Natural Hazards
https://doi.org/10.1007/s11069-018-3394-9

ORIGINAL PAPER

Sub-seasonal extreme rainfall prediction in the Kelani River basin of Sri Lanka by using self-organizing map classification

J. F. Vuillaume[1,2] · S. Dorji[1,3] · A. Komolafe[1,4] · S. Herath[1,5]

Received: 22 July 2017 / Accepted: 8 June 2018

Abstract

The availability of several multi-model and ensemble sub-seasonal forecasts online has generated a growing interest in extreme rainfall prediction and early warning. Developing countries located in the tropics like Sri Lanka are good examples of complex meteorological zones where early warning system progress is crucial for flood damage mitigation. This study investigates the potentials and advantage of the recently available Sub-seasonal to Seasonal (s2s) database provided by a consortium of weather forecasting institutes using self-organizing map classification. The results (1) highlight the relation between teleconnection indexes such as the Madden–Julian Oscillation and the spatiotemporal rainfall pattern, (2) illustrate that heavy rainfall event frequencies depend on the type of the cluster, (3) find that the performance of s2s forecasts varies among cluster and (4) provide corrective bias coefficient to forecast water volume in the basin for each cluster. This study highlights the interest of s2s forecast for extreme rainfall prediction and advocates for the release of real-time s2s data that can provide useful information for early warning in developing country such as Sri Lanka.

Keywords s2s · Self-organizing map (SOM) · Rainfall extremes · Bias correction · Early warning · Sri Lanka

Abbreviations

CMA	China Meteorological Agency
ECMWF	European Center for Middle-Range Weather Forecast
ENSO	El Nino Southern Oscillation
ERA-interim	European Reanalysis Interim

✉ J. F. Vuillaume
vuillaume@student.unu.edu

[1] United Nations University, Institute for the Advance of Sustainability, UNU-IAS, Tokyo, Japan

[2] Global Hydrology and Water Resources Engineering, 4-6-1 Komaba, Meguro-ku, Tokyo 153-8505, Japan

[3] National Center for Hydrology and Meteorology, Thimphu, Bhutan

[4] Remote Sensing and Geosciences Information System (GIS), Federal University of Technology, Akure, Ondo-State, Nigeria

[5] Ministry of Megapolis and Western Development, Government of Sri Lanka, Colombo, Sri Lanka

Published online: 21 June 2018

MJO	Madden–Julian Oscillation
NCEP	National Center for Environmental Protection
RMM	Real-time Multivariate MJO indices
S2S	Sub-seasonal to Seasonal
SOM	Self-organized map

1 Introduction

An efficient early warning system for heavy rainfall and improving flood prevention is the main aim of several weather and water agencies around the globe. Generally, the prediction quality decreases with the increase in lead time. However, an optimal early warning system should take advantage of information available at a critical timescale that allows proactive disaster mitigation intervention. A range of forecast lead-time data is already available from day to week and season, but sub-seasonal scale is poorly covered. Several sub-seasonal forecast systems from 1 to 4 weeks lead time were developed to cover a lead-time gap in current forecast systems with the intention to provide useful informations to several key sectors (Vitart et al. 2017). After implementation of sub-seasonal forecast systems by several forecast agencies as an operational experiment tool, data are usually made available for further research and improvement. However, the accuracy of this relatively "new" and "large" lead-time forecast product needs to be evaluated in diverse areas and weather conditions such as mid-latitude and the tropics. The skills of the s2s may potentially driven by lower assimilation quality, the overweight on mid-latitude forecast data in the global models and the presence of complex weather phenomena such as square line, heavy thunderstorm, and monsoon.

Lately, Vigaud et al. (2017) explored the s2s potential for the boreal summer season over North American, West African and Asian for the 1999–2010 period with the CMA, ECMWF and NCEP models. The authors computed the correlation between the MJO and s2s weeks 3–4 rainfall. The result findings indicated that Asia region presents the highest score with the ECMWF. Furthermore, Wang et al. (2017) examined the rainfall anomaly in California during February 2016, as well as in January and February 2017 and obtained results that "support the broader notion that what is unpredictable atmospheric noise at the seasonal timescale can become predictable signal at the sub-seasonal timescale". Finally, Olaniyan et al. (2018) considered the s2s ensemble mean rainfall forecast skills in a traditional "poor performance" region such as West Africa. The authors concluded that the s2s "has almost 75% synchronization with observation, implying that the model will thrice make correct forecast out of four forecasts on Nigeria and furthermore, that "quantitatively, the model predicts rainfall accumulation best over the Sahel and least over the Guinea areas".

There are potentially several applications that can benefit from s2s forecasts (White et al. 2017), such as the humanitarian sector (Goddard et al. 2014), public health for heat waves (Osman and Alvarez 2017), wind energy (Pinson 2013), water resources related to monsoon forecast (Jie et al. 2017). The flood prevention area is probably the most advanced application area. As an example, the Australian seasonal real-time operational is used as a support for extreme rainfall and flood forecasts as illustrated by White et al. (2015). Recently, Liu et al. (2017) assessed the performance of the sub-seasonal forecast using mainly the Madden Julian Oscillation (MJO) as a reference. First, the correlation analysis indicates that the forecast skill presents higher correlation with the Indian Ocean

Dipole (IOD) teleconnection index than the El Niño-Southern Oscillation (ENSO). Furthermore, MJO forecast skill is associated with the IOD, and the autumn period shows the highest skill. The authors further recommended the need to improve the physics of the region to solve the low predictability limit posed by the maritime continent.

Synoptic climatology is the study of the relationships between atmospheric circulations with the regional climate. Synoptic climate can be studied with canonical correlation analysis (CCA), principal component analysis (PCA), analogue methods like empirical orthogonal functions (EOF) and newer methods like self-organizing maps (Yarnal et al. 2001). The self-organizing map (SOM) is a relatively new field of synoptic climatology analysis with diverse applications (Sheridan and Lee 2011). The SOM (Kohonen 1982) is a nonlinear unsupervised neural network method that can identify patterns in gridded data and reorganize them in a map. SOM is effective at handling missing data, gives consistent results and can describe linear or nonlinear data distribution functions (Hewitson and Crane 2006). As examples, the applications of SOM in climate from multi-model studies can be dedicated to seasonal predictions (Guitiérrez et al. 2005), extreme climate events analysis (Cavazos 2000) or, Indian Summer Monsoon prediction (Borah et al. 2013). SOMs are objective circulation characterization patterns computed from realistic physics data using a wide range of method including the k-mean (Moron et al. 2016). It needs to span a range of sufficient and manageable circulation feature to capture the variability of circulations. Several authors have investigated the link between SOMs and extreme precipitation in regions like Japan (Ohba et al. 2015), Alaska (Glisan et al. 2016) or Australia (Li et al. 2016) with a converging view. SOM permits the link between derived circulation and spatial distribution of rainfall or temperature while providing useful analysis of the change in the circulation patterns over the last decades with the association to extremes.

Ohba et al. (2015) investigated the use of SOM for extreme rainfall over Japan with the Japanese 55-year Reanalysis Project (JRA-55). The SOMs were used to investigate the role of circulation and atmospheric moisture on extreme events. Their results indicated that ENSO affects the frequency of the clusters in relation to the heavy rainfall events. The study also suggested that the inter-decadal variability of frequency for heavy rainfall events corresponds to changes in frequency distributions of clusters and is not due to one particular cluster. Glisan et al. (2016) used the WRF model to investigate extreme daily precipitation using SOM in a polar region, Alaska. They showed that the SOM aids in determining high-frequency nodes, and hence, circulations are conducive to extremes. Their conclusion suggests that multiple circulation patterns are responsible for extreme days, but are differentiated by the location of the frequency of the extreme events in Alaska. These gaps are due to a combination of global weather system of mid-latitude driven by a large financial resource in the North and a lack of ground data in the South. This gap has a direct practical consequences on the quality of the forecast available in the region as well as the potential to correct bias error in the forecast based on accurate observation. The use of cluster to optimize rainfall forecast was illustrated for Japan by Vuillaume and Herath (2017) using forecast range from 1 to 10 days. The results suggested that bias correction at a station scale was improved with a clustering method such as the weather-type approach by a 10–20% rainfall RMSE reduction. Furthermore, a similar approach was used to optimize dynamical weather forecast based on weather-type approach with regional mean rainfall forecast improvement (Vuillaume and Herath, 2018). The study indicated an improvement of 10–20% of RMSE per weather type.

This study focused on the skilful predictability of extreme rainfall event in Sri Lanka at the sub-seasonal lead time (1–4 weeks). In addition, it used a self-organizing cluster weather classification to investigate the performance of the prediction with different weather situations. The majority of the tropical rainfall variability of the region is associated with the MJO phase anomaly which is complex to predict. The cluster method permits the classification of weather with a similar pattern that has specific forecasts skills and can allow the determination of more specific bias correction method than without classification.

2 Study area

Sri Lanka is located south-east of India in the Indian Ocean as shown in Fig. 1. The Kelani river basin is located south-west of Sri Lanka. Colombo is located in the downstream part of the Kelani river basin. The topography plays a significant role in the rainfall concentration time. The climate pattern of Sri Lanka is characterized by rainfall peaks around May and September and results in thunderstorms associated with convective cells and the location of the Inter-tropical Convergence Zone (ITCZ) over Sri Lanka (Zubair 2003). Further, both the monsoon and tropical cyclones from the Bay of Bengal contribute to heavy rainfall from October to November. On the west side, the hillslopes are the main driver of orographic rainfall that occurs from June to September when the westerly winds prevail (Zubair 2003). The short concentration time enables heavy rainfall to quickly create flood events. The Kelani river basin discharges to the sea 144 km downstream towards Colombo. The basin is subjected to heavy rains which results in the rapid rise of the Kelani stream flow and poses frequent flood hazard disasters for Colombo (Zubair 2003).

This study focused on the skilful predictability of extreme rainfall event in Sri Lanka at the sub-seasonal lead time (1–4 weeks). The majority of the tropical rainfall variability of the region is associated with the MJO phase anomaly which is complex to predict. The cluster method permits the classification of weather with a similar pattern that has specific

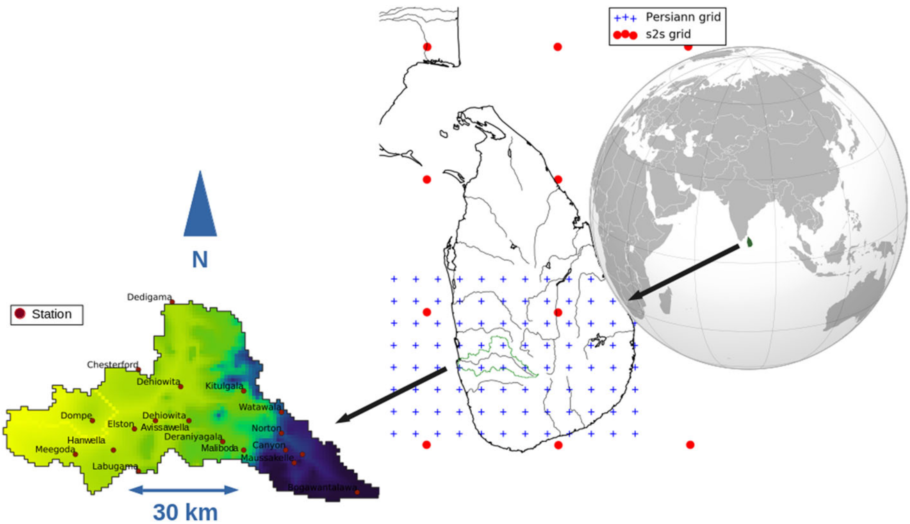

Fig. 1 Map of the Kelani river basin and the s2s/PERSIANN rainfall grid point over Sri Lanka. The stations used in the Kelani basin are indicated by red circle (left). The s2s grid forecast and the PERSIANN observation are marked by red and blue dots, respectively, over the map of Sri Lanka (middle). The Kelani basin is deliminated by green line over Sri Lanka

forecasts skills and can allow the determination of more specific bias correction method than without classification.

3 Methodology

The general framework of the study is illustrated in Fig. 2. First, the European Reanalysis (ERA-40 interim) was used to compute the k-mean cluster classification. Twenty rainfall station observations (1981–2009) from the Kelani river basin were used to identify rainfall events and the contribution of clusters in both water volume and extreme events. Both the observation and forecast data were post-processed to compute anomaly. Then, the rainfall anomalies of both observed PERSIANN and ECMWF s2s forecast for the weeks one, two and three were computed. The study identifies poor forecast clusters that need further attention and computed potential bias correction coefficients for watershed cumulative volume.

3.1 Data

Several data sets from different sources were used as summarized in Table 1. The ERA40 interim is the main source of gridded data used to compute the SOM and organized clusters. It consists of a large gridded data set with several fields optimized by data assimilation and several observation sources (Uppala et al. 2005). For this study, only the wind velocities U and V components at the level 200, 700 and 925 hPa pressure levels are used. The Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks—Climate Data Record (PERSIANN-CDR) provided by the Center for Hydrometeorology and Remote Sensing (CHRS) from the University of California, Irvine consist of daily rainfall data at 0.25° resolution from 1983 to present at the latitudes 60°S–60°N. The data addressed the need for a consistent, long-term, high-resolution global precipitation data set for the study, especially for extreme precipitation events. The PERSIANN algorithm used GridSat-B1 infrared data and adjusted with the Global Precipitation Climatology Project (GPCP) (Ashouri et al. 2015). The data present a good performance for the tropical regions in particular in reducing bias. They are used for

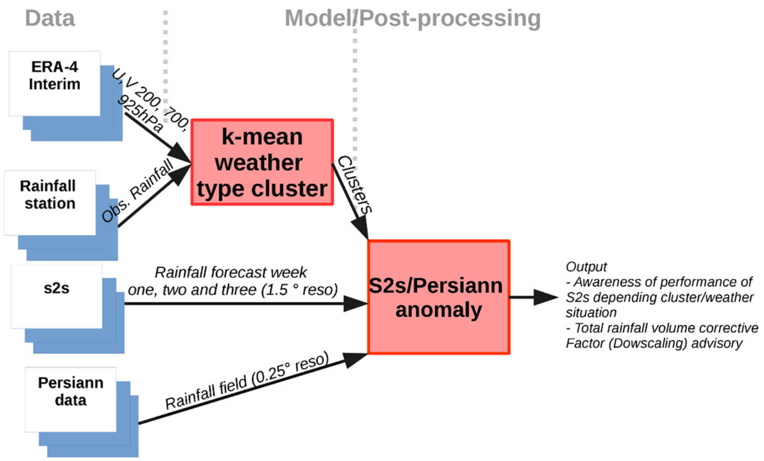

Fig. 2 Framework of the study

Table 1 Acronym, provider, resolution, period and duration of the data used in this study

Data set	Provider	Resolution (°)	Availability	Length (in years)
ERA4-interim	ECMWF	0.75	1979/01/01–2017/03/31	38
PERSIANN-CDR	CHRS California	0.25	1983/01/01–2016/08/31	35
S2S	CMA/ECMWF/NCEP	2.5	2015/01/01–2017/07/06	2.5
Stations	Sri Lanka government	Point	1981/01/01–2009/12/31	29

quality check of the s2s anomaly forecast. Further, additional rainfall observations over Sri Lanka at several rain gauge stations monitored by the Department of Meteorology Sri Lanka were used.

The s2s project consists of a database containing sub-seasonal to seasonal forecasts from 11 operational centres (Vitart et al. 2017). For convenience, the ECMWF data such as the large ensemble size (51 members), forecast period (weekly) and re-forecast frequency (two per week) and ensemble re-forecast size (11) are used. This study has to be considered as a prospect for further investigation when more data will become available.

3.2 Precipitation extremes

The extreme events analysis is strongly limited to the s2s date available. A minimum rainfall threshold of 40 mm/day is defined such that each cluster could at least have three representative case studies. Table 2 summarizes the cases used, the associated clusters, the date and the rainfall average among twenty locations in the basin (computed from

Table 2 Summary of event used from the s2s database, with case number, cluster #, date and average rain among station in the Kelani basin

Case #	Cluster #	Date	Ave. rain (mm)	Case #	Cluster #	Date	Ave. rain (mm)
1	1	2015-10-23	46.4	16	4	2015-04-07	44.8
2	1	2015-10-24	90.9	17	4	2015-04-19	61.7
3	1	2015-12-07	46.7	18	4	2015-04-20	65.9
4	1	2016-05-14	40.1	19	4	2015-05-08	77.7
5	2	2015-04-24	41.8	20	4	2015-05-09	130.8
6	2	2015-04-29	40.6	21	4	2015-11-06	40.1
7	2	2015-12-02	40.2	22	4	2015-11-29	45.0
8	2	2016-05-13	41.0	23	4	2015-12-08	54.0
9	2	2016-05-15	47.4	24	4	2016-04-26	41.4
10	3	2015-10-16	44.6	25	4	2016-04-30	41.0
11	2	2015-10-18	41.3	26	4	2016-05-08	58.0
12	2	2015-10-19	48.4	27	4	2016-05-09	40.1
13	2	2015-10-25	40.3	28	4	2016-05-10	61.5
14	2	2016-03-28	40.6	29	5	2015-06-04	41.5
15	2	2016-04-19	72.7	30	5	2015-09-03	42.8
				31	5	2016-05-27	45.3

PERSIANN grid). A stricter selection of extreme rainfall will be less representative of clusters and risk omitting extreme rainfall associated with some clusters. This will be confirmed by the study of historical data of the Kelani river basin from 1981 to 2009.

3.3 Teleconnection

Teleconnections in climate science are used to refer to wind patterns that connect different regions. Teleconnection refers to persistent, large-scale anomalies of pressure and weather circulations. The climate system exhibits several teleconnections such as the ENSO. It is also implied that some indexes can partly explain the variability of rainfall patterns. Several studies show the link between teleconnections such as the ENSO phase and rainfall pattern in Sri Lanka (Suppiah 1996; Zubair et al. 2008). However, the timescale of the ENSO is not convenient to assess the skill score of the sub-seasonal forecast. The Madden–Julian Oscillation (MJO) is used because it was identified as the principal mode of large-scale sub-seasonal variability in the tropical area (Madden and Julian 1971, 1994). The MJO is an equatorial eastward moving intra-seasonal disturbance of clouds, rainfall, winds, and pressure with a frequency of 30–60 days. It is associated with phases which are either enhancing or suppressing convective rainfall. The location of the group of clouds is the base of the phase classification of the MJO. The MJO can modulate the strength of the monsoons. The MJO is traditionally divided into eight phases which are expressed as a dipole-like pattern composites of summer monsoon rainfall anomalies observed in the Indian surrounding region, between Indian region and the Equatorial Indian Ocean (EIO). This dipole-like pattern consists of positive anomalies during the active convection (Real-time Multivariate Madden–Julian Oscillation RMM phases 3, 4, 5 and 6) and negative rainfall anomaly during suppressed convection phases (7, 8, 1 and 2) of the MJO (Mishra et al. 2017).

Teleconnections can be used for different purposes such as the verification of model quality in predicting large-scale changes. Therefore, MJO index can be used with rainfall pattern to confirm the presence of a pressure oscillation that has strong impacts on rainfall pattern. This pattern has one of the primary influences in the Indian Ocean. It needs to be predicted by sub-seasonal forecast system because s2s systems are developed and evaluated on their quality to forecast the MJO. Further, many recent studies investigated the link between MJO and rainfall pattern in several locations such as West Africa (Niang et al. 2017), South America (Giovannettone 2017; Shimizu et al. 2017) and the maritime continent in Indonesia (Mishra et al. 2017). In this study, we used available MJO data provided by the Bureau of Meteorology (Australia) and rainfall grid. The correlations index was computed online with the International Research Institute Columbia (IRI) tools.

3.4 Self-organizing map (SOM) for weather type

Self-organizing maps from the years 1990 to 2000 were used as a tool to assess climate and weather situation using systematic classification (k-mean method). The k-mean method is one of the methods available for SOM generation; it is an optimal method for low-dimensional unsupervised learning such as weather cluster. K-mean method uses the nearest mean velocity to create different clusters and classifies weather data. This study used the k-mean 2D classification over 6 parameters of the 3 levels of wind direction over u and v direction. Due to non-availability of recent data of the outgoing longwave radiation

(OLR), it was not used but could constitute an additional allowable variable to improve the classification.

3.5 s2s and post-processing

The treatment of s2s data in itself is a method because the multiple dimensions of the data make them complex to manipulate and extract optimal information. The study uses the method developed during the s2s workshop 2015 at the International Center for Theoretical Physics (Trieste, Italy). The method was modified to estimate the cumulative water volume anomaly at the basin scale for both observed (PERSIANN data) and forecast data. The method is summarized in the following steps:

1. Select extreme rainfall using rainfall threshold at the station location
2. Retrieve Observed PERSIANN gridded rainfall data for 20 years and retrieve real-time ensemble forecast and re-forecast (1986–2016) for the same date
3. Estimate the rainfall accumulation by computing the difference between time steps
4. Compute the mean of the ensemble over the region
5. Compare the ensemble mean with re-forecast mean for similar day of the year (number of years depend on of the provider method and the latest re-forecast data set availability, regular updates or "on the fly")
6. Compute the anomaly of the observations
7. Plot the anomaly and retrieve values.

The anomaly are computed against similar date but at different year times for 15–20 years depending on the model used. It allowed the estimation of anomaly for specific events at different forecast ranges typically 7–11 days, 12–18 days and 19–25 days.

4 Results

The description of the results starts with the teleconnection pattern between the MJO and the two rainy periods of Sri Lanka. Then, the results obtained by self-organizing map determination using the k-mean methods for 1979–2016 over Sri Lanka are summarized. Finally, the skills of the s2s forecast are assessed for the 31 largest rainfall events between January 2015 and December 2016.

4.1 MJO and rainfall correlation

The correlation map between the MJO phases and the rainfall anomaly pattern is computed with the online processing tools provided by the International Research Institute Columbia (IRI). Figures 3 and 4 illustrate the correlation between the anomaly rainfall and the MJO index, classified in phases for two periods (May–September and December–February). Both seasons highlight similar temporal variabilities of the seasonal rainfall distribution but with different spatial distributions for phases 1 and 8.

Overall, the phases 2–4 are characterized by positive rainfall anomaly, while the phases 5–7 by negative rainfall anomaly. The two periods correspond to the two monsoons periods observed in the region. For the monsoon period of May–September, the south-west area of Sri Lanka presents a strong positive rainfall anomaly for phases 2–4 and a negative anomaly for the phases 5–8. During the period of December–February, the north-east

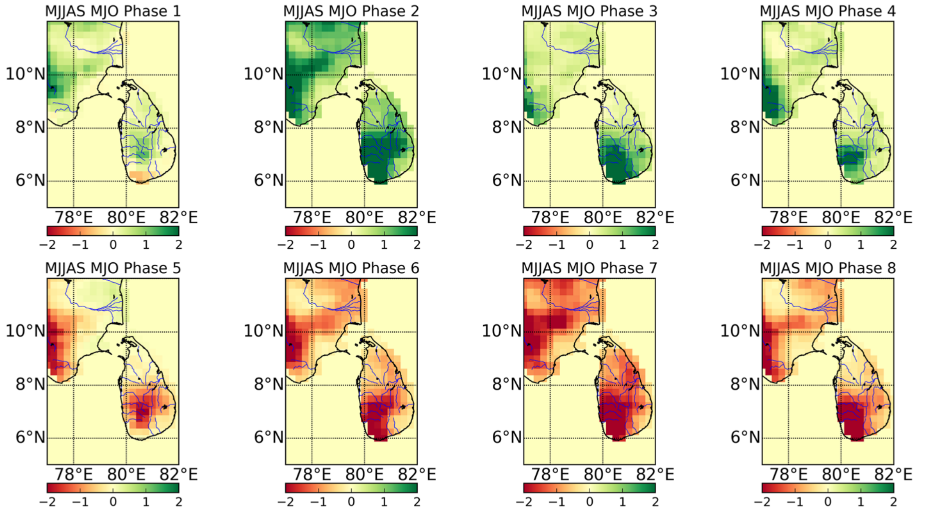

Fig. 3 Rainfall anomaly distribution (1981–2016 CHIRPS data; Funk et al. 2015) classified by Madden–Julian Oscillation phase index for the period of May–June–July–August–September over Sri Lanka. The red colour indicates a strong positive anomaly and the green a negative anomaly between the rainfall location and each the MJO phases

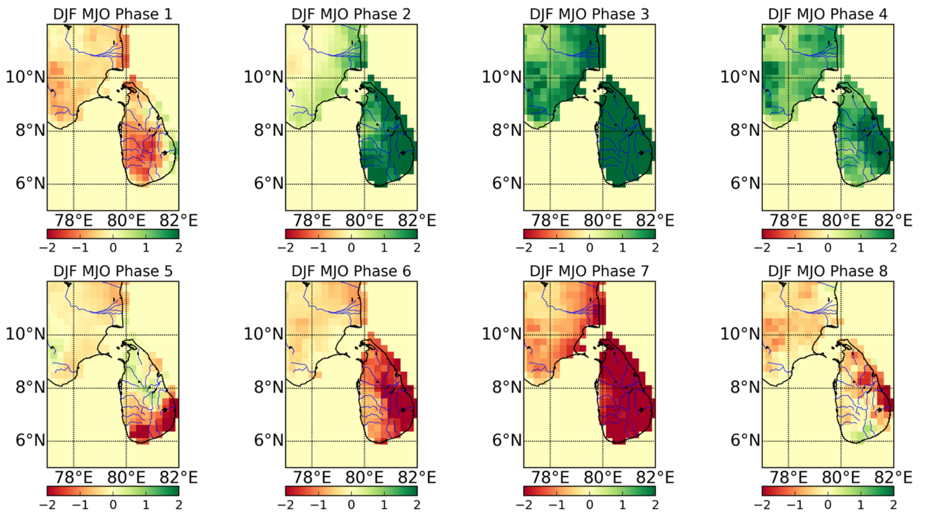

Fig. 4 Rainfall anomaly distribution (1981–2016 CHIRPS data; Funk et al. 2015) classified by Madden–Julian Oscillation phase index for the period of December–January–February over Sri Lanka. The red colour indicates a strong positive anomaly and the green a negative anomaly between the rainfall location and each the MJO phases

monsoon occurs and affects the whole country. The divergence of intensity or anomaly can also be observed between the west and east sides of the island. The phases 3, 4, 6, 7 and 8 exhibit large effects on the south-west coast of Sri Lanka while other part being relatively unaffected. Moreover, the phases 1, 5 and 8 of the December–February monsoon exhibit opposite behaviour between east–west or NW and SE which is strengthened by the orography of the island.

The results indicate that: (1) CHIRPS rainfall data observation adequately represents the monsoon pattern; (2) the distribution in time and space of the rainfall anomaly and therefore rainfall is strongly controlled by the MJO index. Therefore, it can be concluded that a sub-seasonal forecast time calibrated on MJO index quality should potentially improve the rainfall patterns in space and time in the tropical region.

4.2 Self-Organized Map for Sri Lanka

The SOM computed with the k-mean cluster approach at 925 hPa is plotted in Fig. 5 and for 200 hPa in Fig. 6. It illustrates the daily variability captured by the SOM. This method used six parameters (U and V at 200, 700 and 925 hPa, respectively) and classified weather events in a more refined feature than weather-type classification based on sea surface pressure. It suggests a link between clusters and the monsoon or the pre-monsoon period (wind shifted northward similarly to the ITCZ). Clusters can be associated with their wind direction and intensity pattern at 200 and 925 hPa. The solution using five SOMs was used because it gathers most of the variability observed in the region. The clusters can be divided into dominant wind direction, such as west, north-west, monsoon type and east. From the clusters classification, the characteristics of rainfall associated with clusters can be analysed such as cluster frequency, water volume, rainfall characteristics, extreme events frequency and extreme spatial distribution.

Figure 7 shows the seasonal variability of clusters frequency and the monthly average rainfall volume. A seasonal cluster variation can be observed both in terms of frequency of

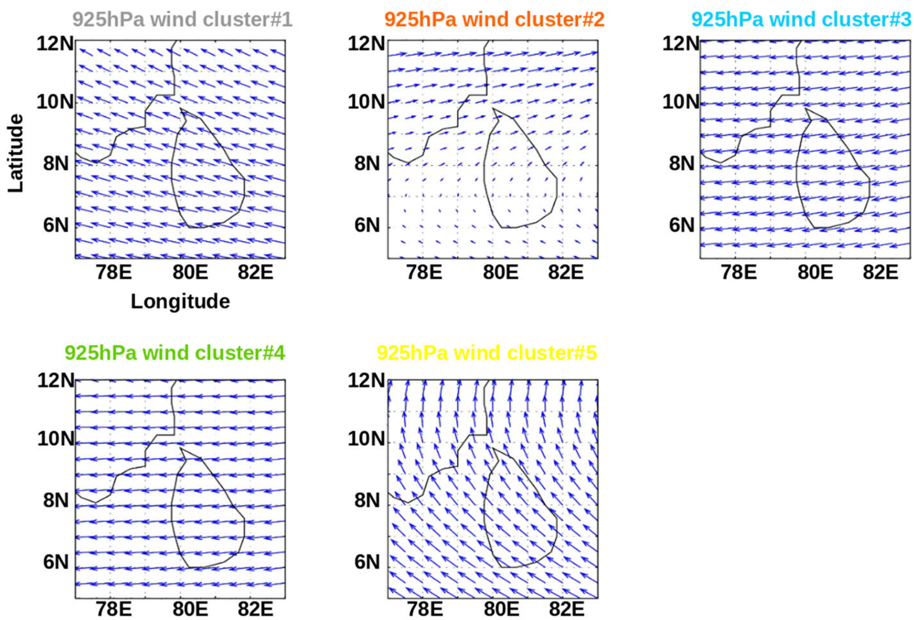

Fig. 5 Self-Organized map (SOM) wind pattern plotted at the 925hPa height for the five cluster solution. The blue blue arrow indicates the main wind direction at 925hPa

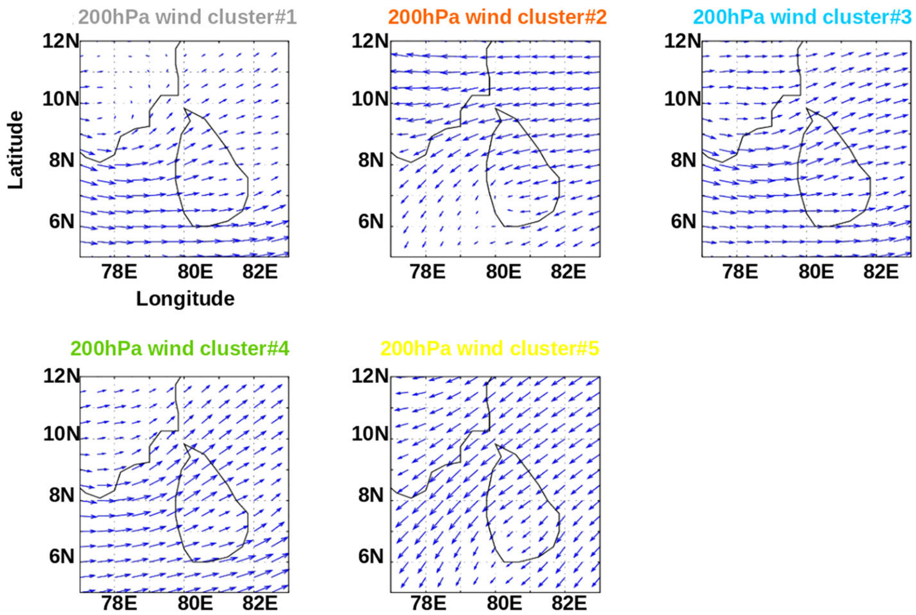

Fig. 6 Self-organized map (SOM) wind pattern plotted at the 925 hPa height for the five cluster solution. The blue arrow indicates the main wind direction 200 hPa

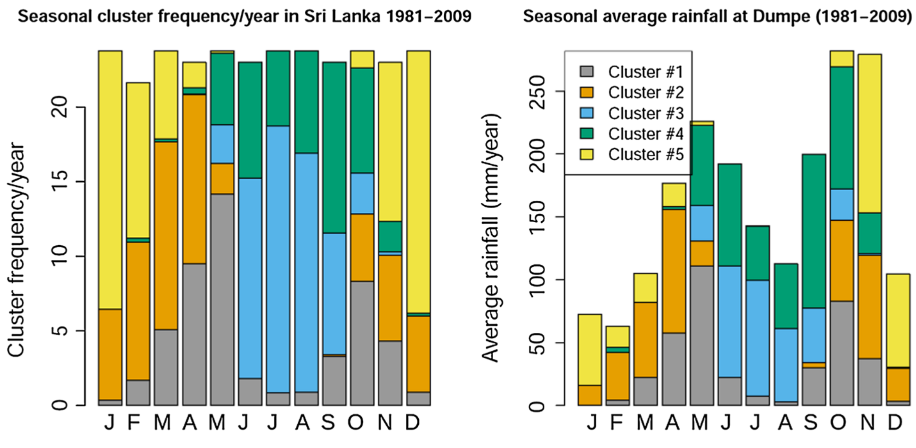

Fig. 7 Left: Seasonal cluster occurrence at Dompe in Sri Lanka. Right: Seasonal cumulative rainfall per cluster type in Sri Lanka from 1981 to 2009

clusters and in rainfall volume. Overall, the clusters #2 and #5 occurred during the NE monsoon with a large extension over the whole country but limited rainfall amount (an average of 100, 70 and 60 mm for the month of December, January and February, respectively). Similarly, the clusters #3 and #4 characterized the May–September/October period characterized by the SW monsoon which strongly affects Dompe location. Furthermore, rainfall can be analysed in terms of scale–shape parameters that characterized in terms of extreme/low rain dominance. Figure 8 summarizes the plot of the Kelani river

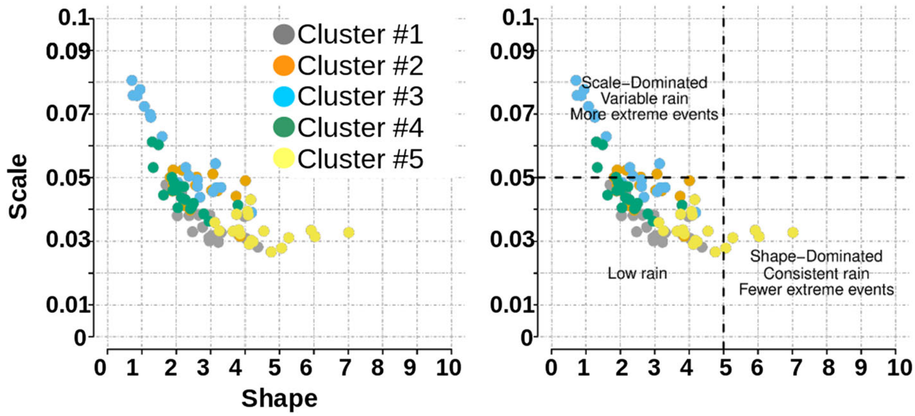

Fig. 8 Classification of rainfall per cluster using the 20 stations available over the period 1981–2009. Left: Shape versus scale plot of rainfall events. Right: Interpretation of the rainfall in terms of rainfall signature from more extreme to fewer extreme and low rainfall

basin stations rainfall in scale and shape classified by clusters (Husak et al. 2007). The figure illustrates the signature of clusters in terms of low/extreme rainfall. Moreover, the cluster #5 is identified as a low rain-few extreme cluster, while the cluster #3 low rain-more extreme. The cluster #1 is representative of low rain system, while clusters #2 and #3 a more complex association that tends to more extreme frequency. The analysis of extreme rainfall frequency in the Kelani river basin is illustrated in Fig. 9 with the threshold of 100, 200 and 300 mm cumulative over 1 day. The figure completes the previous Fig. 5 with a frequency analysis of the extreme vents. It confirmed that the clusters #2, #3 and #4

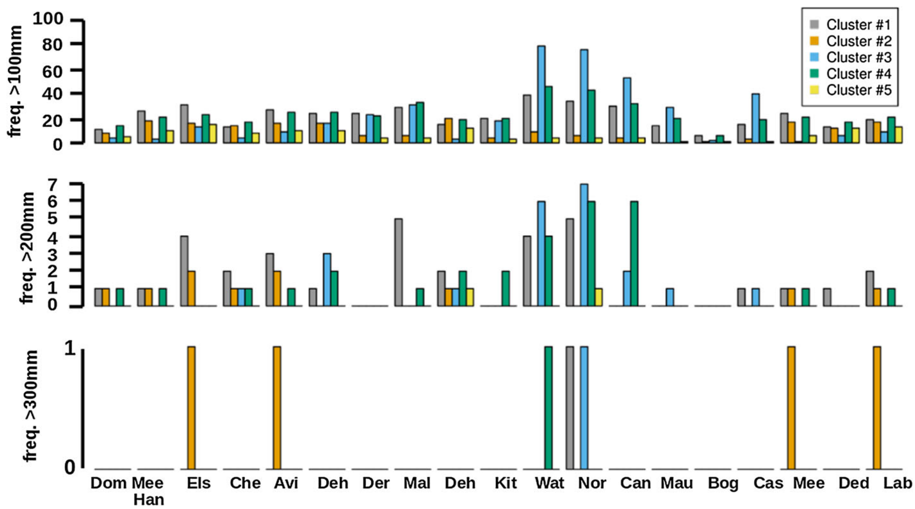

Fig. 9 Occurrence of extreme events per cluster at the meteorological stations located in the Kelani basin (see Fig. 1 for the exact location of the stations): Top > 100 mm/d, Middle > 200 mm/d, Bottom > 100 mm/d. Station names are as follows: *Dom.* Dompe, *Mee./Han.* Meepe/Hanwella, *Els.* Elston, *Che.* Chesterford, *Avi.* Avissawella, *Deh.* Dehiowita, *Der.* Deraniyagala, *Mal.* Maliboda, *Deh.* Dehiowita, *Kit.* Kitulgala, *Wat.* Watawala, *Nor.* Norton, *Can.* Canyon, *Mau.* Maussakelle, *Bog.* Bogawantalawa, *Cas.* Castlereigh, *Mee.* Meegoda, *Ded.* Dedigama, *Lab.* Labugama

presented several extremes with intensity larger than 100–150 mm. In addition, rare event of rainfall higher than 300 mm can be associated with clusters #1, #2, #3 and #4 with the highest frequency for cluster #2.

Finally, the spatial distribution of extreme rainfall higher than 200 mm is illustrated in Fig. 10 over the Kelani river basin with the basin elevation. First, the map shows the increase in occurrence in the eastward direction of the basin similarly to the increase in altitude. Secondly, the variability in the frequency of each cluster in the basin indicated that westward locations are affected by extreme associated with clusters #2, #1 and #4 and that a large majority of extreme are associated with clusters #3, #4 and #1 in the east part of the basin. As a result, the figure seems to indicate that large variation of extreme cluster occurs at the scale of a basin reflecting the complexity of extreme weather system in Sri Lanka but that can be captured by the SOM approach both temporally and spatially.

4.3 s2s skills in Sri Lanka

The computation of the s2s quality for 31 high rainfall events in the Kelani river basin was computed. The events were selected when one of the stations in the Kelani river basin reached a daily cumulative rainfall higher than 40 mm/day. Events were classified per weather types. The weekly s2s ECMWF forecast for weeks one, two and three lead time was extracted as shown as the example in Fig. 11.

Figure 12 presents the results of (a) the %hit ratio of each cluster for the weeks 1–3 and (b) the water volume corrective factor. The results indicate that clusters #2, #4 and #5 performed well in forecasting anomalies correctly for each of the lead times. However, the clusters #1 and #3 performed poorly for most of the events at all the week's lead time, except for the 1st and 2nd week of 7 December 2015, the 1st week of 14 May 2016, 16 October 2015 and 28 March 2016, respectively. These results strongly suggest that the s2s ECMWF forecast performance varies according to the cluster (Fig. 12a). Moreover, it illustrates the variability of cluster performance compared to the "all-cluster" agglomerated plot.

Generally, s2s forecast data underestimated rainfall observation similarly to most of the forecasts due to the coarse grid resolution (Fig. 12b, ratios are higher than 1). It is a

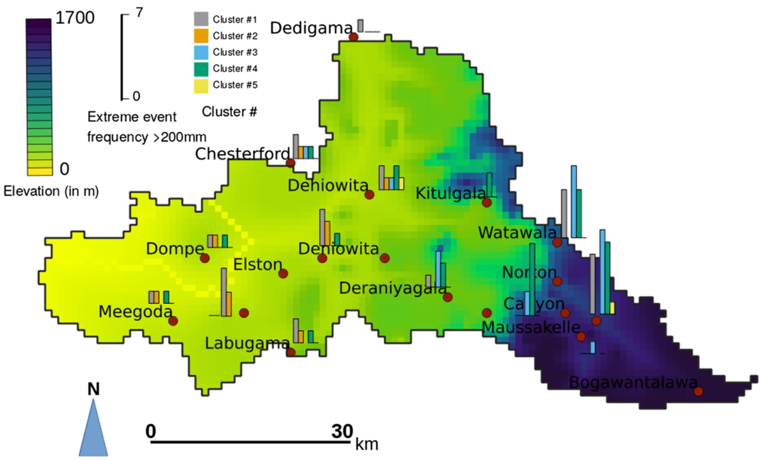

Fig. 10 Spatial distribution of extreme rainfall during 1981–2009 associated with their cluster type

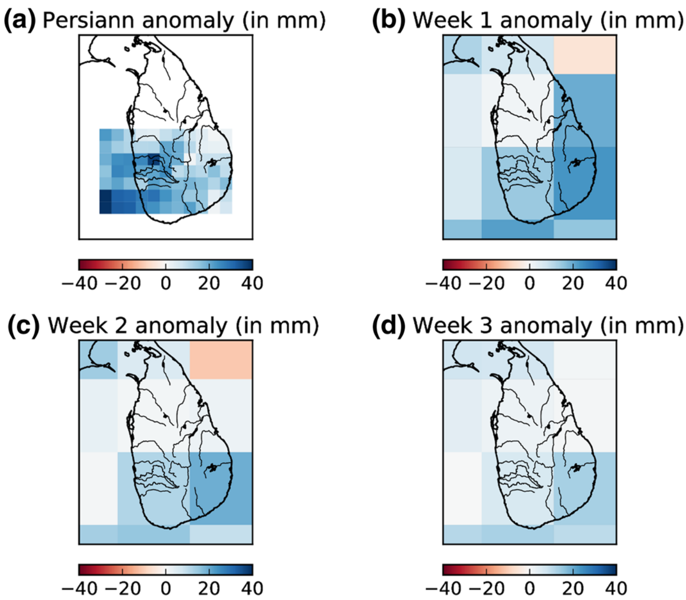

Fig. 11 Illustration of rainfall anomaly observation and s2s rainfall forecast anomaly (in mm) **a** PERSIANN rainfall anomaly spatial distribution of the 23-10-2015 with 0.25° resolution. **b–d** ECMWF rainfall ensemble mean anomaly for the week one (**b**), two (**c**) and three (**d**)

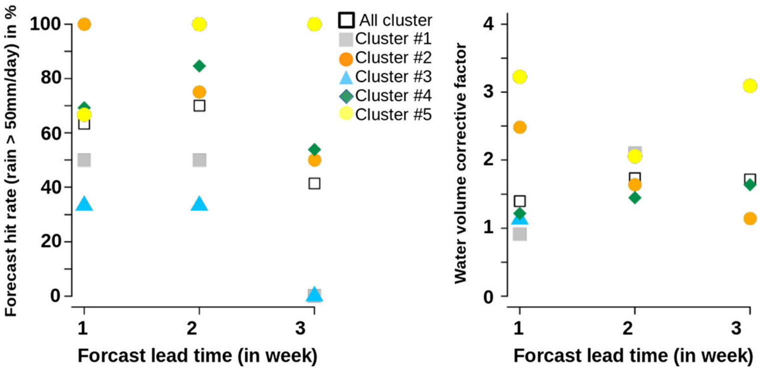

Fig. 12 Left: Summary plot of the forecast hit rate for rainfall event higher than 50 mm/day in % for each weekly lead time. Right: Summary plot of the potential water volume corrective factor or bias correction that need to be applied at each weekly lead time per clusters

characteristic of the low ability of the seasonal forecast to model convective rainfall and coarse orographic resolution even when a sub-grid parameterization is used. Moreover, the clusters ratio of PERSIANN rainfall observation versus s2s forecast does not systematically follow a linear regression curve such as suggested by the "all-cluster" plot. Then, the cluster #2 indicates a decrease in bias correction at longer lead times, while clusters #1 and #4 an increase. The cluster #3 presented no skills at weeks 2 and 3, while cluster #5 a decrease from these bias error coefficients are presented as an illustration of the current situation and will strongly depend on the increase in s2s data records. Nevertheless, statistical corrective bias coefficient is crucial for the design of an adequate early warning system since statistical downscaling is required. For clusters that systematically forecast negative anomaly instead of positive one, the corrections were not possible, and therefore, careful survey should be conducted upon detection of those clusters.

5 Discussion

Several limitations can be drawn from this study regarding the s2s forecast data availability. It impacts the estimation of the skills score per weather type and the total watershed volume bias corrective coefficient. Thus, formulating the rainfall threshold of 40 mm/day to define extremes is a concern. Furthermore, the interpretation of the clusters is complex because several clusters coexist at certain seasons and only their frequency permits a partial analysis. There is also a need to consider that trend of change that occurs in the region during the analysis period of cluster classification (1981–2009) and the s2s forecast (2015–2016).

5.1 Data limitation and clustering method

The results obtained for the s2s forecast are only limited to 2 years due to the recent release of s2s data. Therefore, this study presents a first step of the potential of s2s for heavy rainfall alerts. It effectively identifies the main opportunities for improvement by investigating on the clusters #1 and #3 in the tropical region. Further, s2s sub-seasonal forecast providers other than the ECMWF could be investigated. However, the ECMWF was frequently run (1 per week) and the large size of the ensemble available (51 members) can model the uncertainty. The number of clusters was kept low to limit the number of cases to analysis and isolate the cluster with poor prediction skill. While both monsoons were studied, circulation patterns like the tropical cyclone were not identified and merged with clusters that capture the general circulation pattern associated with cyclonic events such as westerly wind.

The OLR data are reported as an important parameter for k-mean circulation cluster classification (Moron et al. 2016). However, recent data and real-time forecast were not available for the same period as U and V parameters. For a real-time operational perspective, the sub-seasonal OLR forecast is not provided. The forecast of clusters will be limited and the choice of the correction coefficient for extreme bias correction uncertain. Therefore, OLR data were avoided for this study.

5.2 SOM decadal variability

This study based its results on the period 1981–2009. However, variation in clusters distribution may occur during this period. A change in clusters frequency will add further constraints for s2s forecast system efficiency and self-organizing map used. Sheikh et al. (2015) indicated that extreme rainfall intensity decadal change can be observed in Sri Lanka. Furthermore, the consecutive dry day's index has generally decreased across the region except in Sri Lanka. Therefore, both extreme rainfall and drought index indicated trends that will most likely be reflected in cluster distribution trend because of the correlation between cluster and rainfall signatures (Fig. 8).

Figure 13 illustrates the decadal variability of the occurrence of the cluster associated with the extremes. The decadal variability of the clusters is computed at the Dompe station

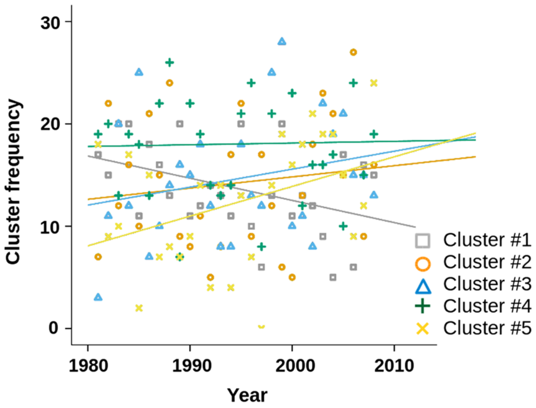

Fig. 13 Scatter plot and trend of decadal change of weather-type frequency in the region of Colombo from 1981 to 2009. Each symbol and colour indicates a cluster category

located in the west of the Kelani basin (see Fig. 1). It illustrated the increase in clusters #2, #3 and #5, the stability of cluster #4 and the decrease in cluster #1. Therefore, it also illustrated a decadal trend change in the Kelani river basin associated with an increase in pro-extreme rainfall clusters and the decrease in low rain ones. Furthermore, the seasonal change of clusters was investigated. Figure 14 illustrates the seasonal change over the last three decades. An increase in clusters #3 associated with extreme rainfall can be noticed in the month of May and October during pre- and post-monsoon period. Furthermore, an increase in cluster #1 was associated with fewer extreme rainfall from October–February during the DJF monsoon. Therefore, the seasonal change suggests an increase in extreme during pre–post-monsoon period and an increase in low rain/few extreme during DJF monsoon.

5.3 Data limitation and clustering method

The results obtained for the s2s forecast are limited to 2 years due to the recent release of s2s data. Therefore, this study presents a first step of the potential of s2s for heavy rainfall alerts. It effectively identifies the main opportunities for improvement by investigating on the clusters #1 and #3 in the tropical region. The further s2s sub-seasonal forecast could be investigated such as BoM, CMA, HMCR, ISAC-CNR, JMA, Méteo France, NCEP, UKMO, ECCC and KMA. However, the ECMWF presents a high frequently run (1 per week) and a large ensemble member size (51 members). The number of clusters was kept low to limit the number of cases to analyse and isolate the cluster with poor prediction

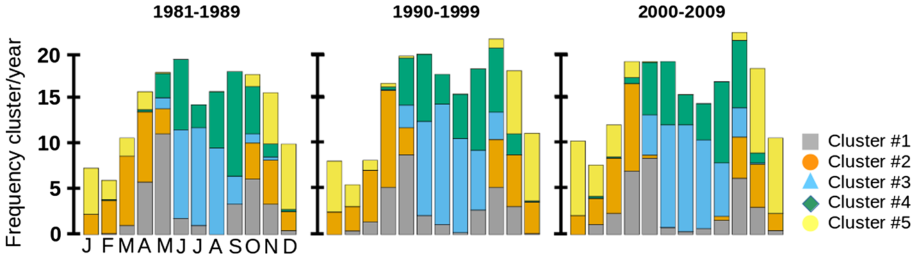

Fig. 14 Seasonal frequency of clusters (in number of cluster per year) for the periods 1981–1989, 1990–1999 and 2000–2009, respectively

skill. While both monsoons were studied, circulation patterns like the tropical cyclone were not identified and merged with clusters that captured the general circulation pattern associated with cyclonic events such as westerly wind.

The k-mean study uses six parameters, but others may have been used such as the outgoing long radiation (OLR). The OLR data are reported as an important parameter for the k-mean circulation cluster classification (Moron et al. 2016). However, recent data and real-time forecast were not available for the same period as U and V parameters. Moreover, regarding operational perspectives, the sub-seasonal OLR forecast is not provided making it use not possible for cluster forecast. Therefore, OLR data were avoided for this study.

6 Conclusions

The study illustrates the potential of the s2s forecast for real-time heavy rainfall early warning system in the Kelani basin in Sri Lanka. The study takes advantage of the self-organized map for climate and adapts it for sub-seasonal weather prediction. It evaluated the frequency of extreme rainfall events associated with specific weather circulation. The weather circulations were estimated using k-mean cluster method with six climate parameters: U and V at 200, 700 and 925 hPa. Then, the quality of the recently released s2s database was evaluated for their potential to predict heavy rainfall event. Then, it was also used to estimate an anomaly bias correction coefficient per cluster.

The results obtained are summarized in Table 3. It shows that (1) the frequency of heavy rainfall is higher for clusters #1, #3 and #4 than #2 and #5 (Fig. 9). (2) The extreme rainfall–cluster associations are spatially localized (clusters #3 and #4 in the east with higher elevation, cluster #2 at low altitude and therefore in the west). The cluster #1 is distributed over the whole basin with peak activity in pre- and post-monsoon period. (3) Performance of the extreme rainfall differs largely among clusters. (4) Different bias correction factors are necessary for each cluster.

The performance of the seasonal forecast is strongly dependent on the type of cluster or weather condition. The extreme associated with the clusters #3 and #1 present low performance, while the clusters #2, #4 and #5 exhibited a good prediction. The s2s data present already some early alert skill that could be used for early warning flood/heavy rainfall/landslide alerts if the correction factor is applied. Furthermore, full awareness should be raised about clusters that are poorly forecast (clusters #1 and #3) while presenting high inundation risk. In addition, the frequency cluster occurred during the last

Table 3 Summary of the results of the s2s forecast performance study

Cluster #	General circulation	Surf. wind direct.	Period	Fct. Perf.	W1-3 Corr. coeff.
#1	SE monsoon	South-easterly	February–May and September–November	Poor	0.9–ND*–ND*
#2	Pre-/post-monsoon	Weak	October–February and September–November	Very good	2.5–1.6–1.2
#3	Monsoon	Easterly	May–October	Very poor	1.1–ND*–ND*
#4	Monsoon	Easterly	May–October	Good	1.2–1.5–1.6
#5	Winter period	North-easterly	November–January	Good	3.2–2–3

ND indicates that the anomaly was not detected by the sub-seasonal forecast. *ND** not determined three decades with seasonal change in the pre- and post-monsoon extreme cluster frequencies. These changes will most likely affect the future performance of the s2s subseasonal forecast. Finally, this study advocates for the full real-time release of s2s data to support meteorological agencies of developing countries with resource constraints.

Acknowledgements The authors thank all the oral contribution speakers and poster for the ICTP s2s workshop held in Trieste in November 2016. Special acknowledgements for code sharing are dedicated to Dr. P. Gonzalez and Prof. V. Moron.

Compliance with ethical standards

Conflict of interest The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

Ashouri H, Hsu K-L, Sorooshian S, Braithwaite D, Knapp KR, Cecil L, Nelson BR, Prat O (2015) Persiann-cdr: daily precipitation climate data record from multisatellite observations for hydrological and climate studies. Bull Am Meteorol Soc 96:69–83. https://doi.org/10.1175/BAMS-D-13-00068.1
Borah N, Sahai AK, Chattopadhyay R, Joseph S, Abhilash S, Goswami BN (2013) A self-organizing map-based ensemble forecast system for extended range prediction of active/break cycles of Indian summer monsoon. J Geophys Res Atmos 118:9022–9034. https://doi.org/10.1002/jgrd.50688
Cavazos T (2000) Using self-organizing maps to investigate extreme climate events: an application to wintertime precipitation in the Balkans. J Clim 13:1718–1732
Funk C, Peterson P, Landsfeld M, Pedreros D, Verdin J, Shukla S, Husak G, Rowland J, Harrison L, Hoell A, Michaelsen J (2015) The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. Sci Data 2:150066. https://doi.org/10.1038/sdata.2015.66
Giovannettone JP (2017) Correlating MJO activity with Argentina rainfall and Atlantic Hurricanes using ICI-RAFT. J Hydrol Eng. https://doi.org/10.1061/(ASCE)HE.1943-5584.0001249
Glisan JM, Gutowski WJ Jr, Cassano JJ, Cassano EN, Seefeldt MW (2016) Analysis of WRF extreme daily precipitation over Alaska using self-organizing maps. J Geophys Res Atmos 121:7746–7761. https://doi.org/10.1002/2016JD024822
Goddard L, Baethgen WE, Bhojwani H, Robertson AW (2014) The International Research Institute for Climate and Society: why, what and how. Earth Perspect 1:1–14. https://doi.org/10.1186/2194-6434-1-10
Guitiérrez J, Cano R, Cofiño AS, Sordo C (2005) Analysis and downscaling multi-model seasonal forecasts in Peru using self-organizing maps. Tellus A 57:435–447. https://doi.org/10.1111/j.1600-0870.2005.00128.x
Hewitson BC, Crane RG (2006) Consensus between GCM climate change projections with empirical downscaling: precipitation downscaling over South Africa. Int J Climatol 26:1315–1337. https://doi.org/10.1002/joc.1314
Husak GJ, Michaelsen J, Funk C (2007) Use of the gamma distribution to represent monthly rainfall in Africa for drought monitoring applications. Int J Climatol 27:935–944. https://doi.org/10.1002/joc.1441
Jie W, Vitart F, Wu T, Liu X (2017) Simulations of the Asian summer monsoon in the sub-seasonal to seasonal prediction project (S2S) database. Q J R Meteorol Soc 143:2282–2295. https://doi.org/10.1002/qj.3085
Kohonen T (1982) Self-organized formation of topologically correct feature maps. Biol Cybern 43:59–69. https://doi.org/10.1007/BF00337288
Li J, Sharma A, Evans J, Johnson F (2016) Addressing the mischaracterization of extreme rainfall in regional climate model simulations—a synoptic pattern based bias correction approach. J Hydrol 556:901–912. https://doi.org/10.1016/j.jhydrol.2016.04.070
Liu X, Wu T, Yang S, Li T, Jie W, Zhang L, Wang Z, Liang X, Li Q, Cheng Y, Ren H, Fang Y, Nie S (2017) MJO prediction using the sub-seasonal to seasonal forecast model of Beijing Climate Center. Clim Dyn 48:3283–3307. https://doi.org/10.1007/s00382-016-3264-7
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.https://doi.org/10.1175/1520-0469(1971)028%3C0702:DOADOI%3E2.0.CO;2

Madden RA, Julian PR (1994) Observations of the 40–50-day tropical oscillation—a review. Mon Weather Rev 122:814–837. https://doi.org/10.1175/1520-0493(1994)122%3C0814:OOTDTO%3E2.0.CO;2
Mishra SK, Sahany S, Salunke P (2017) Linkages between MJO and summer monsoon rainfall over India and surrounding region. Meteorol Atmos Phys 129:283. https://doi.org/10.1007/s00703-016-0470-0
Moron V, Gouirand I, Taylor M (2016) Weather types across the Caribbean basin and their relationship with rainfall and sea surface temperature. Clim Dyn 47:601–621. https://doi.org/10.1007/s00382-015-2858-9
Niang C, Mohino E, Gaye AT, Omotosho JB (2017) Impact of the Madden Julian Oscillation on the summer West African monsoon in AMIP simulations. Clim Dyn 48:2297–2314. https://doi.org/10.1007/s00382-016-3206-4
Ohba M, Kadokura S, Yoshida Y, Nohara D, Toyoda Y (2015) Anomalous weather patterns in relation to heavy precipitation events in Japan during the Baiu Season. J Hydrometeorol 16:688–701. https://doi.org/10.1175/JHM-D-14-0124.1
Olaniyan E, Adefisan EA, Oni F, Afiesimama E, Balogun AA, Lawal KA (2018) Evaluation of the ECMWF sub-seasonal to seasonal precipitation forecasts during the Peak of West Africa Monsoon in Nigeria. Front Environ Sci 6:4. https://doi.org/10.3389/fenvs.2018.00004
Osman M, Alvarez MS (2017) Subseasonal prediction of the heat wave of December 2013 in Southern South America by the POAMA and BCC-CPS models. Clim Dyn 6:1–15. https://doi.org/10.1007/s00382-016-3474-z
Pinson P (2013) Wind energy: forecasting challenges for its operational management. Stat Sci 28(4):564–585. https://doi.org/10.1214/13-STS445
Sheikh MM, Manzoor N, Ashraf J, Adnan M, Collins D, Hameed S et al (2015) Trends in extreme daily rainfall and temperature indices over South Asia. Int J Climatol 35:1625–1637. https://doi.org/10.1002/joc.4081
Sheridan SC, Lee CC (2011) The self-organizing map in synoptic climatological research. Prog Phys Geogr 35:109–119. https://doi.org/10.1177/0309133310397582
Shimizu MH, Ambrizzi T, Liebmann B (2017) Extreme precipitation events and their relationship with ENSO and MJO phases over northern South America. Int J Climatol 37:2977–2989. https://doi.org/10.1002/joc.4893
Suppiah R (1996) Spatial and temporal variations in the relationships between the Southern Oscillation phenomenon and the rainfall of Sri Lanka. Int J Climatol 16:1391–1407. https://doi.org/10.1002/(SICI)1097-0088(199612)16:12h1391:AID-JOC94i3.0.CO;2-X
Uppala SM, Kallberg PW, Simmons AJ, Andrae U, Bechtold VDC, Fiorino M et al (2005) The era-40 re-analysis. Q J R Meteorol Soc 131:2961–3012. https://doi.org/10.1256/qj.04.176
Vigaud N, Robertson AW, Tippett MK (2017) Multi-model ensembling of subseasonal precipitation forecasts over North America. Mon Weather Rev 145:3913–3928
Vitart F, Ardilouze C, Bonet A, Brookshaw A, Chen M, Codorean C et al (2017) The subseasonal to seasonal (s2s) prediction project database. Bull Am Meteorol Soc 98:163–173. https://doi.org/10.1175/BAMS-D-16-0017.1
Vuillaume JF, Herath S (2017) Improving global rainfall forecasting with a weather type approach in Japan. Hydrol Sci J 62:167–181. https://doi.org/10.1080/02626667.2016.1183165
Vuillaume JF, Herath S (2018) Dynamic downscaling based on weather types classification: an application to extreme rainfall in south-east Japan. J Flood Risk Manag. https://doi.org/10.1111/jfr3.12340
Wang S, Anichowski A, Tippett MK, Sobel AH (2017) Seasonal noise versus subseasonal signal: forecasts of California precipitation during the unusual winters of 2015–2016 and 2016–2017. Geophys Res Lett 44:9513–9520. https://doi.org/10.1002/2017GL075052
White CJ, Franks SW, McEvoy D (2015) Using subseasonal-to-seasonal (s2s) extreme rainfall forecasts for extended-range flood prediction in Australia. Proc Int As Hydrol Sci 370:229–234. https://doi.org/10.5194/piahs-370-229-2015
White CJ, Carlsen H, Robertson AW, Klein RJ, Lazo JK, Vitart F, Coughlan De Perez E, Ray AJ, Murray V, Bharwani S, Macleod D, James R, Fleming L, Morse AP, Eggen B, Graham R, Kjellström E, Becker E, Pegion KV, Holbrook NJ, Mcevoy D, Depledge M, Perkins-kirkpatrick S, Brown TJ, Street R, Jones L, Remenyi TA, Hodgson-johnston I, Buontempo C, Lamb R, Meinke H, Arheimer B, Zebiak SE (2017) Potential applications of subseasonal-to-seasonal (S2S) predictions. Meteorol Appl 24(3):315–325. https://doi.org/10.1002/met.1654
Yarnal B, Comrie AC, Frakes B, Brown DP (2001) Developments and prospects in synoptic climatology. Int J Climatol 21:1923–1950. https://doi.org/10.1002/joc.675

Zubair L (2003) Sensitivity of Kelani streamflow in Sri Lanka to ENSO. Hydrol Process 17:2439–2448. https://doi.org/10.1002/hyp.1252

Zubair L, Siriwardhana M, Chandimala J, Yahiya Z (2008) Predictability of Sri Lankan rainfall based on ENSO. Int J Climatol 28:91–101. https://doi.org/10.1002/joc.1514