MAUSAM, **70**, 2 (April 2019), 277-292

551.577.3 (540)

Multi-scale variability and predictability of Indian summer monsoon rainfall

ANDREW W. ROBERTSON, VINCENT MORON*, NICOLAS VIGAUD, NACHIKETA ACHARYA,
ARTHUR M. GREENE and D. S. PAI**

International Research Institute for Climate and Society (IRI), Columbia University, Palisades, NY, USA

**Aix Marseille Univ, CNRS, IRD, INRA, Coll France, CEREGE, Aix-en-Provence, France*

***India Meteorological Department, Pune – 411 005, India*

e mail : awr@iri.columbia.edu

सार – यह शोध पत्र आई आर आई में लेखकों तथा उनके सहयोगियों द्वारा पिछले एक दशक से किए गए शोध की समीक्षा करता है जो मौसमी और उपमौसमी समय पैमाने के अनुसार भारतीय ग्रीष्मकालीन मॉनसून वर्षा (ISMR) पूर्वानुमान और प्रागुक्तिता से संबंधित है। आइ एम डी के ग्रिडेड डेटा के आधार पर कृषि के लिए स्थानीय पैमाने पर दैनिक ISMR के अनुभवजन्य विश्लेषणों से पता चलता है कि ऋतु में बारिश के दिनों की संख्या और ऋतु की कुल वर्षा का पूर्वानुमान सही लगाया गया है, इसके अलावा 'जलवायु के अंतर्गत मौसम' का पूर्वानुमान एक महत्वपूर्ण मौसमी मॉड्यूलेशन से गुजरता है, और मॉनसून के शुरूआती तथा बाद के चरणों में सबसे अधिक और जुलाई-अगस्त की मुख्य मॉनसून की अवधि में सबसे कम वर्षा होती है। ISMR के कैलिब्रेटेड मल्टी मॉडल सीजनल फोरकास्टिंग में नए शोध को नॉर्थ अमेरिकन मल्टी मॉडल एनसेंबल और आई एम डी के ग्रिडेड डेटा के आधार पर प्रस्तुत किया गया है, इसमें वर्ष 2018 के पूर्वानुमानों का केस स्टडी के रूप में उपयोग किया गया है; इन पूर्वानुमानों को वास्तविक समय में टरकाइल-श्रेणी के संभाव्यता प्रारूप में जारी किया गया था और जून से सितंबर तक प्रत्येक कैलेण्डर माह की शुरूआत में 2018 के मॉनसून ऋतु के लिए अद्यतन किया गया था। ISMR के उपमौसमी मल्टीमॉडल संभाव्यता पूर्वानुमान 2-3 सप्ताह की अवधि (8-21 दिनों के लीड टाइम) के लिए तैयार और विश्लेषित किए जाते हैं, जिसमें 2018 के मॉनसून आगमन का उदाहरण के लिए उपयोग किया गया है, इन 2-3 सप्ताह के ग्रिडेड ISMR पूर्वानुमानों के हिंडकास्ट कौशल को मॉनसून ऋतु के आरंभिक और बाद के चरणों में अधिकतम दिखाया है जो IMD डेटा के अनुभवजन्य निष्कर्षों के अनुरूप है। अंत में बहु कालिक पैमाने पर वर्षा की संगठित भिन्नता की व्याख्या करने के लिए भारत के मॉनसूनी स्टेशनों के संजाल में दैनिक वर्षा की परिवर्तिता के लिए हिडेन मार्कोव मॉडल (HMM) का उपयोग किया गया है जो भारतीय ग्रीष्मकालीन मॉनसून वर्षा की विशेषता को दर्शाता है।

ABSTRACT. This paper reviews research done by the authors and their collaborators at IRI and beyond over the past decade on predictability and prediction of Indian summer monsoon rainfall (ISMR) on seasonal and sub-seasonal timescales. Empirical analyses of the daily ISMR characteristics at local scales pertinent to agriculture, based on IMD gridded data, reveal that the number of rainy days in the season is more predictable than the seasonal rainfall total; furthermore, this "weather-within-climate" predictability undergoes an important seasonal modulation and is highest in the early and late phases of the monsoon and lowest in the July-August core monsoon period. New research in calibrated multi-model seasonal forecasting of ISMR is presented based on the North American Multi-Model Ensemble and gridded IMD data, using the 2018 forecasts as a case study; these forecasts were issued in real-time in tercile-category probability format and were updated for the remainder of the 2018 monsoon season at the beginning of each calendar month from June to September. Sub-seasonal multimodel probabilistic predictions of ISMR in the weeks 2-3 range (8-21 day lead times) are constructed and analyzed, using the onset of the 2018 monsoon as an example; the hindcast skill of these week 2-3 gridded ISMR forecasts is shown to be substantial in the early and late stages of the monsoon season, consistent with the empirical findings from IMD data. Lastly, a hidden Markov model (HMM) of daily rainfall variability at a network of stations over monsoonal India is used to interpret the organized variation of rainfall across the multiple temporal scales that characterize ISMR.

Key words – ISMR, ENSO, NMME, HMM.

1. Introduction

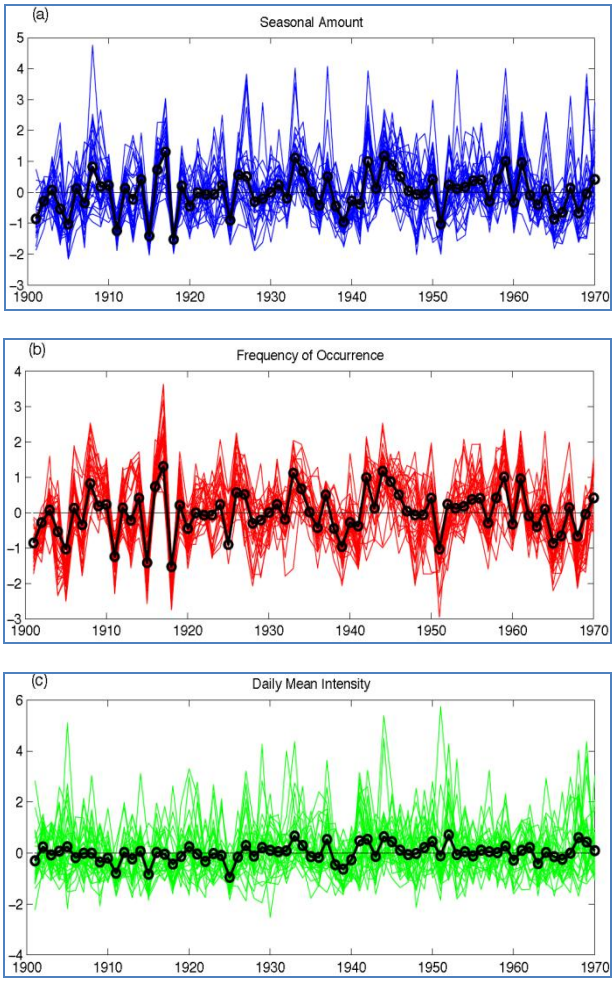
Meteorologists and climatologists have striven to the predict Indian summer monsoon rainfall (ISMR) for more than a century (Blanford, 1884; Walker, 1910), motivated originally by the disastrous famines of the late 19th century. More than a century later, its prediction still presents a formidable challenge and it has become

recognized that the problem is made complex by the many time and space scales on which monsoon rainfall varies (Gadgil, 2003). These range from the sub-daily scales of individual thunderstorms and mesoscale convective complexes (<100 km) to monsoon depressions with time scales of days (100-1000 km) to intraseasonal active and break phases of the monsoon simultaneously impacting much of the subcontinent. On time scales of a season and longer, sea surface temperature (SST) anomalies over the Pacific and Indian Oceans and the El Nino-Southern Oscillation (ENSO) in particular play important roles, but interannual predictability remains modest and the presence of decadal-scale variations has hindered its quantification (Kumar *et al*., 1999).

Forecasts between medium-range weather and seasonal climate are receiving increasing interest due to a better understanding of climate phenomena on sub-seasonal to seasonal (S2S) time-scales (from 2 weeks to a season ahead), modeling advances and their potential relevance to important societal decision-making lead times (Robertson and Vitart, 2019). The importance of the boreal summer intraseasonal oscillation (BSISO) (Lee *et al*., 2013) in causing active and break phases of the monsoon over the Indian sub-continent, together with recent increases in the skill of S2S forecast models in predicting the Madden-Julian Oscillation (Vitart, 2017), which is closely related to BSISO, has created optimism for ISMR forecast potential up to several weeks ahead. The sub-seasonal forecast time range is commonly referred to in India as the extended-range, up to 30 days ahead (Sahai *et al*., 2019).

From the end-user perspective, it is often the daily statistics of local rainfall that matter more than the all-India seasonal total. Monsoon onset and withdrawal dates along with the active and break phases can be critical from the agricultural point of view; this requires analysis of daily rainfall data in the context of the highly pronounced seasonal cycle of the monsoon.

Motivated by these multi-scale use-oriented considerations, this paper reviews research done by the authors and their collaborators over the past decade on seasonal and sub-seasonal (including weather within climate) rainfall predictability and prediction over India. Section 2 discusses the interannual predictability of monsoon daily rainfall characteristics at station scale, as revealed empirically from observational analyses of the spatial coherence of sub-seasonal to seasonal ISMR anomalies. In Section 3, we then consider seasonal forecast skill of summer monsoon precipitation derived from modern day ensemble prediction systems and document its real-time performance during the 2018 monsoon season. Section 4 presents sub-seasonal precipitation prediction in the weeks 2-3 range (8-21 day lead times) and the modulation of skill across the season, including the 2018 monsoon onset. Section 5 then summarizes results from stochastic models of rainfall variability at station scale, providing a bridge between the local daily scales of rainfall and the slower potentially more predictable sub-seasonal to seasonal and longer variation. A summary and conclusions are presented in Section 6.



Figs. 1(a-c). Interannual variability of rainfall characteristics at 28 stations over Gujarat and Rajastan, 1900-1970, based on data from the Global Daily Climatology Network (GDCN), archived at the National Climatological Data Center (NCDC). Each curve has been standardized by subtracting its mean and dividing by its standard deviation. The black curves show the 28-station average of the individual station standardized curves, *i.e*., the standardized anomaly index (SAI). (a) seasonal total rainfall, (b) daily rainfall frequency (number of wet days receiving at least 1 mm) and (c) mean daily intensity (mean amount falling on wet days)

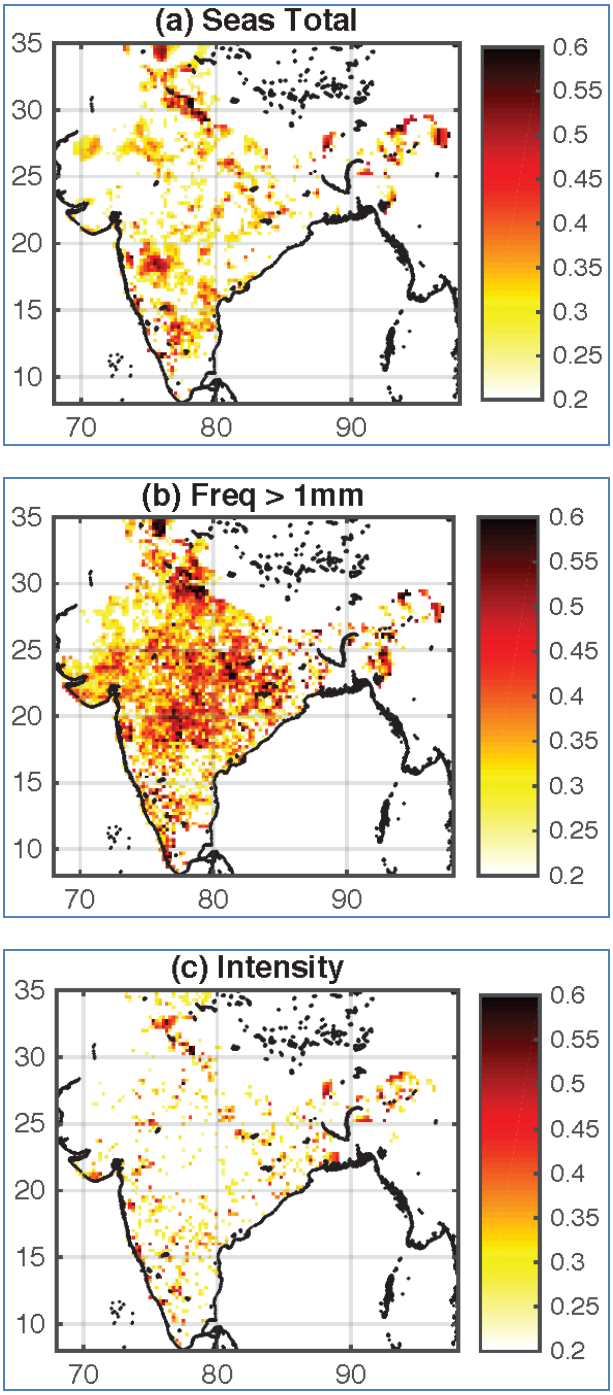
2. Empirical analyses of ISMR characteristics at local scales

Seasonal average rainfall anomalies can vary greatly from location to location, as illustrated in Fig. 1(a), which shows the year-to-year variability (1900-1970) of the June-September rainfall averages for a set of 28 stations in Gujarat and Rajasthan. There is considerable spread between the stations, even though the anomalies have been standardized in the figure by subtracting the station average from each station and dividing by its standard deviation. Nonetheless, a spatially-coherent interannual "signal" emerges in Fig. 1(a), defined by the spatial average of standardized anomalies, *i.e.*, the standardized anomaly index (SAI) (Moron *et al.*, 2006, 2007). The spatially-coherent interannual signal is more marked in some years than in others and thus might be expected to be related to the climate-drivers of monsoon variations; indeed, the SAI in Fig. 1(a) is significantly correlated with ENSO.

The interannual variance of the SAI (var[SAI]) provides an empirical measure of the spatial coherence of the seasonal anomalies between stations; var[SAI] ranges between near-zero for station-independent variations to near 1 for identical ones and is closely related to the number of spatial degrees of freedom (DoF). The DoF generally ranges between 1 for perfectly co-varying station variations to the number of stations in the network for independent variations (Moron *et al.*, 2006; 2007).

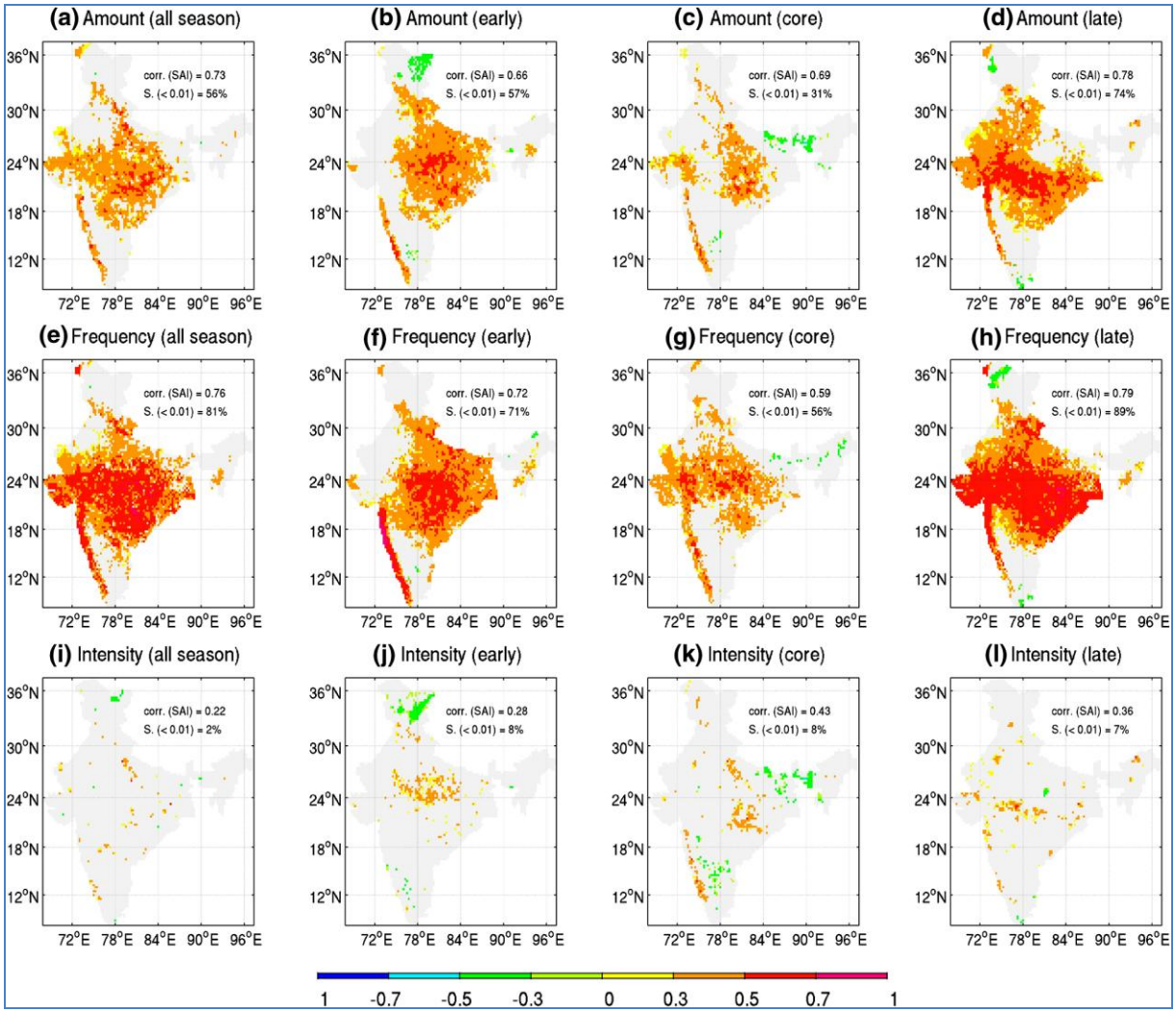
The seasonal total of rainfall (S) at a location can be factored into the product of number of days with rain (frequency of occurrence of rainy days, O) and the average amount of rainfall falling on rainy days (mean intensity, I) across the season[1]. Figs. 1(b&c) shows similar station time series for O and I. The spatial coherence of seasonal anomalies of O between stations is visibly larger than for S, while it is very much smaller for I. The number of spatial degrees of freedom of O, S and I equal respectively 2.7, 5.3 and 14.1 (noting that the DoF of 28 time series of white noise over 71 yearly samples is ~ 20 (instead of 28) due to the finite length of the time series). Thus, the spatial coherence between stations is clearly intermediate for S, while it is highest for O.

[1]Note that I should not be strictly identified with instantaneous rain rate since the wet events last usually far less than a day. In consequence I depends on the instantaneous rain rate as well as on the duration of the wet events during the day. Nevertheless, the interannual variations of I may mostly reflect those of instantaneous rain rate, which will be strongly impacted by the extreme rainfall, assuming the mean duration of wet events does not vary too much across the years. Wet days are often defined with a threshold of 1 mm to exclude inconsequential daily amounts.



Figs. 2(a-c). Predictability of Jun-Sep seasonal rainfall characteristics. Anomaly correlation skill of (a) seasonal total rainfall, (b) daily rainfall frequency and (c) mean daily intensity, as a function of simultaneous Indo-Pacific SSTs. Rainy days are defined as days with >1 mm rainfall. Only grid points that are statistically significant at the 99% confidence level are plotted

The differing spatial coherence characteristics of S, O and I reflect differences in their potential seasonal



Figs. 3(a-l). Correlations between time-averaged Monsoon Circulation Index (MCI) given by the first principal component of an EOF of unfiltered daily winds at 850 hPa, precipitable water and vertical velocity at 500 hPa from NCEP Reanalyses (from May 16, 1948 to October 31, 2014) and gridded rainfall amount, frequency and mean intensity (rows). The correlations are calculated separately for the whole May 16-October 31 season and its 3 sub-seasonal stages (columns): early (May 16-June 30), core (July 1-August 31) and late (September 1-October 31). The values in the upper right corner of each panel show the correlation value between MCI and the SAI of the Monsoonal India (MI)-averaged rainfall characteristic (corr. SAI) and the fraction of MI area with significant cor-relations at p = 0.01 according to a random-phase test (S < 0.01). Insignificant correlations at p = 0.01 are masked out as light gray. Reproduced from Moron *et al*. (2017)

predictability across monsoonal India, assuming that spatially-coherent variations (maximum for O) are a prerequisite for climate predictability while incoherent variations (as for I) imply negligible potential predictability. This is demonstrated in Figs. 2(a-c) which show the anomaly correlation skill of "forecasts" constructed by regressing fields of June to September contemporaneous Indo-Pacific sea surface temperatures (SST) against gridded fields of 0.25° rainfall across India.

In Figs. 2(a-c) a cross-validated canonical correlation analysis (CCA) approach is used to "predict" the chosen Jun-Sep daily rainfall statistic based on tropical Indo-Pacific SSTs averaged over the same season. This method is used to quantify the extent to which seasonal anomalies of Indian summer monsoon rainfall can be inferred from contemporaneous seasonal anomalies of SST and provides an upper bound on the SST-related potential predictability. The CCA regularizes the high-dimensional regression problem between a spatial field of predictors (here SST fields) and predictands (here the seasonal rainfall statistic) by reducing the spatial dimensionality via principal component (PC) analysis and thus minimizes problems of over-fitting and multi-colinearity (Tippett *et al*., 2003). Here we truncated the PC expansions of the SST and rainfall-statistic fields so as to retain 90% of the

variance of each. The SSTs were limited to the [40° E-70° W, 30° S-30° N] domain and all 4964 of the 0.25° rainfall grids over India were selected. The skill of the rainfall-SST regression relationship of S, O and I clearly reflects the differences in the spatial coherence of the seasonal anomalies between the three quantities: the seasonal predictability of O is highest, while that of I is near-zero. The rather low predictability of the seasonal rainfall total can thus partly be explained as a result of the "contamination" of the signal (reflected in O) by unpredictable small-scale noise in the intensity component. The impact of extreme daily rainfall amounts on seasonal forecast skill has been discussed by Stephenson *et al*. (1999).

An additional analysis of rainfall potential predictability at near-local scales is shown in Figs. 3(a-l), which shows how the relationship between the leading regional-scale empirical atmospheric mode of variation and near-local scale rainfall evolves in time between the early, core and late stages of the SW monsoon season, defined here as early (May 16-June 30), core (July 1-August 31) and late (September 1-October 31). The strength of the relationship weakens considerably during the core phase of the monsoon. This is true for both S and O while the cross-scale relationship is very weak in all monsoon stages for I. As discussed in Moron *et al.* (2017), the drop in spatial coherence during the core phase may be related to increases in the number and intensity of wet "patches" of daily rainfall while their mean spatial extent remains similar throughout the monsoon season. The core stage, with large values of precipitable water, is characterized by very variable daily rainfall amounts across grid points, as a consequence of the near exponential distribution of daily rainfall.

To summarize these empirical findings from historical daily rainfall data, the potential S2S predictability of station-scale rainfall is larger for daily rainfall frequency than for monthly or seasonal rainfall total. In addition, this potential predictability is not constant through the monsoon season, but is larger in the early and late parts of the season. In the next two sections we present results from multi-model dynamical ensemble prediction systems, first for seasonal forecasts and then for sub-seasonal ones.

3. Seasonal forecasts from the North American Multi-Model Ensemble (NMME)

3.1. *Methodology*

General circulation models (GCMs) represent the physical processes within the atmosphere and have been used increasingly to predict ISMR in recent decades.

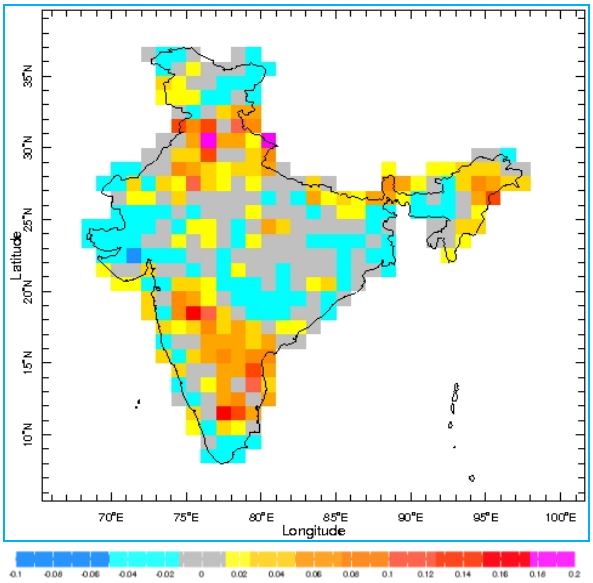
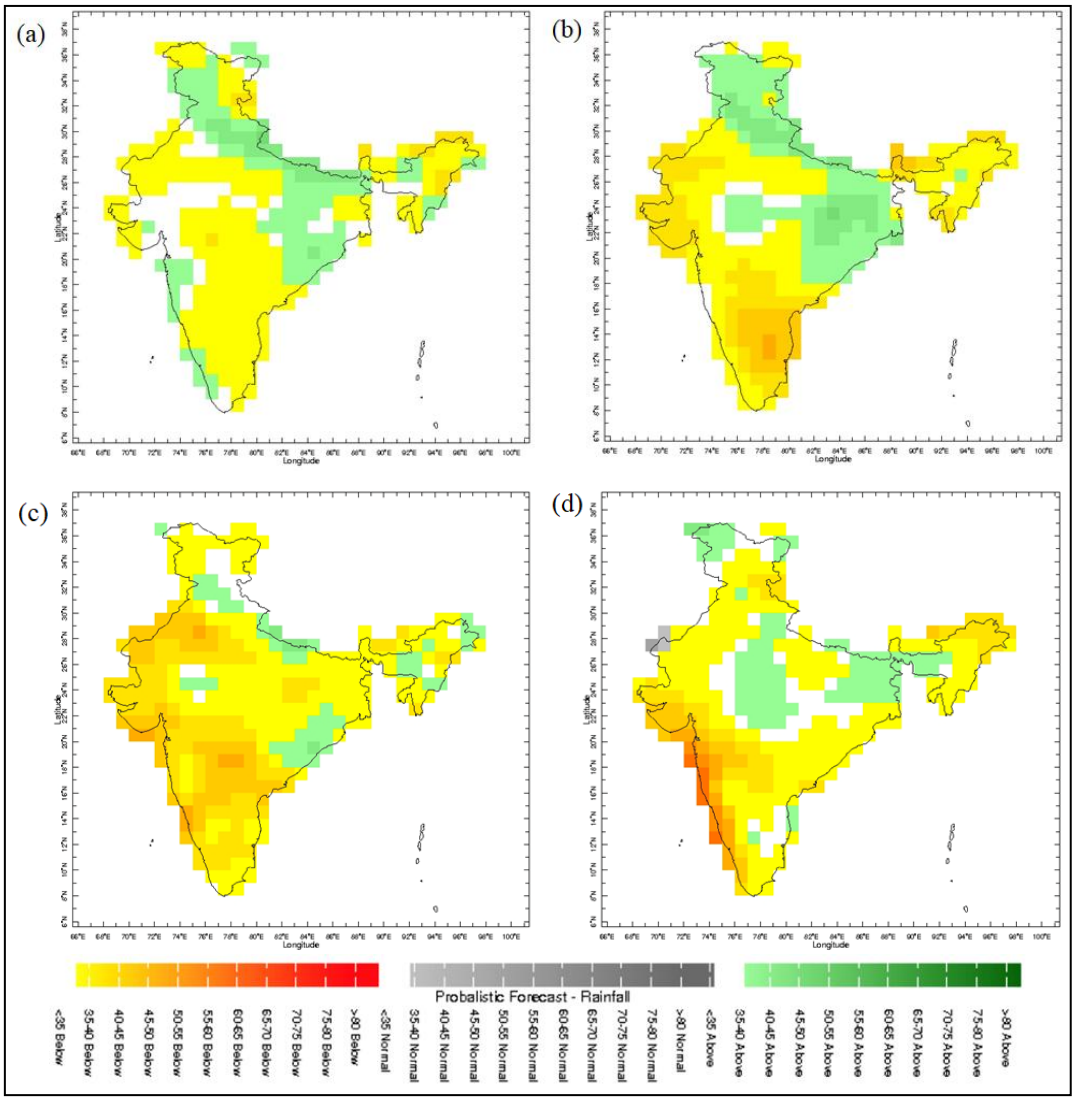


Fig. 4. Seasonal forecast Ranked Probability Skill Score (RPSS) of calibrated MME forecasts for June-September precipitation, from early June, based on 1982-2010 hindcasts

Several recent studies have analyzed GCM performance in terms of ISMR, revealing large biases (systematic and random) in simulating the inter-annual variability of ISMR (Acharya *et al.*, 2011; 2014b). Acharya *et al.* (2013a) compared different bias correction methods on GCM outputs and concluded that two methods, quantile-quantile mapping and standardized-reconstruction techniques, performed best among them. With the availability of climate predictions produced by multiple GCMs, multimodel ensemble (MME) forecasting has drawn attention recently to predict ISMR. Several approaches have been attempted to combine multiple GCM forecasts into a single MME forecast of ISMR with higher skills than the individual models. These include the simple ensemble mean, singular value decomposition (SVD)-based multiple linear regression (Kar *et al.*, 2012), supervised principal component regression (SPCR) (Nair *et al.*, 2013) and CCA (Singh *et al.*, 2013).

Most of the techniques mentioned in the previous paragraph generate a deterministic forecast without an explicit measure of its inherent uncertainty. However, beyond the weather scale especially, a single deterministic rainfall forecast is not sufficient. The user community should be given probabilistic forecasts that quantify the uncertainty within the prediction. Fewer studies have described probabilistic prediction of ISMR at monthly and seasonal scale (Kulkarni *et al.*, 2012; Acharya *et al.*, 2013a; 2014b). These studies use a non-parametric or parametric approach with respect to the occurrence of three categories of seasonal total rainfall below, near and above normal as defined by the climatological base



Figs. 5(a-d). Seasonal precipitation 2018 forecast probabilities for (a) June-September, (b) July-September, (c) August-September & (d) September. The forecasts are initialized near the start of the first month of each forecast period

period. The non-parametric method (Acharya *et al.*, 2013a) commonly uses the "counting method" to estimate the forecast probabilities, *i.e.*, the percentage of GCM ensemble members falling into each tercile category gives the forecast probability, whereas the parametric method (Kulkarni *et al.*, 2012; Acharya *et al.*, 2014a) converts a deterministic forecast from a GCM into probabilities by fitting a probability distribution function (PDF): the mean and variance of the distribution is determined from GCM outputs (Tippett *et al.*, 2007). Acharya *et al.* (2013b) describe and compare a number of different schemes for making a non-parametric probabilistic MME to predict ISMR. The imprecision of the counting method is a drawback when the ensemble size is small. In addition, the ensemble spread is typically underestimated resulting in over-confident forecast probabilities (Becker and van den Dool, 2015). Most studies adopting the parametric method to make probabilistic predictions of ISMR (Kar *et al.*, 2012; Kulkarni *et al.*, 2012; Acharya *et al.*, 2014a), convert the deterministic prediction into probabilities assuming a Gaussian PDF. The variance of the distribution can be estimated using either the ensemble spread (variance within the MME), error residuals (error between the MME mean and corresponding observation), or the correlation method (based on anomaly correlation between the MME mean and corresponding observation) (Kulkarni *et al*., 2012).

Since neither observations of seasonal ISMR nor the models forecasts follow a Gaussian distribution, logistic regression (LR) is an alternative approach (Hamill *et al*., 2004). LR is a nonlinear regression method where

probability itself can be considered as the predictand rather than a measurable physical quantity. Extension of LR by including the predictand threshold in the regression equation as an additional predictor (Wilks, 2009), namely Extended Logistic Regression (ELR), allows the derivation of full predictive distributions and avoids the problem of potentially inconsistent forecast probabilities for different quantiles. In this study, an ELR-based calibration method is implemented in models from the North American Multi-Model Ensemble (NMME) project (Kirtman *et al*., 2014) with gridded rainfall data from IMD. Though the ELR method has been successfully applied in the past to ensemble weather and sub-seasonal forecast, this is the first time to our knowledge that it has been used to produce seasonal probabilistic forecast of ISMR. The ELR method also allows generating the forecast in more flexible format in addition to commonly used tercile probability forecast. The flexible format enables users to extract information for those parts of the forecast distribution of greatest interest to them, especially the probability of extremely dry/wet conditions.

The lead-0 (using initial conditions of June for forecasting the mean rainfall of June-September) hindcast runs (spanning 1982-2010) from nine NMME models were used; their basic characteristics and references are discussed in Barnston and Tippett (2017). These NMME monthly real-time forecast and hindcast datasets are available on a common 1° grid at http://iridl.ldeo.columbia.edu/SOURCES/.Models/.NMME. The 1° gridded daily rainfall dataset from IMD is used as observational reference. Rajeevan *et al.* (2006) describe the detailed procedure to generate such gridded products from rain gauge stations.

The ELR based calibration method is implemented in the following way to obtain the probabilities of the occurrence of tercile categories, *viz*., below, near and above normal of seasonal mean ISMR. The ELR was fitted to each NMME model's predicted precipitation dataset together with the observed as:

$$\ln\left[\frac{p}{1-p}\right] = f(x) + g(q) \qquad (1)$$

with $f(x) = b_0 + b_1\overline{x_{ens}}$ and $g(q) = b_2 q$

Here, p is the cumulative probability of not exceeding the quantile q, b are regression coefficients and $\overline{x_{ens}}$ is the mean of all ensemble members for an individual model.

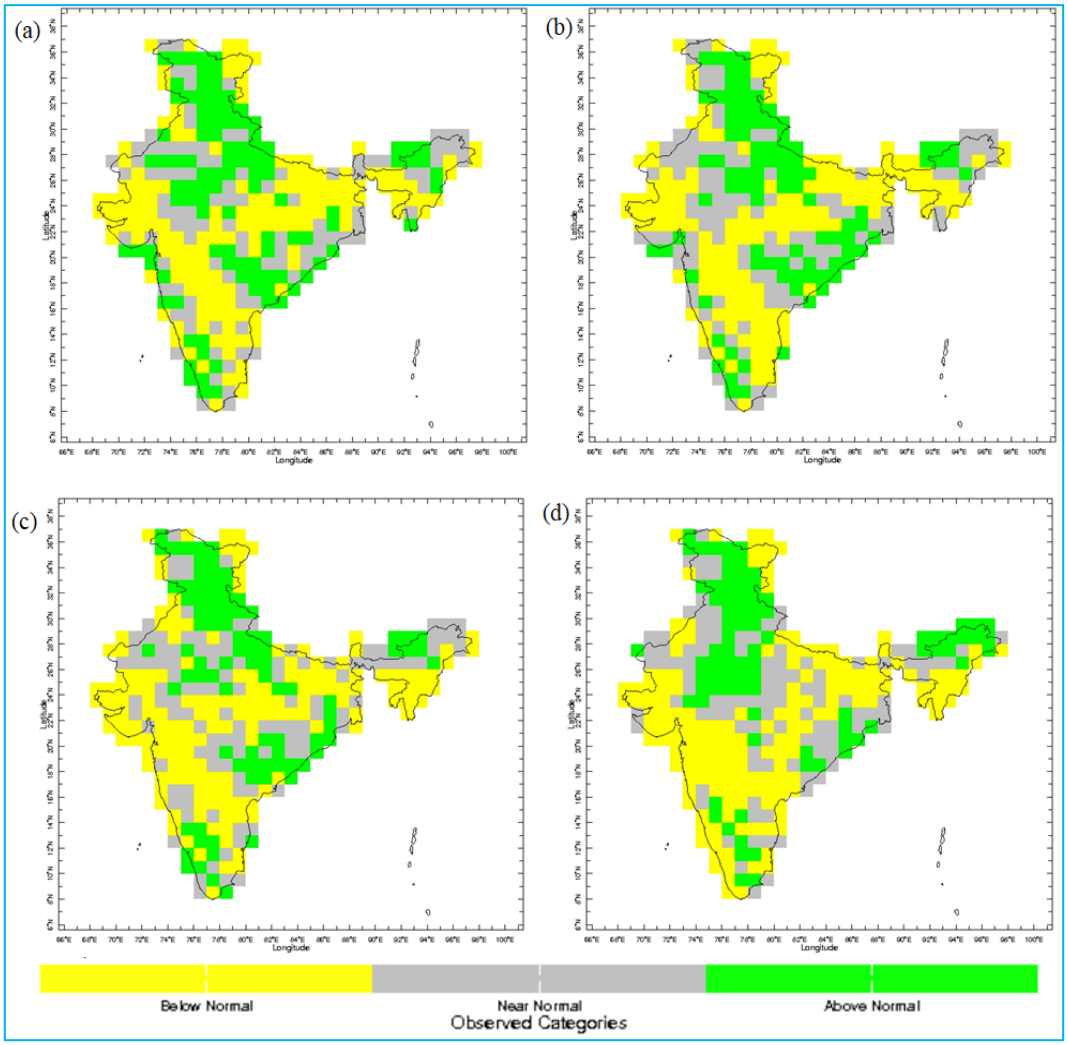
These resultant tercile probabilities from each of individual GCM are then averaged together with equal weight to obtain the final multi-model ensemble (MME) forecast. The above-developed prediction scheme is applied in a leave-one-out cross-validation manner. Since the dataset contains only 29 years data (1982-2010), one year is omitted and the model is developed for the rest of the 28 years and the predictand variable is estimated for the omitted year. In this way, the cross-validated series for the predictand variable is generated and then validated against the observed rainfall data using skill scores mention in next section.

3.2. *Skill assessment and 2018 monsoon seasonal forecasts*

The skill of the ELR-based NMME prediction is evaluated against observations using the ranked probability skill score (RPSS) (Weigel *et al*., 2007), a proper score, used to quantify to which extent the probabilistic predictions are improved upon the climatological forecast of equal probabilities for each tercile category (*i.e*., 1/3 for each tercile category). A positive (negative) value of RPSS indicates that the prediction system is better (worse) than the climatological guess, while zero implies no improvement over the climatological forecast. The average RPSS over the 29 years of the calibrated forecast is shown in Fig. 4. Subndstantial positive RPSS values are achieved south of 20° N over central peninsula India and over the northern Indo-Gangetic plain near 30° N. However, skill is generally low across the core monsoon zone.

For the monsoon 2018, this newly developed prediction system has been tested experimentally. The forecast was provided to IMD at the beginning of the season (early June) for the June-September, with monthly updates issued at the beginning of each month for the remainder of the season (*i.e*., July-September, August-September and September). These probability forecasts are shown in Figs. 5(a-d). Each map shows the forecast probability of the tercile-category that is dominant in the forecast, on a 1° grid; yellow-brown areas show a tilt of the odds toward the below-normal category, with above-normal forecasts in greens-blues. The forecasts in Figs. 5(a-d) indicate increased odds of below-normal rainfall in the western parts of India and for above-normal over the Gangetic plain to the east and Deccan plateau to its south. The monthly forecast updates (panels b-d) are quite similar in this overall pattern to the June forecast for the season as a whole, although the dry probabilities in the west are enhanced in the updates.

How successful were these forecasts for 2018? The tercile category of the observed rainfall at each grid point for the respective forecast periods are shown in Figs. 6(a-d). These verification maps were computed



Figs. 6(a-d). Observed 2018 precipitation categories for (a) June-September, (b) July-September, (c) August-September and (d) September

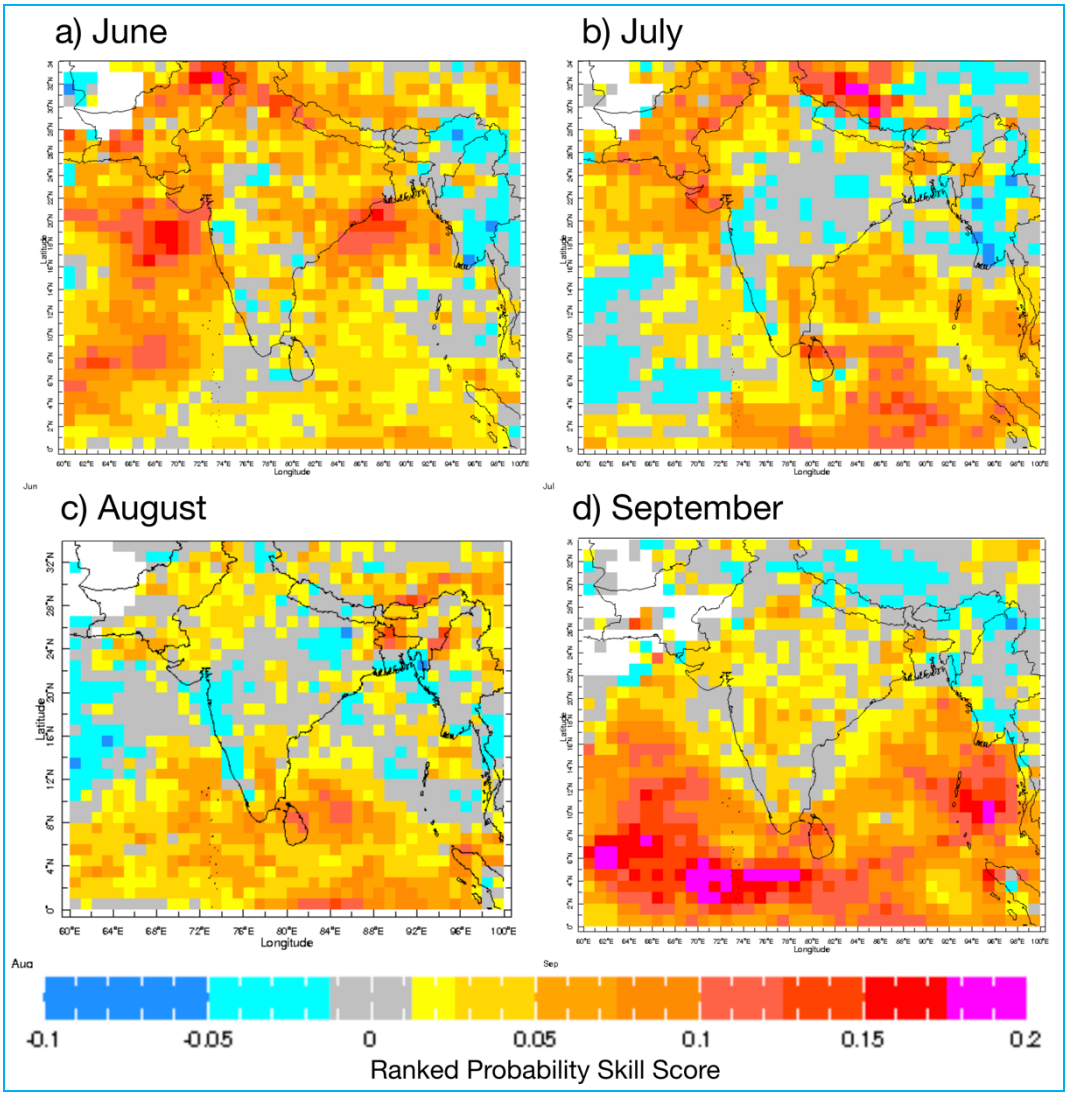
using NCEP CPC unified gridded data on the same 1° grid as for the forecasts, with the category thresholds calculated based on historical data for the 1982-2010 period. As with the updated forecasts in Figs. 5(a-d), the verification maps are quite similar to one another. They exhibit broad similarities with the forecasts which captured the below-normal rainfall conditions in western parts of the peninsula and above-normal rainfall over the Deccan plateau and northern states. On the other hand the dry conditions over Bihar were not well captured.

To summarize, probabilistic seasonal forecasts of monsoon rainfall at 1° gridded scale based on the NMME set of models and calibrated with extended logistic regression show promise for some regions of India as evidenced here by hindcast skill estimates, as well as by their performance during 2018.

4. Subseasonal forecasts from S2S models

Despite some encouraging indications from earlier studies (Vitart and Molteni, 2009), the sub-seasonal forecast skill in gridded precipitation over India during the summer monsoon season as a whole was found by Li and Robertson (2015) to be relatively low beyond a week ahead. Nonetheless, recent advances reported by Sahai *et al.* (2019) indicate considerable skill in the latest IMD extended range ensemble prediction system.

We have examined whether sub-seasonal probabilistic forecast skill can be enhanced through effective calibration and multi-model ensemble techniques over North America (Vigaud *et al.*, 2017a), boreal summer monsoon regions (Vigaud *et al.*, 2017b) and the

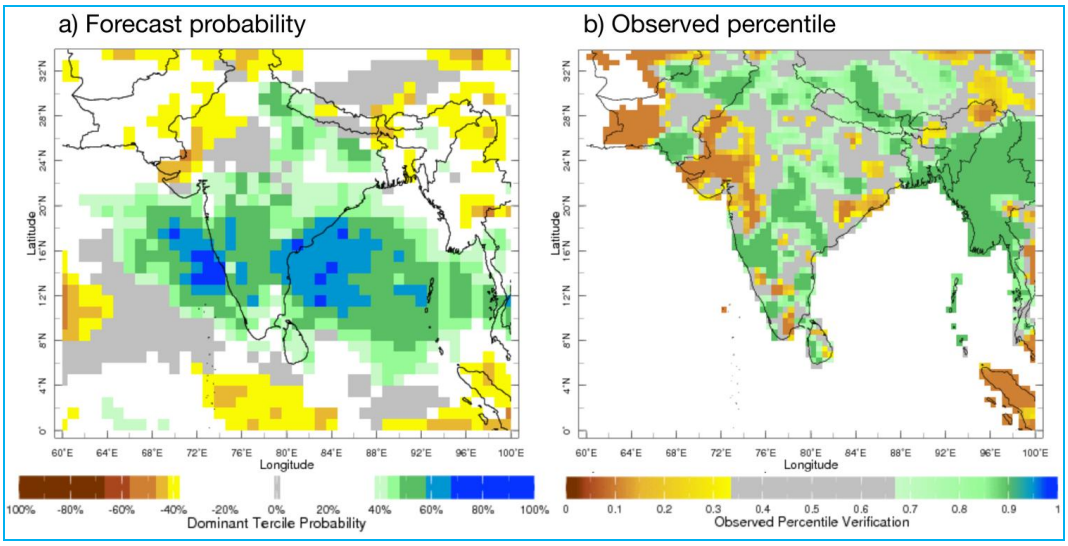


Figs. 7(a-d). Ranked Probability Skill Score (RPSS) for weeks 2-3 hindcasts starting in June, July, August and September

East Africa-West Asia sector (Vigaud *et al*., 2018), as has been previously demonstrated for seasonal (Robertson *et al*., 2004b) and medium range (Hamill, 2012) forecasts. Extended logistic regression (ELR, as described in Section 3) was used to obtain the forecast directly as a probability of exceeding/non-exceeding a chosen quantile, as a function of the GCM's ensemble mean forecast separately for each model, grid point, start and lead. As shown in Vigaud *et al.* (2017a), when applied to the 0.33 and 0.67 quantiles to produce weekly precipitation tercile probabilities, this approach has the advantage of producing logically consistent sets of forecasts, in the sense that cumulative probabilities for smaller predictand thresholds do not exceed those for larger thresholds (Wilks, 2009).

4.1. *Skill assessment*

Figs. 7(a-d) shows the ELR-calibrated skill for sub-seasonal hindcasts, obtained from a 3-model multi-model ensemble of the ECMWF, NCEP CFSv2 and CMA models from the S2S Database (Vitart *et al*., 2017), using the methodology described in Vigaud *et al*. (2017a). The skill is expressed using RPSS, at a forecast lead time of 2-3 weeks ahead (bi-weekly target periods from [d+8, d+21] for a forecast issued on day d); the different panels show the skill separately for forecast weekly starts in each calendar month of the June-September season. Generally over India, there is substantial and widespread positive RPSS skill for forecast starts in June and September, covering most of India, but with a pronounced decrease in



Figs. 8(a&b). (a) Week 2-3 forecast from May 24, 2018, shown in terms of the probability of the dominant forecast tercile category. (b) Percentile of the observed two-week precipitation for the target period May 31 - June 13, 2018, calculated with respect to the 1999-2010 period

July and August. This seasonal modulation of the sub-seasonal skill with a drop in July-August is consistent with the drop in the spatial coherence and thus potential predictability of precipitation anomalies during the core phase of the monsoon in July-August, as calculated empirically from data in Figs. 3(a-l).

Higher RPSS values on land during June and September suggest potentially higher skill during the onset and cessation of the monsoon rather than during the core season (Vigaud *et al*. 2017b, Moron *et al*., 2017). At this stage however, skill and predictability of onset date are yet to be explicitly examined. We can hypothesize that a weak-to-moderate signal associated with emerging El Niño conditions (*i.e*., delayed local-scale onset, mostly across central India is expected with warm ENSO and anomalously warm Indian ocean (Moron and Robertson, 2013) may combine with the MJO/BSISO signal, making the onset partly predictable in years when these two signals are pronounced.

4.2. *Sub-seasonal forecast for May 31-June 13, 2018, initialized on May 24*

Figs. 8(a&b) shows the forecast for the May 31-June 13 fortnight period, derived from the 3-model S2S MME forecast initialized on Thursday May 24, 2018 (*i.e*., weeks 2-3 lead). The map shows the forecast probability of the tercile-category that is dominant in the forecast, on a 1.5° grid [similar in format to Figs. 5(a-d)]. The forecast indicates an enhanced probability of above-normal precipitation of 50-60% (compared to 33%) over much of the Indian peninsula, with 60-80% probabilities over the Bay of Bengal and over the western Arabian Sea. These enhanced probabilities coincide with the onset phase of the monsoon over the Indian peninsula, as is indicated in IMD's assessment (Fig. 9). In regards to predictability sources for this event, Figs. 10(a&b) shows the observed evolution of the MJO, which in boreal summer is closely related to the BSISO, during 24 May - 7 July in the plane of the RMM indices (Wheeler & Hendon, 2004), along with that of the ECMWF model ensemble mean initialized on Thursday May 24. The strong MJO activity observed from phase 3 at forecast start to phase 6 prior to the week 2-3 forecast target period is well captured by the ECMWF model. This MJO evolution is consistent with the observed northward progression of monsoon onset. Vigaud *et al*. (2017b) found that MJO phase 3 at the forecast start, when convection is increased over the west Indian Ocean, was associated with substantial skill at week 3-4 leads. This together with the drop in MJO strength in phase 6 during the target period, when convection is enhanced over the western Pacific but reduced over Indian monsoon regions, could explain the skill in forecasting the timing of rain onset shown in Figs. 8(a&b).

To summarize this section, the skill of a sub-seasonal forecast MME constructed from three S2S models shows substantial skill across much of India in the early and late phases of monsoon evolution, consistent with the empirical estimates of predictability derived in Section 2. In particular, the onset phase of the monsoon in 2018 is well captured by the MME forecast, which appears attributable, atleast in part, to the successful forecast of the MJO propagation (and implied northward progression of BSISO over India) by the ECMWF model.

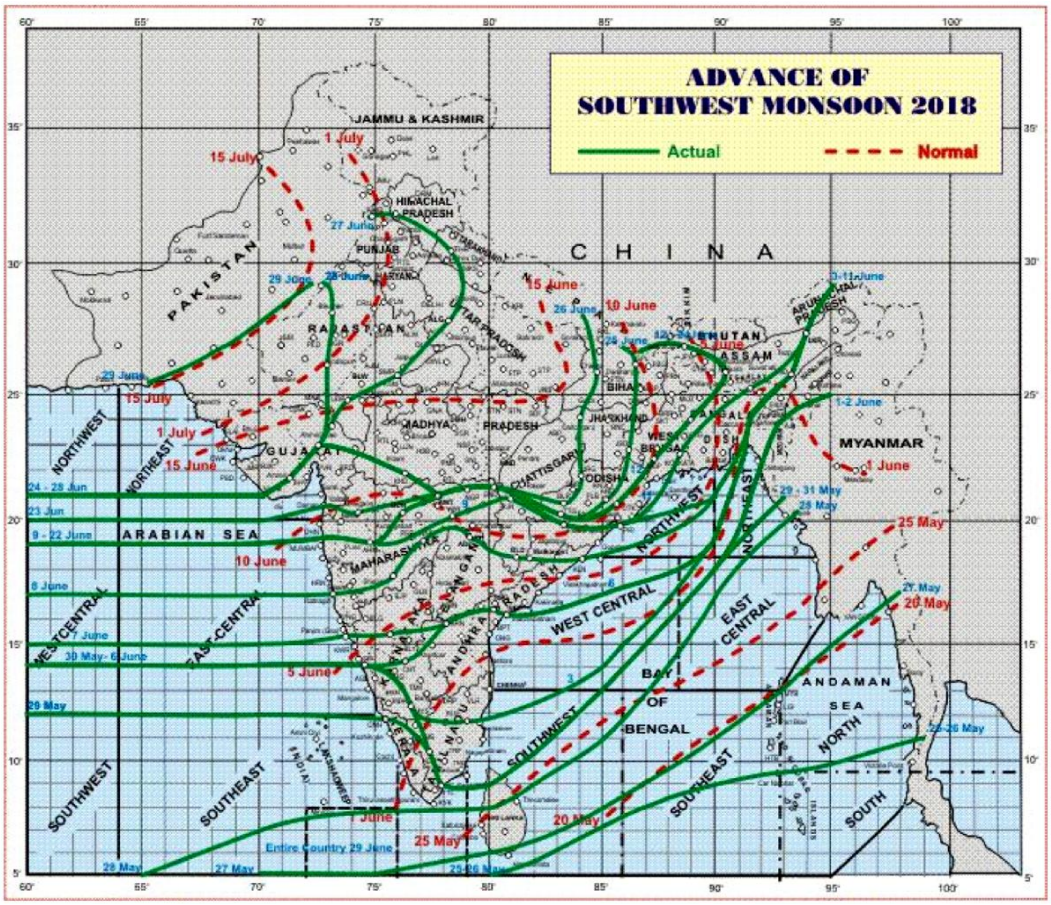
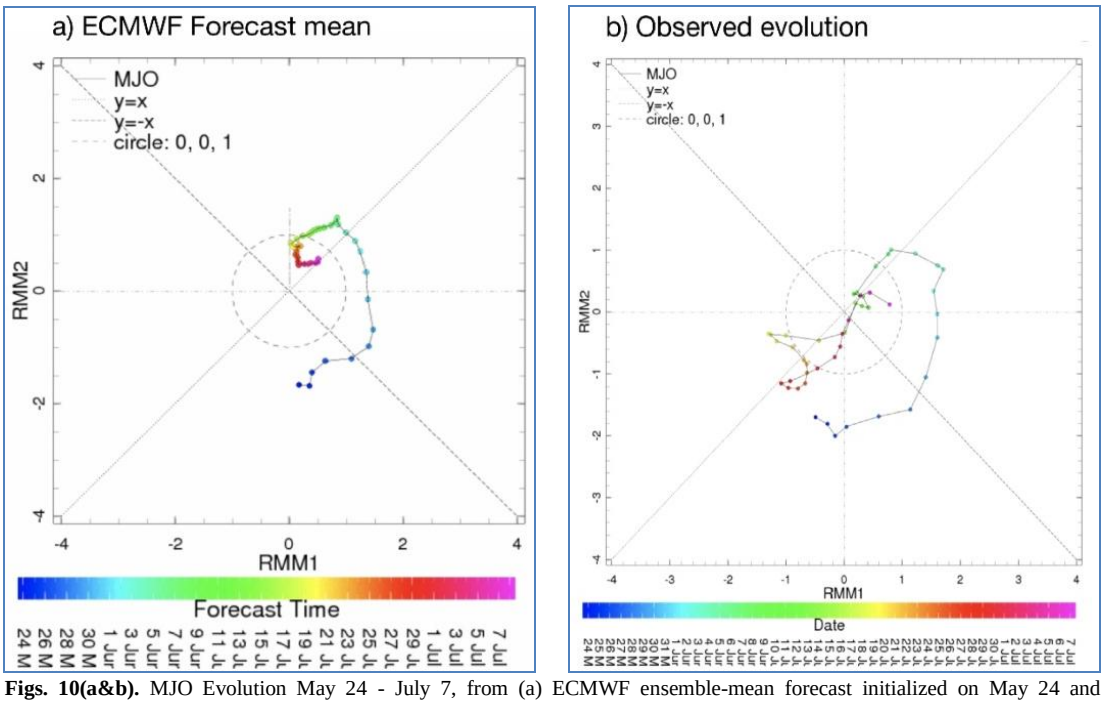
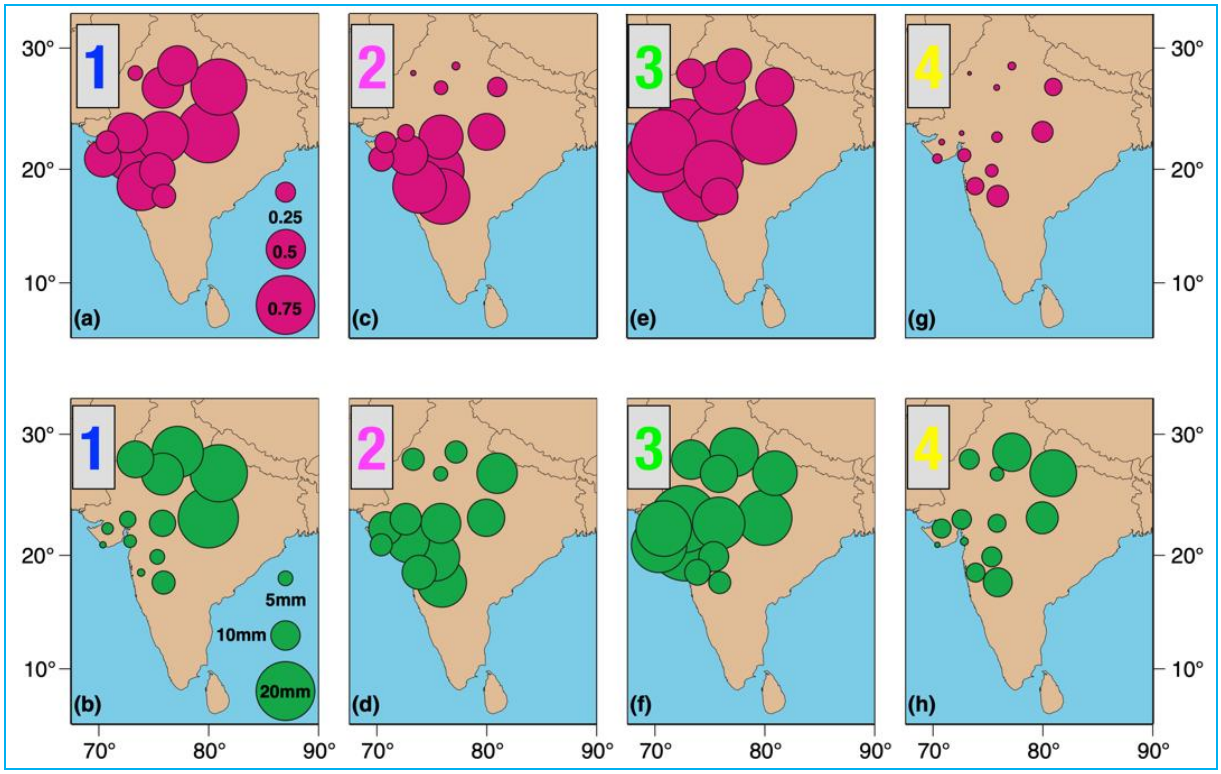


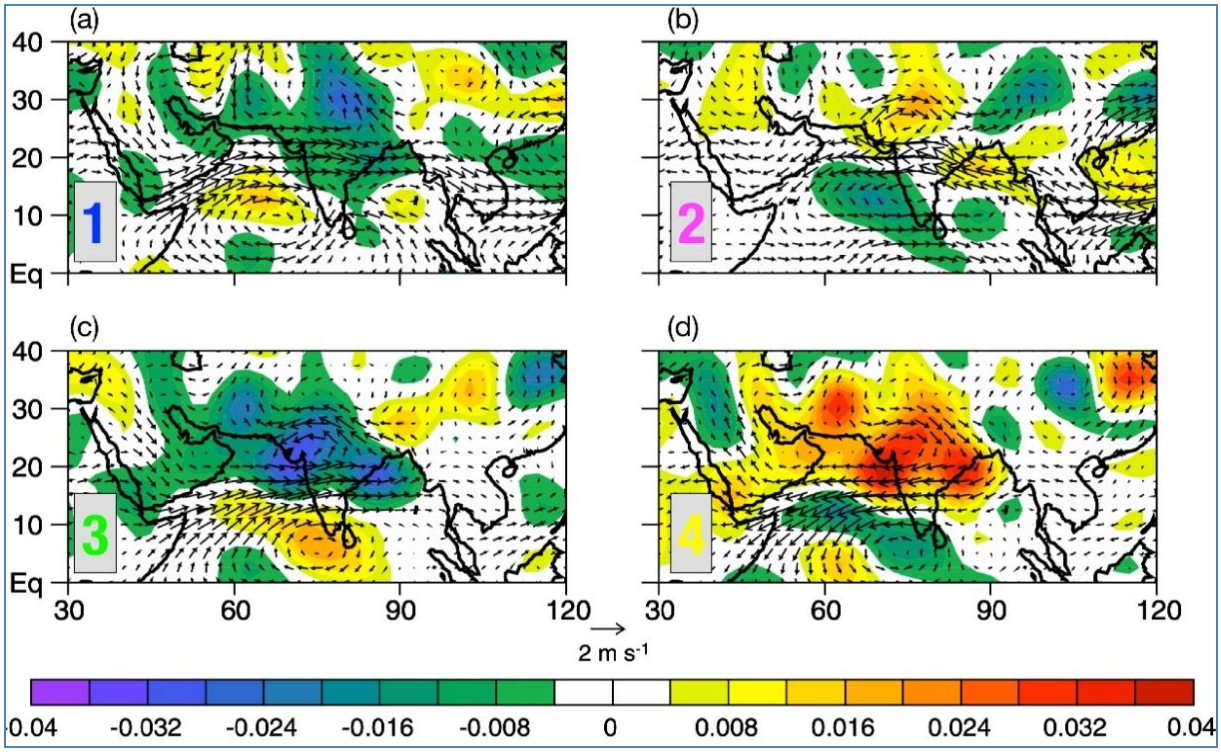
Fig. 9. Monsoon onset progression during 2018



Figs. 10(a&b). MJO Evolution May 24 - July 7, from (a) ECMWF ensemble-mean forecast initialized on May 24 and (b) Australian Bureau of Meteorology observed estimates



Figs. 11(a-h). HMM rainfall occurrence probabilities (top row) and mean daily intensities (bottom, in mm), for each of the 4 states, ordered from left to right. Values are computed conditional on a daily amount of at least 0.1 mm. The legend in the leftmost panel applies to all plots on that row



Figs. 12(a-d). HMM meteorological anomaly composites for each state, with respect to the mean June-September mean (1951-1970). Shown are winds at 850 hPa (vectors, ms^{1}) and 500 hPa vertical velocity (colors, Pas^{1}), anomalous ascent regions are shaded green-blue

5. Stochastic models of station rainfall variability

5.1. *A Homogeneous Hidden Markov model for Indian summer monsoon rainfall*

In this section we highlight the application of a stochastic model of daily rainfall as a tool to provide further insight into the S2S predictability of ISMR at the station scale and as a downscaling tool. The hidden Markov model (HMM) is a class of dynamic Bayesian networks, made popular in speech recognition but later applied to model daily rainfall due to its probabilistic nature and Markovian property (Hughes and Guttorp, 1994; Norris, 1997; Robertson *et al*., 2004a).

The HMM combines features of cluster and mixture models and adds a temporal dimension, which, in the present analysis, proceeds on a daily time step. The model groups together days having similar daily rainfall patterns across a network of stations and characterized as sets of daily rainfall distributions. The patterns, which are observed, are conceptualized as expressions of "hidden states", which themselves cannot be directly observed, but whose characteristics must be inferred from the observed data. The states progress from day to day with a Markovian dependence, the temporal evolution of daily rainfall also constituting inferential "evidence" used to fit the model. The HMM thus comprises both the distributional patterns and a state transition matrix; rainfall, in effect, is modeled as a "mixture of clusters" unfolding in time in Markovian fashion.

In Greene *et al*. (2008) a four-state HMM is fit to a 70-year record of daily Indian summer monsoon rainfall over a network of 13 stations, the rainfall data obtained from the Global Historical Climate Network (Legates and Willmott, 1990). Figs. [11(a-h)&12(a-d)] show, respectively, the state decomposition in terms of rainfall probabilities and intensities and atmospheric analyses, in the form of composited 850 hPa horizontal winds and 500 hPa vertical velocity, corresponding to these states. (Atmospheric data from the NCEP-NCAR reanalysis was utilized.) Clear patterns are evident, the most obvious contrast occurring between states 3 and 4, which capture overall wet and dry conditions, respectively [Figs. 11(e&f) and 12(c) for state 3 and Figs. 11(g&h) and 12(d) for state 4]. The "dry" state, characterized by anticyclonic circulation and subsidence over most of the Indian subcontinent, was shown in Greene *et al*. (2008) to toward the Himalayan escarpment, with high rainfall probabilities (but, apparently paradoxically, low mean amounts) to the southwest. In fact states 1 and 2, taken together, are consistent with northward propagation correspond well with monsoon breaks. State 2 shows a strong S-N gradient, while state 1 shows heavy rainfall characteristic of BSISO. Onshore winds impinging on the western escarpment, along with some ascent, in state 1 [Fig. 12(b)] help to explain the seemingly paradoxical probability-intensity relation.

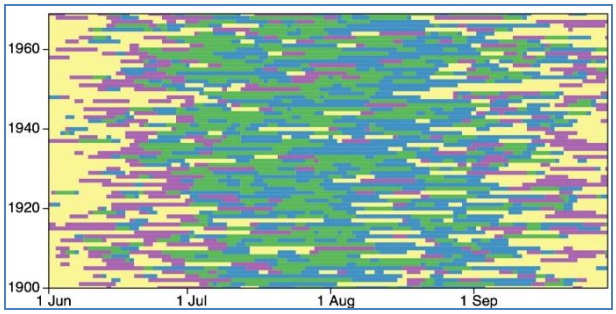
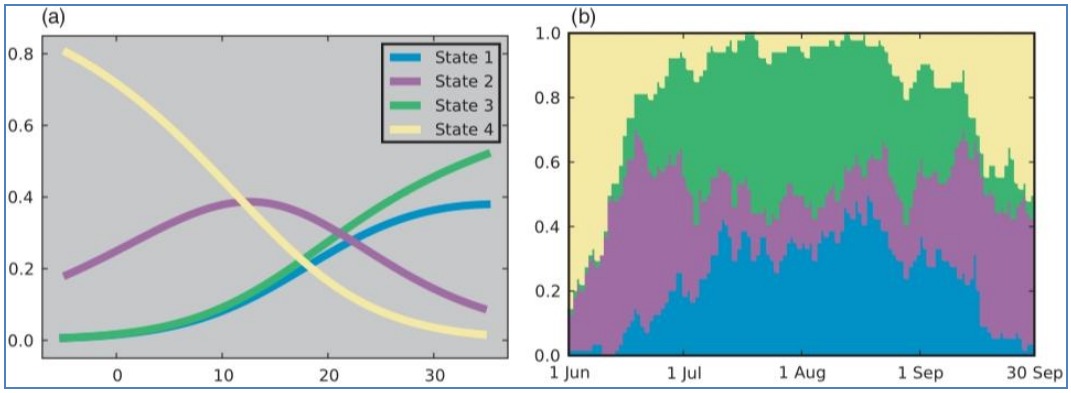


Fig.13. HMM daily sequence of the 4 hidden states in Fig. 11. From Greene *et al*. (2008).

The daily time sequence of the four hidden rainfall states is shown in Fig. 13, where the rows show the June 1-September 30 daily evolution for each monsoon season 1900-1970. The multiple time scales of the monsoon are made strikingly visible in this plot, highlighting the monsoon's seasonality and intraseasonal active and break phases, in terms of the time sequence of the 4 rainfall states, together with interannual monsoon variations which are expressed as changes from year to year in the frequencies of the states. A similar approach was taken by Pal *et al*. (2014) to model winter rainfall in Northwest India, illuminating the role of Western Disturbances, as well as their dynamics.

5.2. *Extension: A non-homogeneous HMM*

As described above, the HMM includes a matrix whose entries are the daily state transition probabilities. That this matrix is time-invariant means that the corresponding relative state frequencies represent averages over the monsoon season as a whole, *i.e*., they are also, of necessity, time-invariant. While this is certainly a useful model, as we have shown, it cannot represent the modulation of relative state frequencies implicit in the existence of a seasonal monsoon cycle. To accomplish this it becomes necessary to introduce some exogenous influence, that can pace, or regulate, the S2S multi-scale evolution of the monsoon. This modality was explored in Greene *et al.* (2011), where an index of monsoon strength, based on vertical shear of the zonal wind, is taken as the exogenous variable. Such a model is denoted a non-homogeneous HMM, or NHMM. The station network in Greene *et al.* (2011) is based on gridded data (Rajeevan *et al*., 2006) but spans a similar region to that discussed in Greene *et al*. (2008), while winds used to compute the index were extracted from a multi-model



Figs. 14(a&b). (a) NHMM state probabilities for the four states as a function of the monsoon index and (b) the seasonal climatology of the NHMM's daily sequence of states, averaged over all years. From Greene *et al.* (2011)

ensemble of GCM simulations, using the multimodel average. This provides a means of climate prediction by using the NHMM together with GCM predictions of the shear index.

The NHMM is fit jointly to the daily rainfall record and the monsoon index. Rather than a stationary transition matrix, the nonhomogeneous model infers a set of relations, between the index and each of the transition-matrix entries. The corresponding (fitted) equilibrium state probabilities are plotted against the index (again, for a four-state model) in Fig. 14(a), while Fig. 14(b) shows a climatology of the actual relative state occupation frequencies, as estimated from the observations. It can be seen from the latter (consistent with Fig. 13) that these frequencies vary considerably (but coherently) over the course of the monsoon season, with, e.g., the dry state (4) predominating in the early part of the season but becoming quite rare during peak monsoon, while the wet state (3) exhibits the inverse behavior, being almost nonexistent at the beginning and end of the season but predominating during the monsoon core. These temporal characteristics correspond well with the inferred smooth curves shown in Fig. 14(a), indicating a good fit of the model to data.

In addition to these useful inferences it was shown in Greene *et al*. (2011) that daily distributions of NHMM-simulated rainfall were largely free of the significant biases of GCM-generated daily rainfall, which tends (at least in lower-resolution GCMs) to be truncated toward low-intensity events and lacks the long tails characteristic of observed distributions. This illustrates the potential utility of the NHMM for downscaling fine-scale information inherent in observational daily rainfall data conditioned on larger-scale sources of climate predictability captured in GCM predictions, but which are often biased at smaller scales, especially in daily rainfall. In this framework scale- and time-specific information is utilized selectively, according to the source best able to provide it.

6. Summary and conclusions

This paper has reviewed research done by the authors and their collaborators at IRI and beyond over the past decade on predictability and prediction of Indian summer monsoon rainfall (ISMR) on seasonal and sub-seasonal timescales. Empirical analyses of the daily ISMR characteristics at local scales pertinent to agriculture, based on IMD gridded data, reveal that the number of rainy days in the season is more predictable and the seasonal rainfall total; furthermore, this "weather-within-climate" predictability undergoes an important modulation and is highest in the early and late phases of the monsoon and lowest in the July-August core monsoon period. New research in calibrated multimodel seasonal forecasting of ISMR is presented based on the North American Multimodel Ensemble and gridded IMD data, using the 2018 forecasts as a case study; these forecasts were issued in real-time in tercile-category probability format and were updated for the remainder of the 2018 monsoon season at the beginning of each calendar month from June to September. Sub-seasonal multimodel probabilistic predictions of ISMR in the weeks 2-3 range (8-21 day lead times) are constructed and analyzed, using the onset of the 2018 monsoon as an example; the hindcast skill of these week 2-3 gridded ISMR forecasts is shown to be substantial and widespread across most of India in the early and late stages of the monsoon season, consistent with the empirical findings from IMD data. Lastly, a hidden Markov model (HMM) of daily rainfall variability at a network of stations over monsoonal India is used to interpret the organized variation of rainfall across the multiple temporal scales that characterize ISMR.

The results presented here suggest several avenues for future study. The empirical finding that ISMR predictability is enhanced for rainfall frequency compared to total rainfall has not yet been

fully exploited on the seasonal scale and has yet to be considered at the sub-seasonal (or extended) range. In both cases more skillful forecasts and thus sharper probability forecasts with higher probability values stand to be more actionable in agricultural management, thorough this will need to be demonstrated. The rainfall states derived from the HMM offer a means toward downscaling of S2S forecasts, such as for use in crop simulation models. The higher predictability and broad sub-seasonal skill found in the early and late phases of the monsoon season should help to identify windows of opportunity with the potential of enhanced forecast utility, particularly when key ENSO and BSISO drivers are active. In such cases, the combination of seasonal forecasts (issued monthly) with sub-seasonal ones (issued weekly) holds promise for skillful monsoon forecasts that are seamless across time scales and of benefit to a broad spectrum of potential users.

Acknowledgments

The authors wish in particular to thank their many colleagues at the IRI and IMD, without whom the use-inspired research reviewed here would not have been possible. This work was supported by a Columbia University Climate and Life fellowship to AWR, by NOAA's International Research and Applications Program (NA). We acknowledge the agencies that support the NMME-Phase II system and we thank the climate modeling groups (Environment Canada, NASA, NCAR, NOAA/GFDL, NOAA/NCEP and University of Miami) for producing and making available their model output. NOAA/NCEP, NOAA/CTB and NOAA/CPO jointly provided coordinating support and led development of the NMME-Phase II system. The work in Section 3 is based on S2S data. S2S is a joint initiative of the World Weather Research Programme (WWRP) and the World Climate Research Programme (WCRP). The original S2S database is hosted at ECMWF as an extension of the TIGGE database.

The contents and views expressed in this research paper are the views of the authors and do not necessarily reflect the views of their organizations.

References

Acharya, N., Chattopadhayay, S., Mohanty, U. C. and Ghosh, K., 2014a, "Prediction of Indian summer monsoon rainfall: A weighted multi-model ensemble to enhance probabilistic forecast skills", *Meteorol. Appl*., **21**, 3, 724-732.

Acharya, N., Chattopadhyay, S., Mohanty, U. C., Dash, S. K. and Sahoo, L. N., 2013a, "On the bias correction of General Circulation Model output for Indian summer monsoon", *Meteorol. Appl*., **20**, 3, 349-356.

Acharya, N., Kar, S. C., Mohanty, U. C., Kulkarni, M. A. and Dash, S. K., 2011, "Performance of GCMs for seasonal prediction over India - A case study for 2009 monsoon", *Theor. Appl. Climatol*., **105,** 3-4, 505-520.

Acharya, N., Kulkarni, M., Mohanty, U. C. and Singh, A., 2014b, "Comparative evaluation of performances of two versions of NCEP Climate Forecast System in predicting Indian summer monsoon rainfall", *Acta Geophysica*, **62**, 199-219.

Acharya, N., Mohanty, U. C. and Sahoo, L. N., 2013b, "Probabilistic multi-model ensemble prediction of Indian summer monsoon rainfall using General Circulation Models: A non-parametric approach", *Comptes Rendus Geoscience*, **345**, 3, 724-732.

Barnston, A. G. and Tippett, M. K., 2014, "Climate information, outlooks and understanding where does the IRI stand?", *Earth Perspectives*, **1**, 1-20

Barnston, A. G. and Tippett, M. K., 2017, "Do statistical pattern corrections improve seasonal climate predictions in the North American Multimodel Ensemble models?", *Journal of Climate*, **30**, 20, 8335-8355.

Becker, E. and Van Den Dool, H., 2015, "Probabilistic seasonal forecasts in the North American Multimodel Ensemble: A baseline skill assessment", *Journal of Climate*, **29**, 8, 3015-3026.

Blanford, H. H., 1884, "On the connection of Himalayan snowfall and seasons of drought in India", *Proc. R. Soc. London, Ser. A*, **37**, 3-22.

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

Greene, A. M., Robertson, A. W., Smyth, P. and Triglia, S., 2011, "Downscaling projections of Indian monsoon rainfall using a nonhomogeneous hidden Markov model", *Quart. J. Royal Meteor. Soc*., **137**, 347-359.

Greene, A., Robertson, A. W. and Kirshner, S., 2008, "Analysis of Indian monsoon daily rainfall on subseasonal to multidecadal time scales using a hidden Markov model", *Quart. J. Royal Meteor. Soc*., **134**, 875-887.

Hamill, T. M., Whitaker, J. S. and Wei, X., 2004, "Ensemble reforecasting: Improving medium-range forecast skill using retrospective forecasts", *Monthly Weather Review*, **132**, 6, 1434-1447.

Hamill, T., 2012, "Verification of TIGGE Multi-model and ECMWF reforecast-calibrated probabilistic precipitation forecasts over the conterminous US", *Mon. Wea. Rev*., **140**, 2232-2252.

Hughes, J. P. and Guttorp, P., 1994, "A class of stochastic models for relating synoptic atmospheric patterns to regional hydrologic phenomena", *Water Resour. Res*., **30**, 5, 1535-1546.

Kar, S. C., Acharya, N., Mohanty, U. C. and Kulkarni, M. A., 2012, "Skill of monthly rainfall forecasts over India using multi-model ensemble schemes", *Int. J. Climatol*., **32,** 8, 1271-1286.

Kirtman, B. P., D. Min, J. M. Infanti, J. L. Kinter, D. A. Paolino, Q. Zhang, H. van den Dool, S. Saha, M. P. Mendez, E. Becker, P. Peng, P. Tripp, J. Huang, D. G. DeWitt, M. K. Tippett, A. G. Barnston, S. Li, A. Rosati, S.D. Schubert, M. Rienecker, M. Suarez, Z. E. Li, J. Marshak, Y. Lim, J. Tribbia, K. Pegion, W. J. Merryfield, B. Denis and E. F. Wood, 2014, "The North American Multimodel Ensemble: Phase-1 seasonal-to-interannual prediction; Phase-2 toward developing intraseasonal prediction", *Bull. Amer. Meteor. Soc.*, **95**, 585-601.

Kulkarni, M. A., Acharya, N., Kar, S. C., Mohanty, U. C., Tippett, M. K., Robertson, A. W., Luo J.-J. and Yamagata, T., 2012, "Probabilistic prediction of Indian summer monsoon rainfall using global climate models", *Theor. Appl. Climatol*., **107**, 441-450.

Kumar, K. K., Rajagopalan, B. and Cane, M. A., 1999, "On the weakening relationship between the Indian monsoon and ENSO", *Science*, **284**, 5423, 2156-2159.

Lee, J. Y., Wang, B., Wheeler, M. C., Fu, X., Waliser, D. E. and Kang, I.-S., 2013, "Real-time multivariate indices for the boreal summer intraseasonal oscillation over the Asian summer monsoon region", *Clim. Dyn*., **40**, 493-509.

Legates, D. R. and Willmott, C. J., 1990, "Mean seasonal and spatial variability in gauge-corrected global precipitation", *Int. J. Climatol*., **10**, 111-127.

Li, S. and Robertson, A. W., 2015, "Evaluation of sub-monthly precipitation forecast skill from global ensemble prediction systems", *Mon. Wea. Rev*., **143**, 2871-2889.

Moron, V. and Robertson, A. W., 2013, "Interannual variability of Indian summer monsoon rainfall onset date at local scale", *Int. Jour. Climatol*., **34**, 4, 1050-1061.

Moron, V., Robertson, A. W. and Pai, D. S., 2017, "On the spatial coherence of sub-seasonal to seasonal Indian rainfall anomalies", *Climate Dynamics*, **49**, 9, 3403-3423.

Moron, V., Robertson, A. W. and Ward, M. N., 2006, "Seasonal predictability and spatial coherence of rainfall characteristics in the tropical setting of Senegal", *Mon. Weather Rev*., **134**, 11, 2148-2148.

Moron, V., Robertson, A. W., Ward, M. N. and Camberlin, P., 2007, "Spatial coherence of tropical rainfall at the regional scale", *Journal of Climate*, **20**, 21, 5244-5263.

Nair, A., Acharya, N., Singh, A., Mohanty, U. C. and Panda, T., 2013, "On the predictability of northeast monsoon rainfall over south peninsular India in General Circulation Models", *Pure and Appl. Geophys.*, **170**, 11, 1945-1967.

Norris, J., 1997, "Markov Chains", Cambridge University Press, Cambridge, p237.

Pal, I., Robertson, A. W., Lall, U. and Cane, M., 2014, "Modeling winter rainfall in Northwest India using a Hidden Markov Model: Understanding occurrence of different states and their dynamical connections", *Clim. Dyn*., **44**, 3-4, 1003-1015.

Rajeevan, M., Bhate, J., Kale, J. and Lal, B., 2006, "High resolution daily gridded rainfall data for the Indian region: Analysis of break and active monsoon spells", *Curr. Sci. India*, **91**, 3, 296-306.

Robertson, A. W. and Vitart, F., 2019, "Sub-Seasonal to Seasonal Prediction: The Gap Between Weather and Climate Forecasting",. 1st Eds., ISBN 978-0-12-811714-9, Elsevier.

Robertson, A. W., Kirshner, S. and Smyth, P., 2004a, "Downscaling of daily rainfall occurrence over northeast Brazil using a hidden Markov model", *J. Climate*, **17**, 4407-4424.

Robertson, A. W., Lall, U., Zebiak, S. E. and Goddard, L., 2004b, "Improved combination of multiple atmospheric GCM ensembles for seasonal prediction", *Monthly Weather Review*, **132**, 12, 2732-2744.

Sahai, A., Chattopadhyay, R., Joseph, S., Krishna, P. M., Pattanaik, D. and Abhilash, S., 2019, "Seamless prediction of monsoon onset and active/break phases", *Sub-Seasonal to Seasonal Prediction: The Gap Between Weather and Climate Forecasting*, A. W. Robertson and F. Vitart, Eds., Elsevier, chap. 20, 421-438.

Singh, A., Acharya, N., Mohanty, U. C. and Mishra, G., 2013, "Performance of multi model canonical correlation analysis (MMCCA) for prediction of Indian summer monsoon rainfall using GCMs output", *Comptes Rendus Geoscience*, **345**, 2, 62-72.

Stephenson, D. B., Kumar, K. R., Doblas-Reyes, F. J., Royer, J.-F., Chauvin, F. and Pezzulli, S., 1999, "Extreme daily rainfall events and their impact on ensemble forecasts of the Indian monsoon", *Monthly Weather Review*, **127**, 9, 1954-1966.

Tippett, M. K., Barlow, M. and Lyon, B., 2003, "Statistical correction of central southwest Asia winter precipitation simulations", *Internat. J. Climatol*., **23**, 1421-1433.

Tippett, M. T., Barnston, A. G. and Robertson, A. W., 2007, "Estimation of seasonal precipitation tercile-based categorical probabilities from ensembles", *J. Climate*, **20**, 2210-2228.

Vigaud, N., Robertson, A. W. and Tippett, M., 2017a, "Multi-model ensembling of sub-seasonal precipitation forecasts over North America", *Mon. Wea. Rev*., **145**, 3913-3928.

Vigaud, N., Robertson, A. W., Tippett, M. and Acharya, N., 2017b, "Sub-seasonal predictability of boreal summer monsoon rainfall from ensemble forecasts", *Front. Env. Sci*., **5**, 67, 1-19.

Vigaud, N., Tippett, M. K. and Robertson, A. W., 2018, "Probabilistic skill of subseasonal precipitation forecasts for the East Africa-West Asia sector during September-May", *Weather and Forecasting*, **33**, 6, 1513-1532.

Vitart, F. and Molteni, F., 2009, "Dynamical extended-range prediction of early monsoon rainfall over India", *Monthly Weather Review*, **137**, 4, 1480-1492.

Vitart, F., 2017, "Madden-Julian Oscillation prediction and teleconnections in the S2S database", *Quart. J. Roy. Meteor. Soc*., **143**, 2210-2220.

Vitart, F., C. Ardilouze, A. Bonet, A. Brookshaw, M. Chen, C. Codorean, M. Déqué, L. Ferranti, E. Fucile, M. Fuentes, H. Hendon, J. Hodgson, H. Kang, A. Kumar, H. Lin, G. Liu, X. Liu, P. Malguzzi, I. Mallas, M. Manoussakis, D. Mastrangelo, C. MacLachlan, P. McLean, A. Minami, R. Mladek, T. Nakazawa, S. Najm, Y. Nie, M. Rixen, A. W. Robertson, P. Ruti, C. Sun, Y. Takaya, M. Tolstykh, F. Venuti, D. Waliser, S. Woolnough, T. Wu, D. Won, H. Xiao, R. Zaripov and L. Zhang, 2017, "The subseasonal to seasonal (s2s) prediction project database", *Bulletin of the American Meteorological Society*, **98**, 1, 163-173.

Walker, G. T., 1910, "On the meteorological evidence for supposed changes of climate on India", *Meteor. Mem*., **21**, 1-21.

Weigel, A. P., Liniger, M. A. and Appenzeller, C., 2007, "The discrete brier and ranked probability skill scores", *Monthly Weather Review*, **135**, 1, 118-124.

Wheeler, M. C. and Hendon, H. H., 2004, "An all-season real-time multivariate MJO index: Development of an index for monitoring and prediction", *Monthly Weather Review*, **132**, 8, 1917-1932.

Wilks, D. S., 2009, "Extending logistic regression to provide full-probability-distribution MOS forecasts", *Meteor. Appl*., **16**, 361-368.