321-333.

Molteni, F. and T.N.Palmer, 1993: Predictability and finite-time instability of the northern winter circulation. *Quart. J. Roy. Meteor. Soc.*, **119**, 269-298.

Moron, V., A.W. Robertson, and M..N. Ward, 2006: Seasonal predictability and spatial coherence of rainfall characteristics in the tropical setting of Senegal. *Mon. Wea. Rev.*, **134**, 3468-3482.

Murakami, T., and T. Nakazawa, 1985: Transition from the Southern to Northern Hemisphere summer monsoon. *Mon. Wea. Rev.*, **113**, 1470-1486.

Pearce, R. P., and U. C. Mohanty, 1984: Onsets of the Asian summer monsoon 1979-82. *J. Atmos. Sci.*, **41**, 1620-1639.

Rajeevan, M., J. Bhate, J.D. Kale and B. Lal., 2006: High resolution daily gridded rainfall data for the Indian region. *Curr. Sci.*, 91 (3).

Ramesh, K. V., and P. Goswami, 2007: Reduction in temporal and spatial extent of the Indian summer monsoon, *Geophys. Res. Lett.*, **34**, L23704.

Rao, Y. P., 1976: Southwest monsoon: Synoptic meteorology. *Meteor. Monogr.*, No.1/1976, India Meteorological Department, 367pp.

Sabre, M., K. Hodges, K. Laval, J. Polcher, and F. Désalmand, 2000: Simulation of Monsoon Disturbances in the LMD GCM. *Mon. Wea. Rev.,* 128 (11), 3752-3771.

Sadourny, R. and Laval, K., 1984: January and July Performances of the LMD General Circulation Model. *New Perspectives in Climate Modelling*. Berger A. Ed. Elsevier 173-198.

Sharma, O. P., H. C. Upadhyaya, Th. Braine-Bonnaire, and R. Sadourney, 1987: Experiments on regional forecasting using a stretched coordinate general circulation model. Short and medium range numerical weather prediction, Special volume of *Meteo. Soc. Japan,* Ed T. Matsuno, 65, 263–271.

Sikka, D.R., and S.Gadgil, 1980: On the maximum cloud zone and the ITCZ over Indian longitude during the southwest monsoon. *Mon. Wea. Rev.,* **108**, 1840-1853.

Soman, M. K., and K. K. Kumar, 1993: Space-time evolution of meteorological features associated with the onset Indian summer monsoon. *Mon. Wea. Rev*., **121**, 1177-1194.

Sperber, K. R., & Palmer, T. N., 1996: Interannual Tropical Rainfall Variability in General Circulation Model Simulations Associated with the Atmospheric Model Intercomparison Project. *J. Climate*, **9**, 2727-2750.

Shukla, J., 1981: Dynamical Predictability of Monthly Mean, *J. Atmos. Sci.*, **38**(12), 2547-2572.

Takagi, T., F. Kimura, and S. Kono, 2000: Diurnal variation of GPS precipitable water at Lhasa in premonsoon and monsoon periods. *J. Meteor. Soc. Japan*, **78**, 175-180.

Tanaka, M., 1992: Intraseasonal oscillation and the onset and retreat dates of the summer monsoon over the east, southeast and western North Pacific region using GMS high cloud amount data. *J. Meteor. Soc. Japan*, **70**, 613-629.

Tiedtke, M. A., 1989: A comprehensive Mass Flux Scheme for Cumulus Parameterization in Large-scale Models. *Mon . Wea. Rev.,* **117**, 1799-1800.

Ueda, H, and T. Yasunari, 1998: Role of Warming over the Tibetan Plateau in Early Onset of the Summer Monsoon over the Bay of Bengal and the South China Sea. *J. Meteor. Soc. Japan*, **76**, 1-12.

Wang, Bin and Lin Ho, 2002: Rainy season of the Asian–Pacific Summer Monsoon. *J. Climate,* 15, 386-398.

Webster, P. J., 1983: Mechanisms of monsoon transition: Surface hydrology effects. *J. Atmos. Sci.,* 40, 2110-2124.

Webster, P. J., V. Magana, T. Palmer, J. Shukla, R. Tomas, M. Yanai, and T. Yasunari, 1998: Monsoons: Processes, predictability and the prospects for prediction. *J. Geophys. Res.*, **103**, 14451-14510.

Wu, G. X., and Y. S. Zhang, 1998: Tibetan Plateau forcing and the timing of the monsoon onset over South Asia and the South China Sea. *Mon. Wea. Rev.*, **126**, 913-927.