

# An Optimum Initial Manifold for Improved Skill and Lead in Long-range Forecasting of Monsoon Variability

K C GOUDA (✉ [kcgouda@csir4pi.in](mailto:kcgouda@csir4pi.in))

CSIR C-MMACS <https://orcid.org/0000-0003-2322-021X>

S Joshi

Nagaraj Bhat

---

## Research Article

**Keywords:** monsoon variability, initial manifold approach

**Posted Date:** March 10th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-233987/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

**Version of Record:** A version of this preprint was published at Theoretical and Applied Climatology on March 23rd, 2021. See the published version at <https://doi.org/10.1007/s00704-021-03589-x>.

# Abstract

Using an initial manifold approach, an ensemble forecast methodology is shown to simultaneously increase lead and realizable skill in long-range forecasting of monsoon over continental India. Initial manifold approach distinguishes the initial states that have coherence from a collection of unrelated states. In this work, an optimized and validated variable resolution general circulation model is being adopted for long range forecasting of monsoon using the multi-lead ensemble methodology. In terms of realizable skill (as against potential) at resolution ( $\sim 60\text{km}$ ) and lead (2–5 months) considered here the present method performs very well. The skill of the improved methodology is significant, capturing 9 of the 12 extreme years of monsoon during 1980–2003 in seasonal (June-August) scale. 8-member ensemble average hindcasts carried out for realizable skill with lead of 2 (for June) to 5 (for August) months and an optimum ensemble is presented.

## 1. Introduction

The conventional formalism generally implies a loss of forecasting skill with increasing lead in the long-range prediction of rainfall (Lorenz 1965; Molteni and Palmer 1993; Goswami 1998). Although ensemble forecasting has been applied extensively in short-range forecasting (Toth and Kalnay 1997; Toth and Kalney 1993; Buizza 1997; Kyouda and Kusunoki 2002), its application in long-range forecasting (LRF) has been comparatively less explored. The dynamical models used for LRF are based on ensemble forecasting (Buizza et al., 1999; Mullen and Buizza 2001; Mullen and Buizza 2002; Saha 2006; Goswami and Gouda 2008,2009; Abhilash et al., 2014). For LRF of monsoon, the initial manifold is generated by sampling states over a period of time based on dynamical considerations in terms of intra seasonal oscillations (ISO) (Madden and Julian 1971; Yasunari 1979; Krishnamurti and Ardanuy 1980; Goswami et al. 2005), whose phases and amplitudes can significantly affect the monsoon (Hendon 1990; 2000). We adopt an optimized and validated general circulation model with a variable resolution for long range forecasting of monsoon (Goswami and Gouda 2009; Gouda et al. 2018; Joshi et al. 2020; Gouda et al. 2020). The forecast skill with ensemble simulations for the seasonal rainfall prediction is generally better with general circulation models (Kirtman and Shukla 2002; Wang et al. 2004; Goswami and Gouda 2009; Joshi et al., 2020; Gouda et al. 2018; Gouda et al. 2020). Improvement in accuracy of forecasting monsoon rainfall at long range has many applications; however, in spite of advances in model physics, resolution and data, there has been very little improvement in operational skill of GCM in long-range forecasting of monsoon (Wang et al 2004; Kang and Shukla, 2006; Moron et al. 2006; Palmer et al. 2004). The difficulties in improving range or accuracy of forecasts are part of the general challenges in predictability; in particular, it is generally accepted that an increase in range of forecast leads to poorer accuracy due to growth of error (Lorenz 1965; Molteni and Palmer 1993). It may therefore be counter intuitive to attempt a simultaneous increase in the lead and accuracy of a forecast. It has been recognized that inclusion of improved forecast methodology, such as bias correction and multi-model ensemble can significantly improve forecast skill (Kirtman and Shukla 2002; Wang et al. 2004; Yun et al. 2003). Besides, the predictability at long range depends also on the scale and degrees of freedom; it has

been shown that non-linearity in presence of a large number of degrees of freedom can improve predictability (Kang and Shukla 2006).

A primary source of errors in forecasts is due to the uncertainties in the initial conditions. A well-known methodology to quantify and reduce these uncertainties is ensemble forecasting (Toth and Kalnay 1997; Toth and Kalney 1993; Buizza 1997). In terms of implementation, the ensemble forecasting generate a number of simulations using multiple lead initial conditions such that, the average of the ensembles is more accurate than the simulation (forecast) obtained from any of the individual member of the ensembles. Generally, researchers consider a large sample of simulations. Using the high spread of the ensemble members which contain the qualitative information, there will be an increase in the reliability of the forecast which in turn provides a good basis for the probabilistic forecast. Thus, unlike in a classical forecast where the system evolves from a given initial state (point in the phase space), in an ensemble forecast the system evolves from a neighborhood of states.

The conventional ensemble forecasting can be said to assume all initial conditions (with comparable leads) in a given neighborhood in state space to be equivalent (Buizza et al. 1999). This is true for an isolated system like a non-linear oscillator whose initial state can be prepared by the observer. It also appears reasonable for short-range forecasting for which the day-to-day synoptic noise can overwhelm any (weaker) underlying structure (Moron et al. 2006). However, it may not be valid for processes like the Indian summer monsoon (ISM) which is a part of a continuous global dynamics. The ISM region is mainly characterized by quasi-periodic oscillation spectrum which influences on the regional circulation pattern. The Madden Julian Oscillation (Madden and Julian 1971) is placed in the lower end of the spectrum of the intra-seasonal Oscillation (ISO). Studies also inferred the wider spectrum of ISO with a range of 20–90 days (Goswami et al. 2005; Krishnamurti and Ardanuy 1980; Yasunari 1979; Hendon 2000) and it is well established theory that ISO is very important for the tropical as well as extra tropical systems including monsoon. ISO presence also implies that the states in the pre-monsoon system are not dynamically disconnected but are embedded in an underlying dynamic. Thus, the initial states belonging to a cycle of an ISO, say 90 days, will have dynamic coherence; we shall call such states a manifold of states to distinguish them from a collection of unrelated states, such as those used in short-range forecasting or simulation of an isolated system. We shall use this dynamic coherence of initial states for improving skill in long-range forecasting of monsoon.

The important and challenging issue of ensemble forecasting is the generation of the best ensemble i.e. creation of a set of initial states which will simulate better results using the dynamical model. In short-range prediction, many techniques evolved to create the ensembles by generating the perturbations (Toth and Kalney 1993; Kyouda and Kusunoki 2002). In LRF, the creation of ensemble of initial conditions is mainly required to represent the dynamical states on different dates. Thus, in case of long-range forecasting of monsoon, the problem of choice of an ensemble can be said to be the choice of an initial manifold. It has been shown that an ensemble with wider spread but longer lead can provide a better monsoon rainfall prediction skill than an ensemble of states, taken over a short initial manifold during pre-monsoon period i.e. March 01-April 30.

## 2. Model, Data And Methodology

The atmospheric General Circulation Model (GCM) adopted from LMD, France with variable resolution is considered for this simulation study. The GCM uses a stretched coordinate with a zoom so we can get higher resolution over a chosen domain. The principle, formulation of the model is described in earlier studies (Sharma et al. 1987; Hourdin et al. 2006) and the version used in this study is the optimum configuration and already used for long-range forecasting of monsoon rainfall (Goswami and Gouda 2009, 2010; Gouda et al., 2018; Joshi et al. 2020, Gouda et al. 2020). The skill evaluation was carried out for seasonal hindcasts. In this ensemble simulation study, ~~study~~ the model is integrated from initial condition date to August end for all the 24 years i.e. 1980 to 2003 in hind cast mode. The simulated ensemble rainfall is represented as follow: **see formulas 1, 2, and 3 in the supplementary files.**

## 3. Results And Discussion

A simple measure of manifold structure is the correlation among the members of the set (ensemble). Correlation of states (globally averaged 850mb zonal wind) for 54 years (1950-2003) during certain (pre-monsoon) period with reference states taken from during the monsoon season (June 1, July 1 and August 1) reveals this structure (Fig. 1, left panel). In particular, the states between 18<sup>th</sup> March-15<sup>th</sup> April and between 15<sup>th</sup> April -15<sup>th</sup> May are significantly correlated with at least two of the reference states. In contrast, there is little such correlation of states in the winter season. (Fig. 1, right panel). Similarly, the average correlation of the states belong to CE with states of June 1, July 1 and August 1 are respectively 0.06, 0.16 and 0.15 (Fig. 1, left panel); thus these states do not have any coherence with the monsoon state. In other words, certain states during pre-monsoon period have better coherence with monsoon states and can be expected to provide better skill in capturing monsoon dynamics. The same analysis averaged over a domain 0-30<sup>0</sup>N and 65-95<sup>0</sup>E) is presented in Fig. 2.

We then consider eight such sub-manifolds of the initial manifold, each characterized by eight members (leads) spread over a period of about fifteen days. As our null hypotheses we consider two ensembles, one is a compact ensemble (CE) of eight leads of closely packed states (April 23-May01) and the other a large ensemble (LE) of states from all the eight test ensembles (March 01-April 30). Thus, the CE, due to its short time span, can not embed the dynamical coherence inherent in an initial manifold characterized by low frequency ISO; the CE is thus akin to a collection of unrelated (synoptic) states. This is clear from the fact that none of the states of the eight leads in the CE has a significant correlation with any of the reference states (Fig. 2). In other words, it is expected that the better sampling of initial states over ISO time scale generally compensates error due to the longer lead and in turn improves the forecast using ensemble with shorter lead and small sampling time scale. A measure of difference among the ensembles is the standard deviation among the members (normalized to ensemble mean) and the same is presented in Fig. 3. It can be seen (Fig. 3) that the initial manifolds with wider spread have larger internal structure than a compact ensemble. Further, there is a gradual decrease in the richness of this

structure beyond set 4 (March 18-April 15; Table 1). Thus, the ensembles with longer leads have the higher dispersion of states. Based on the results of Fig. 1 and Fig. 3, therefore, we expect the ensemble states between March 18 – April 15 to have highest skill and shall be referred to as optimum initial manifold (OIM) in subsequent discussion. It needs to be emphasized that the number of states in an ensemble should be chosen carefully to ensure that the results are stable with respect to the size of the ensemble; however, it has been shown that for the model configuration, dispersion among the forecasts from different ensembles saturates at an ensemble size of 6 (Goswami and Gouda, 2009). It may be mentioned that in terms of climatological seasonal cycle, most ensembles performed comparably well (Fig. 4). However, in terms of inter-annual variability in area-averaged (75-85°E, 8-28°N) seasonal (June-August) rainfall, defined as departure from corresponding 24 years (1980-2003) mean, OIM outperforms both CE and LE (Fig. 5). The OIM has a phase synchronization of 67% with a correlation coefficient of 0.44 between all-India seasonal (JJA) rainfall anomalies, significant at 99% confidence level for the degrees of freedom involved.

The model simulations are compared with the coupled models from international centres like Asia Pacific Economic Cooperation (APEC) Climate Center (APCC) and National centre for environmental prediction. The Seamless Coupled Prediction System (SCoPS) of APCC (Ham et al. 2019) and the Climate Forecast System (CFSv2) of NCEP (Saha et al. 2010; 2014) are used for the comparison of rainfall simulations with the IMD observation. We have used the 3-6 month lead predictions i.e. March lead for the months of June-August for the year 1982 to 2003 and compared the inter-annual variability of seasonal (June-August) rainfall over continental India using the APCC and NCEP long-range simulations and IMD observations as presented in the figure S1 which clearly indicates the simulation with APCC model has zero correlation and only 45% phase synchronization while the NCEP (CFSv2) has 0.43 correlation with 59% of Phase which is lower compared to the OIM used in our study as mentioned earlier.

A number of other parameters have been considered to quantify the skill of the forecasts (Table 1). Comparison of skill of the forecasts for different initial manifolds in terms of these parameters shows OIM to have highest and significant skill (Table 1). The total number of failures ( $N_{UW}+N_{OW}+N_M+N_{FF}+N_{FD}$ ) for OIM is 9, followed by 13 for the ensemble with starting initial state of March 01; However, this ensemble scores only 54% in terms of phase synchronization (Table 1). As the phase synchronization is a binary (i.e. 0 or 1) process, the random forecast would be expected to result in a 50% success rate. Thus, this ensemble scores only marginally better than a random forecast.

An important consideration in evaluation of skill is the performance for extreme years. Examination of the skill for the years with amplitude of anomaly more than 5% at different scales shows (table 2) OIM to perform significantly better than most of the other ensembles; only two ensembles have higher (by one) cases of extreme years in phase.

It is important to note that a larger ensemble (LE), with initial states including those of the optimum ensemble, does not provide a better forecast; the optimum initial manifold has significantly better skill than the LE. To measure the effectiveness of ensemble forecasting, the ratio of error ( $\epsilon$ ) in the ensemble

forecast and the average forecast error from the ensemble members are computed and presented in figure 6. The initial manifolds 3-5 have among the lowest values of this error ratio; on the other hand, the CE ensemble has this ratio higher than one, while for the LE, this ratio is comparable to that of OIM.

## Conclusion

The simulations carried out here in the hind cast mode using the monthly climatological SST shows that the ensemble forecasting skill is minimal but realizable unlike the simulations with observed SST having potential skill. This ensemble forecasting shows that the forecast dispersion is mainly due to atmospheric internal dynamics (atmospheric states) and the land surface processes. It is logical to expect that use of an OIM in a coupled ocean-atmosphere model will further improve the realizable skill of long-range forecasting of monsoon. While it has been recognized that prescription of sea surface temperature is a major factor that determines monsoon simulation (Kirtman and Shukla 2002; Wang et al., 2004), it should be noted, however, that the primary requirement for this, an ocean-atmosphere coupled model with sufficient skill in forecasting monsoon, still does not exist. While there have been a number of studies on potential skill (using observed SST for assessing skill) on variables like seasonal mean rainfall using several models, fewer studies have considered realizable skill (as in an actual forecast environment). In terms of realizable skill at resolution and lead considered here, the present method out performs any reported skill so far (Kirtman and Shukla 2002; Wang et al., 2004). This manifold approach can be used for the long range prediction of seasonal monsoon rainfall over the continental India.

## Declarations

**Conflicts of interest:** The author declares that he has no conflict of interest.

**Funding :** This work is supported by the projects funded by the Department of Science and Technology (DST) and National Mission on Himalayan Studies (NMHS) of Ministry of Environment, forest and climate change, Govt. of India

**Author's Contribution:** KCG conceived the presented idea and performed computations. KCG, SJ and NB analyzed the results from the computations and wrote the manuscript.

**Availability of data and material:** The APCC model simulation, NCEP reanalysis datasets are freely available on the web sites. Gridded rainfall observation data are collected from India Meteorological Department.

**Code availability:** Model results are available from the corresponding author upon request.

**Ethics approval:** Not applicable

**Consent to participate:** Not applicable

**Consent for publication:** Not applicable

**Acknowledgement:** We acknowledge NCEP and IMD for the data support. The authors acknowledge the APCC MME Producing Centers for making their hindcast/forecast data available for analysis, the APEC Climate Center for collecting and archiving the data, as well as for producing APCC MME predictions. We acknowledge Department of Science and Technology for the project support.

## References

1. Abhilash S, Sahai AK, Borah N *et al.* (2014) Prediction and monitoring of monsoon intraseasonal oscillations over Indian monsoon region in an ensemble prediction system using CFSv2. *Clim Dyn* **42**, 2801–2815. <https://doi.org/10.1007/s00382-013-2045-9>
2. Buizza R (1997) Potential forecast skill of ensemble prediction, and spread and skill distribution of the ECMWF Ensemble Prediction System. *Mon Weather Rev* **125**: 99-119
3. BuizzaR, Hollingsworth A, Lalaurette F & Ghelli A (1999) Probabilistic predictions of precipitation using the ECMWF Ensemble Prediction System. *Weather and Forecasting* **14**, 168–189.
4. Buizza R, Miller M, Palmer TN (1999) Stochastic representation of modeluncertainties in the ECMWF Ensemble Prediction System. *Quarterly Journalof the Royal Meteorological Society* **125**: 2887–2908. <https://doi.org/10.1002/qj.49712556006>.
5. Goswami BN (1998) Interannual Variations of Indian Summer Monsoon in a GCM: External Conditions versus Internal Feedbacks. *J Climate* **11**: 501-522
6. Goswami P, Sijikumar S, Mandal A (2005), Seasonal cycle and intraseasonal oscillations in the interannual variability over the monsoon region. *Geophys Res Lett* **32**: L06810, doi:10.1029/2004GL022171
7. Goswami P, Gouda KC (2009) Comparative Evaluation of two Ensembles for Long-range Forecasting of Monsoon Rainfall. *Mon Weather Rev* **137** (9): 2893-2907
8. Goswami P, Gouda KC (2010) Evaluation of a Dynamical Basis for Advance Forecasting of the Date of Onset of Monsoon Rainfall over India. *Mon Weather Rev* **138** (8): 3120–3141.
9. Gouda KC, Nahak S, Goswami P (2018) Evaluation of a GCM in seasonal forecasting of extreme rainfall events over continental India. *Weather and Climate Extremes* **21**: 10-16.
10. Gouda KC, Nahak S, Goswami P (2020) Deterministic seasonal quantitative Precipitation forecasts: Benchmark skill with a GCM. *Pure Appl Geophys* <https://doi.org/10.1007/s00024-020-02463-7>
11. Hendon H, Liebmann HB, Newman M, Glick JD, Schemm T (2000) Medium range forecast errors associated with active period of Madden-Julian Oscillation. *Mon Weather Rev* **128**: 69-86

12. Hendon H, Liebmann B (1990) The intraseasonal (30-50 day) oscillation of the Australian summer monsoon. *J Atmos Sci* 47:2909-2923
13. Krishnamurti TN, Ardanuy P (1980) The 10- to 20 day westward propagating mode and breaks in the monsoons. *Tellus* 32: 15-26
14. Ham et al. (2019) A newly developed APCC SCoPS and its prediction of East Asia seasonal climate variability. *Clim Dyn* 53, 3703-3704. <https://doi.org/10.1007/s00382-019-04894-y>
15. Hourdin F et al. (2006) The LMDZ4 general circulation model: Climate performance and sensitivity to parameterized physics with emphasis on tropical convection. *Clim Dyn* 27(7-8): 787-813.
16. Joshi S, Gouda KC, Goswami P (2020) Seasonal rainfall forecast skill over Central Himalaya with an atmospheric general circulation model. *Theor Appl Climatol* 139:237-250 DOI 10.1007/s00704-019-02971-0
17. Kalnay E et al. (1996) The NCEP/NCAR 40-year reanalysis project. *Bull Amer Meteor Soc* 77: 437-471
18. Kang IS, Shukla J. (2006) Dynamic seasonal prediction and predictability of the monsoon. In: *The Asian Monsoon*. Springer Praxis Books. Springer, Berlin, Heidelberg. [https://doi.org/10.1007/3-540-37722-0\\_15](https://doi.org/10.1007/3-540-37722-0_15)
19. Kirtman B, Shukla J (2002) Interactive coupled ensemble: A new coupling strategy for CGCMs. *Geophys Res Lett*, 29:1367, doi: 10.1029/2002GL 014834
20. Kyouda M, Kusunoki S (2002) Ensemble Prediction System. Outline of the Operational Numerical Weather Prediction at the Japan Meteorological Agency. *JMA* 59-63
21. Lorenz EN (1965) A study of the predictability of a 28-variable atmospheric model. *Tellus* 17: 321-333
22. Madden RA, Julian PR (1971) Detection of a 40-50 day oscillation in the zonal wind in the tropical Pacific. *J Atmos Sci* 28: 702-708
23. Molteni F, Palmer TN (1993) Predictability and finite-time instability of the northern winter circulation. *Quart J Roy Meteor Soc* 119: 269-298
24. Moron V, Robertson VAW, Ward MN (2006) Seasonal predictability and spatial coherence of rainfall characteristics in the tropical setting of Senegal. *Mon Weather Rev* 134: 3468-3482
25. Mullen SL & BuizzaR (2001) Quantitative precipitation forecasts over the United States by the ECMWF Ensemble Prediction System. *Monthly Weather Review* 129,638–663.
26. MullenSL & BuizzaR (2002) The impact of horizontal resolution and ensemble size on probabilistic forecasts of precipitation by the ECMWF Ensemble Prediction System. *Weather and Forecasting* 17,

173–191.

27. Palmer TN et al. (2004) Development of a european multimodel ensemble system for seasonal-tointerannual prediction (DEMETER). *Bull Amer Meteor Soc* 85: 853–872.
28. Rajeevan M, Bhate J, Kale JD, Lal B (2006) High resolution daily gridded rainfall data for the Indian region. *Current Science* 91(3): 296-306
29. Saha S & Coauthors (2006) The NCEP Climate Forecast System. *Climate*, **19**, 3483–3517, <https://doi.org/10.1175/JCLI3812.1>.
30. Saha, S, S. Moorthi, H. Pan, X. Wu, J. Wang, and Coauthors (2010) The NCEP Climate Forecast System Reanalysis. *Bulletin of the American Meteorological Society*, **91**, 1015–1057, [doi:10.1175/2010BAMS3001.1](https://doi.org/10.1175/2010BAMS3001.1).
31. Saha, S, S Moorthi, X Wu, J Wang, and Coauthors (2014) The NCEP Climate Forecast System Version 2. *Journal of Climate*, **27**, 2185–2208, [doi:10.1175/JCLI-D-12-00823.1](https://doi.org/10.1175/JCLI-D-12-00823.1)
32. Sharma, OP, Upadhyaya, HC, Braine-Bonnaire Th, Sadourney R (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
33. Toth Z, Kalnay E (1993) Ensemble forecasting at NMC: The generation of perturbations. *Bull Amer Meteor Soc* 74: 2317–2330
34. Toth Z, Kalnay E (1997) Ensemble forecasting at NCEP: the breeding method. *Mon Weather Rev*, 12: 3297-3319
35. Yasunari T (1979) Cloudiness fluctuation associated with the Northern Hemisphere summer monsoon. *J Meteor Soc Japan* 57: 227-242
36. Yun WT, Stefanova L, Krishnamurti TN (2003) Improvement of the multi-model super ensemble technique for seasonal forecasts. *J climate* 22: 3834-3840
37. Wang B, Kang IS, Lee JY (2004) Ensemble simulations of Asian-Australian monsoon variability during 1997/1998 El Nino by 11 AGCMs. *J Climate* 17: 803-818

## Tables

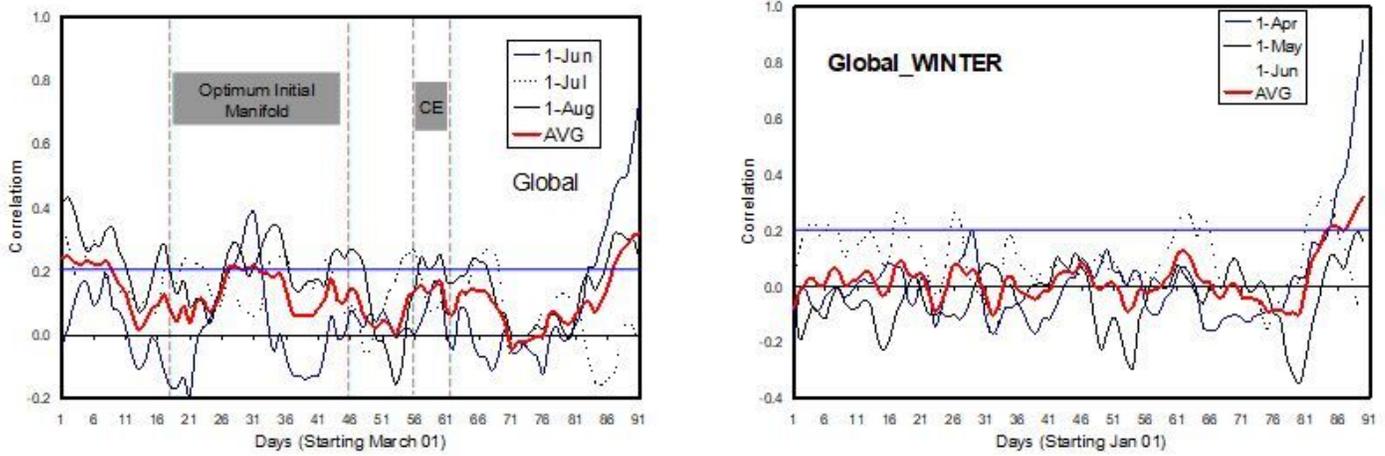
Table 1: Comparative evaluation of 9 initial manifolds (ensembles) with two test ensembles for 24 years (1980-2003)

Set	Leads		Warning						IAV	
	Starting Lead	Ending Lead	N <sub>UW</sub>	N <sub>OW</sub>	N <sub>M</sub>	N <sub>FF</sub>	N <sub>FD</sub>	Total Failures	Phase Sync (%)	CC
1	Mar01	Apr01	2	5	4	1	1	13	54	0.15
2	Mar07	Apr07	4	5	4	1	0	14	67	0.33
3	Mar15	Apr11	2	7	6	1	2	18	63	0.43
4 (OIM)	Mar18	Apr15	4	3	2	0	0	9	67	0.44
5	Mar21	Apr18	4	5	6	1	2	18	58	0.2
6	Mar25	Apr21	3	5	7	2	2	19	67	0.27
7	Mar30	Apr27	5	8	8	4	2	27	58	0.08
8	Apr01	Apr28	4	9	7	3	1	24	50	0.09
9	Apr07	Apr29	6	7	9	3	0	25	42	0.05
CE	Apr23	May01	4	6	7	1	1	19	63	0.4
LE	Mar01	Apr30	8	6	8	2	0	24	54	0.1

Table 2: Comparative evaluation of 9 initial manifolds (ensembles) with two test ensembles for 12 extreme (amplitude of anomaly in all-India seasonal rainfall > 5%) years. The number in the bracket for each year indicates the observed % of anomaly for the year. “y” indicates that the simulation and observation values of rainfall anomalies are in phase.

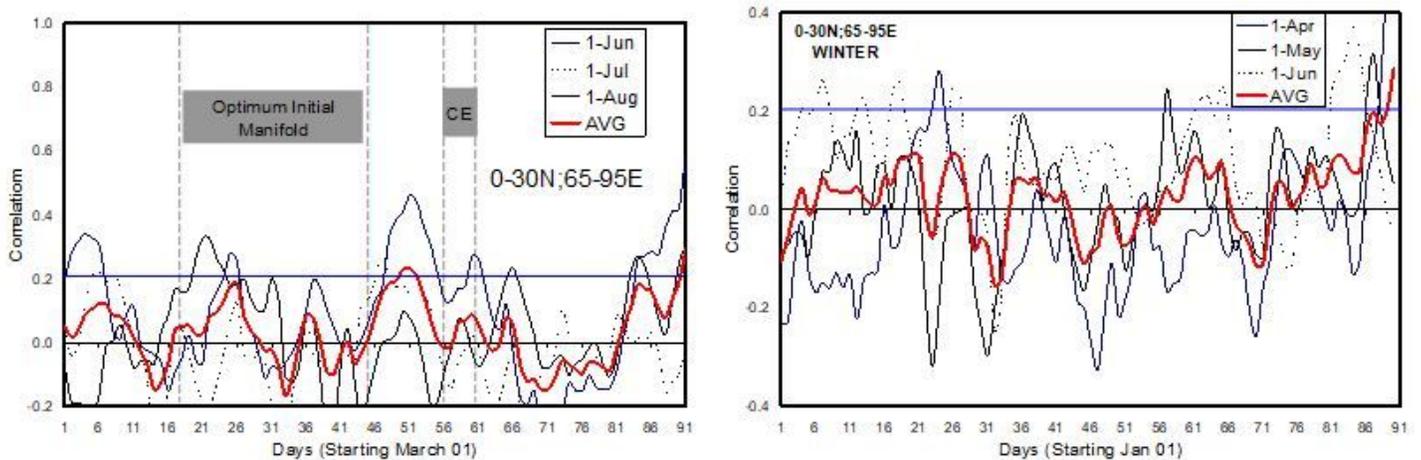
Ensemble	1980 (12.5)	1982 (-9)	1986 (-7)	1987 (-13)	1988 (13)	1990 (6)	1994 (11)	1998 (7)	1999 (-9)	2000 (-7)	2002 (-18)	2003 (9)	No of year
1	y	y	y	y			y	y	y	y			8
2	y		y	y			y	y	y	y		y	8
3	y		y	y			y	y	y	y		y	8
4 (OIM)	y	y	y	y			y	y	y	y		y	9
5	y		y	y		y	y	y	y	y			8
6	y	y	y	y		y	y	y	y	y		y	10
7		y	y	y		y		y	y	y		y	8
8		y	y	y			y	y	y	y		y	8
9		y	y	y			y	y	y	y			7
CE	y	y	y	y	y		y	y		y	y	y	10
LE	y		y	y			y	y	y	y			7

## Figures



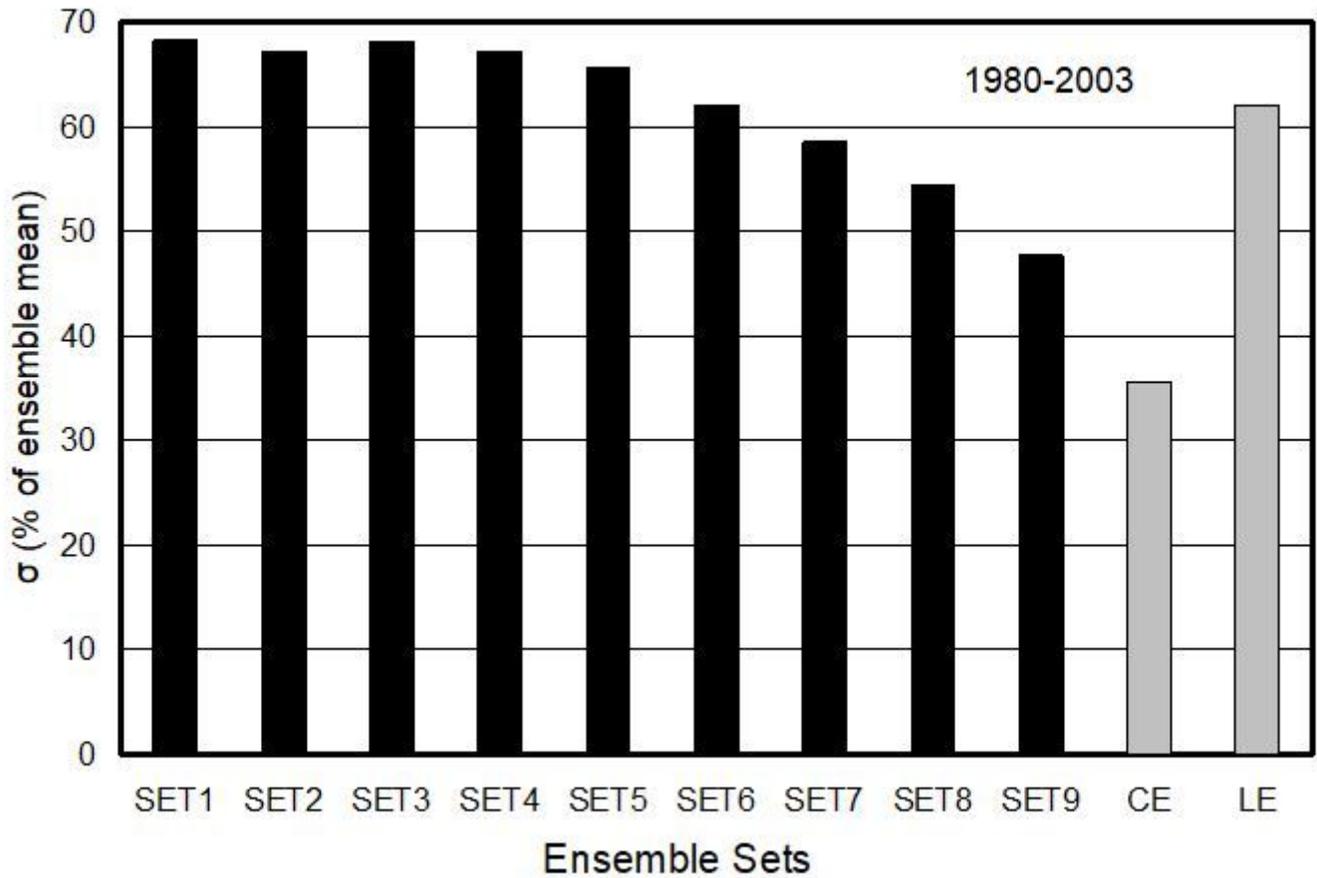
**Figure 1**

Structure of the initial manifold as revealed by auto correlation among the initial states (globally averaged 850 mb zonal wind). The correlation is calculated between the area-averaged states on a given date and those for a reference date (taken as June1, July 1 and August 1 (left panel) and April1, May1 and June1(right panel) for 54 years (1950-2003). The significance at 99% confidence level for the degrees of freedom involved is 0.23. The initial states are taken from NCEP Reanalysis. While these correlations are at the significance level for the states belonging to March 20 – April 15 in general during the monsoon period, they are essentially insignificant for the winter season (right panel).



**Figure 2**

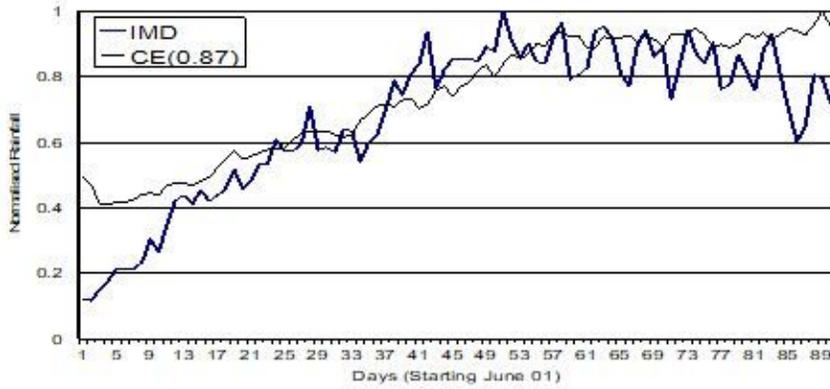
Same as figure 1 but 850 mb zonal wind averaged over area 0-30N and 65-95E).



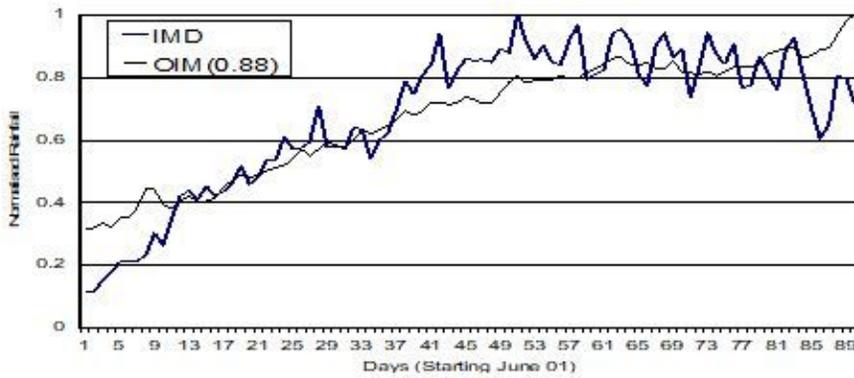
**Figure 3**

Comparison of the structure of initial manifolds (in terms of standard deviation among ensemble members as % of the ensemble mean) for the nine initial manifolds with the compact ensemble (CE) and the large ensemble (LE) respectively. The higher structure implies larger dispersion among the initial states and hence larger dispersion among the forecasts from the members.

Compact Ensemble (Apr23-May01)



Optimum Initial Manifold (Mar18-April15)



Large Ensemble (Mar01-April 30)

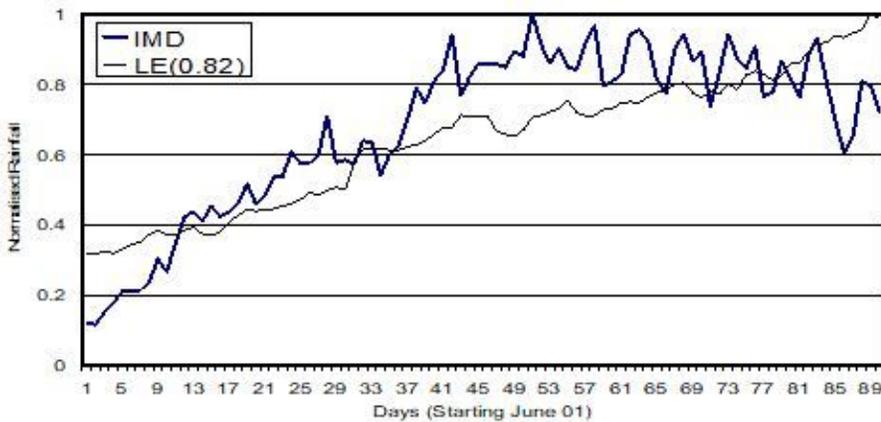
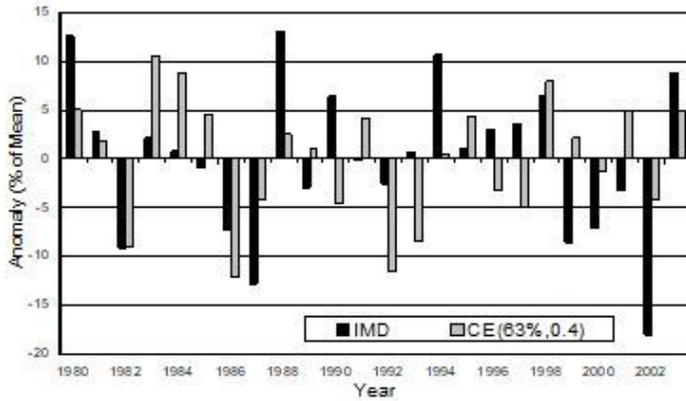


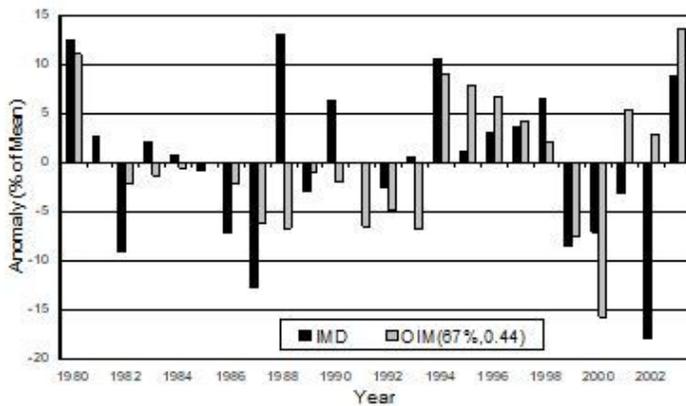
Figure 4

Seasonal cycle of the daily rainfall climatologies (normalised to maximum value in season) for the period June to August from IMD (thick solid line) compared with CL (top panel), optimum initial manifold (middle panel) and LE (bottom panel). The rainfall values are considered in the area average over the region 75-85oE; 8-28oN. The correlation coefficient between IMD observation and simulations are given in bracket.

Compact Ensemble (Apr23-May01)



Optimum Initial Manifold (Mar18-April15)



Large Ensemble (Mar01-April 30)

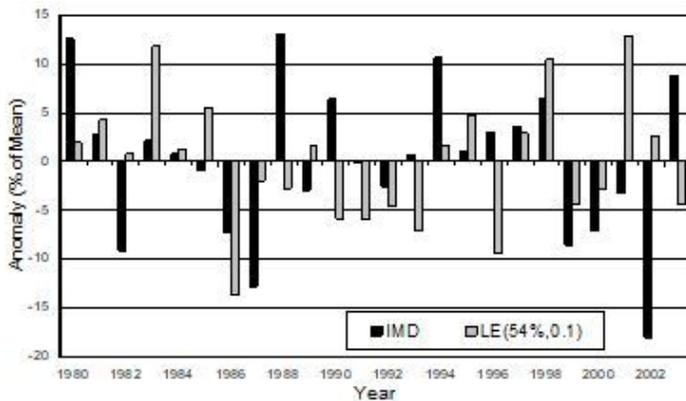
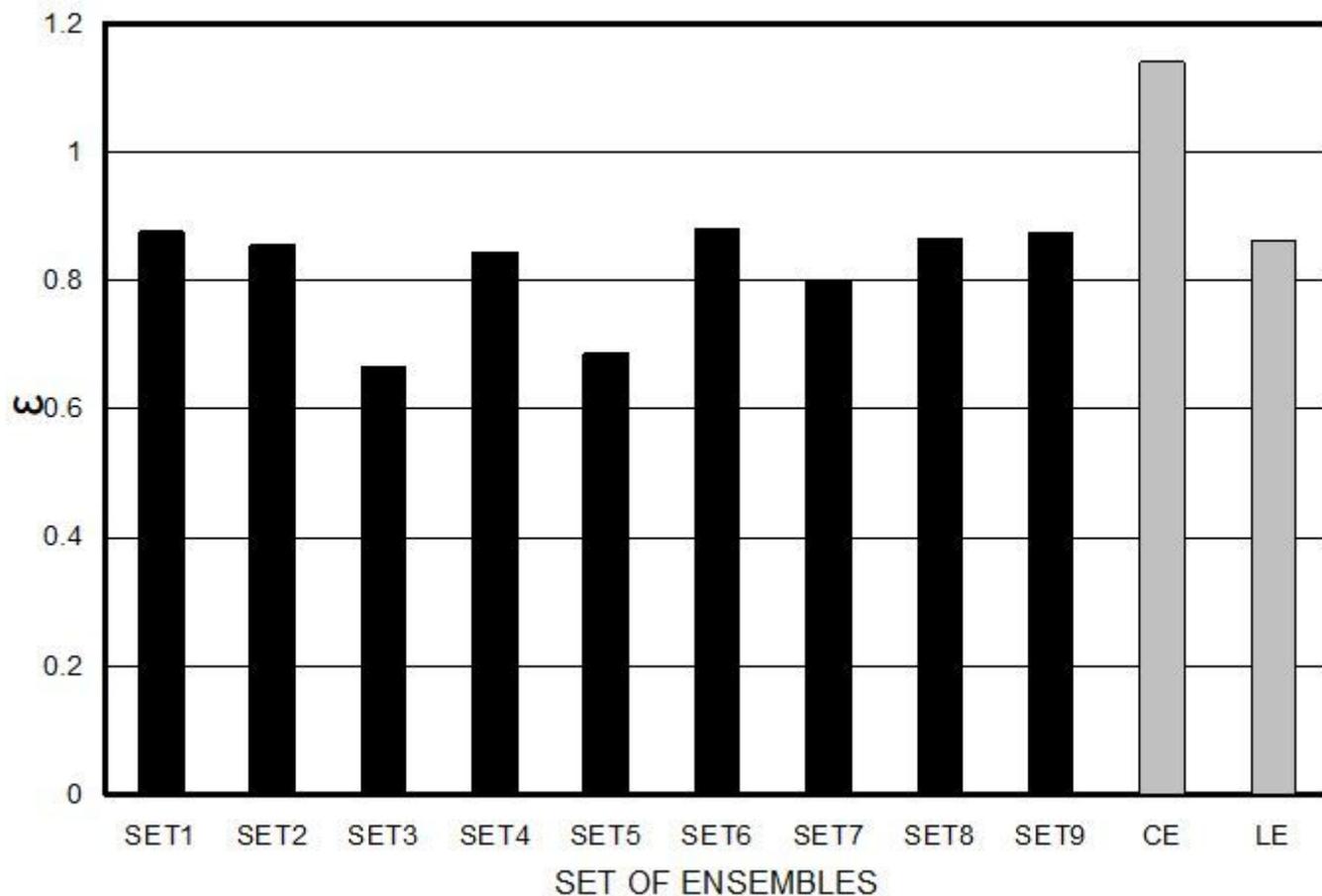


Figure 5

Inter annual variability in the all India monsoon seasonal (June-August) rainfall anomaly (expressed as % of respective mean) for (a) CE (top panel), (b) OIM (middle panel) and (c) LE (bottom panel). The two numbers in the bracket in each panel represent, respectively, the phase synchronization (%) and correlation coefficient with observed (IMD) rainfall anomalies.



**Figure 6**

Ratio of error ( $\epsilon$ ) in ensemble forecast to average error among ensemble members for different initial manifolds. The rainfall forecasts are ensemble-average forecasts for different sets of the initial manifold and the ratio is computed over area 75-85oE; 8-28oN.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [formulas.docx](#)
- [SNEHTAAC11FEB2021FigS1.docx](#)