

## Reviewer 1 (RC1): Adway Mitra

MS No.: egusphere-2025-4585

---

The paper focuses on identifying spatial patterns of Indian Monsoon rainfall using K-means clustering. These patterns are identified with specific phases of Indian Monsoon, such as the break phase with low overall rainfall, and several other phases, each of which is characterized by rainfall over some particular region like the Indo-Gangetic Plane, Northeastern India or Kutch and Thar. It is argued that these 10-odd clusters represent the spatial distribution of rainfall on most of the days in the study period, it is assumed that these patterns follow Markovian dynamics, and transition probabilities are worked out with some physical justifications. The change in the frequency of each of the clusters/patterns and the intensity of rainfall associated with them is studied across the entire study period, and such change is considered as a marker of climate change. I have the following observations about this paper:

We thank Dr. Mitra for his insightful comments. Author reply to each question is given below.

---

### Comment 1: Originality

This is not the first work that aims to identify a set of "canonical" patterns of rainfall. MISO or Monsoon Intra-Seasonal Oscillations (Suhas et al, 2013, "An Indian monsoon intra-seasonal oscillations (MISO) index for real time monitoring and forecast verification") and the binary Random Fields (Mitra et al, 2019, "Spatio-temporal patterns of daily Indian summer monsoon rainfall", Sharma et al, 2020, "Spatio-temporal relationships between rainfall and convective clouds during Indian monsoon through a discrete lens"). The patterns identified here have significant similarity with these patterns, and hence some sort of qualitative and quantitative comparison is needed, along with a justification of why at all more patterns like this are needed.

We thank the reviewer for bringing these studies to our attention. We think independent methodologies converging on similar patterns actually strengthens the case that these modes of monsoon rainfall variability are physically robust rather than artifacts of any particular technique.

We agree that rainfall-regime identification is not new in itself, and we have revised the Introduction to position our study more clearly relative to previous work. The novelty of this paper does not lie only in identifying spatial rainfall patterns, but in the post-clustering diagnostic framework built on those patterns. Specifically, this study (i) derives regimes directly from 58 years of daily 0.25° IMD rainfall without dimensionality reduction, (ii) associates each regime with composite synoptic anomalies of wind, sea-level pressure, and moisture, (iii) quantifies regime persistence and sequential evolution using transition probabilities, and (iv) decomposes regional rainfall changes into cluster frequency and intensity contributions. While some regimes are qualitatively similar to those identified in earlier studies, that convergence supports their physical robustness. We have now added discussion of Suhas

et al. (2013) and Mitra et al. (2019) in the Introduction and clarified how the present framework differs in aim and analysis.

---

## Comment 2: Choice of K-means

It is not very clear what was considered as the feature vectors for clustering. Is it vectorized form of the rainfall values of each day? If so, we may be missing out on the spatial aspects due to vectorization (2D to 1D). Also, what is the nature of the space on which we are doing the clustering? What kinds of clusters are desired, and is K-means the ideal way to do it? What is the within-cluster variance and across-cluster separations?

The daily rainfall field is vectorized from 2D to 1D for clustering, which is purely a programming convention and does not result in any loss of spatial information. K-means operates in Euclidean space where distances between data points are independent of data dimensionality or spatial arrangement — each grid point serves as an independent axis in this high-dimensional space, and the distance between any two daily vectors therefore inherently reflects their spatial similarity across the entire rainfall field. This holds for most clustering methods unless spatial coordinates are explicitly encoded as additional features.

Regarding the choice of k-means: We understand that application-specific requirements influence the optimal choice which is why we do not advocate one clustering method over another. K-means was selected based on our prior experience demonstrating that it yields physically interpretable and stable rainfall clusters over Australia (Raut et al., 2014) and because it is extensible for climate model ensemble comparison (Raut et al., 2017), directly facilitating the model evaluation applications we envision for this dataset. We have also worked with SOMs, DBSCAN, and GMMs in other contexts, and in our experience k-means performs comparably to these methods for gridded rainfall fields.

Regarding within-cluster variance and between-cluster separation: these are directly and quantitatively addressed in the elbow plot (Figure 2) through WCSS and BCSS as explicit functions of k. The clear elbow at  $k = 11$  confirms an optimal balance between intra-cluster compactness and inter-cluster distinctiveness. Cluster stability was further verified through repeated initializations with fewer than 0.1% of days changing cluster affiliation, providing strong evidence that the clusters are robust and well-separated.

---

## Comment 3: Markovian Dynamics

What is the basis of considering the first-order Markovian dynamics of the clusters? What is the physical justification of it? How many days does each cluster persist and why? The kind of transition dynamics that has been estimated - how can it be justified using known processes (oscillations etc) associated with the Monsoon?

Markov chain framework is used for computing the transition probability matrix and serves as a compact, quantitative diagnostic of regime persistence and sequencing (Franzke et al. 2008), one that can be directly compared against transition matrices derived from model simulations. This provides a

rigorous methodology for evaluating whether models correctly reproduce not just the spatial patterns of ISM rainfall regimes, but also their sequential evolution, that mean-state comparisons alone cannot capture. Here it is not intended as a predictive tool.

The transition probability matrix is a way of showing repetition of each cluster and reveals both persistence and transience in a physically meaningful way. Break monsoon (Cluster 3) shows the highest self-transition probability ( $\sim 0.75$ ), consistent with observed break durations of several days to weeks while Depression-related clusters (11, 4, 9) show lower persistence ( $\sim 0.4$ ) with sequential transition 11, 4, 9, 3. This is a metastate as described in (Franzke et al. 2008) and mirrors the well-documented westward propagation of depressions from the Bay of Bengal through central India toward the northwest, followed by a break.

As per reviewer's request we did a quick computation that shows the mean duration of each cluster repetition (in days) as follows.

Cluster	Mean Duration (days)
1	1.54
2	1.53
3	3.93
4	1.61
5	1.74
6	2.06
7	2.22
8	2.07
9	1.64
10	1.78
11	1.73

Franzke, Christian, et al. "A hidden Markov model perspective on regimes and metastability in atmospheric flows." *Journal of Climate* 21.8 (2008): 1740-1757.

---