Integrating Wavelet-Entropy SOM Clustering with Association Rule Mining for Assessing Extreme Monthly Precipitation Under Major Teleconnection Forcing in Shikoku, Southern Japan

○Seyed Amirreza Tabataba Vakili, Sameh A. Kantoush, Sohei Kobayashi, Mohamed Saber, Hadir Abdelmoneim, and Vahid Nourani

1- Introduction

Extreme precipitation has exerted substantial hydrological and societal impacts in southern Japan, particularly over Shikoku Island, where steep topography and the interaction of multiple oceanic and atmospheric circulation systems generated pronounced spatial variability in rainfall (Mikami, 2023). The island is influenced by the Baiu frontal system, tropical cyclones, and moisture transport from surrounding seas, resulting in strong contrasts between Pacific-facing, mountainous, and leeward regions. Mean annual precipitation during 1989–2024 exceeded 3000 mm over the southern Pacific-facing areas, while markedly lower totals characterized the northern regions adjacent to the Seto Inland Sea (Figure 1). This pronounced spatial heterogeneity indicated that clustering the region into homogeneous areas and regional-scale analysis was essential for understanding the mechanisms governing extreme precipitation over Shikoku Island.

2- Methodology

Monthly precipitation (MP) data from selected stations covering 1989–2024 were analyzed. To account for strong spatial heterogeneity and reduce bias caused by the high dimensionality and nonstationary nature of precipitation time series, the wavelet-entropy-based Self-Organizing Map (WE–SOM) clustering framework was employed (Nourani et al., 2015; Lee et al., 2024). Each precipitation series was decomposed using the Discrete Wavelet Transform, and wavelet entropy (WE) was calculated for the resulting subseries to summarize multi-scale variability into a compact feature set (Nourani et al., 2023), following Shannon's entropy, providing a compact representation of multi-scale variability while reducing redundancy. For each cluster, the centroid was selected as the representative station. To investigate the large-scale and regional drivers of extreme monthly precipitation (EMP), a sensitivity analysis was conducted using sea surface temperature (SST) anomalies, oceanic atmospheric teleconnection indices (OATI), and lower-tropospheric dynamical variables. Redundant predictors were screened using a correlation-based dissimilarity metric. EMP was then identified using a percentile-based threshold derived from empirical cumulative distribution functions (ECDF). All retained predictors were discretized into categorical classes based on their mean and standard deviation before association rule mining (ARM). ARM was applied to identify combinations of antecedent climatic and dynamical conditions specifically associated with EMP, which classic linear and nonlinear approaches could not effectively capture, since they analyze the entire time series rather than focusing on rare EMP events. Each rule was evaluated using support and confidence criteria. Only rules exceeding predefined minimum support and confidence thresholds were retained, allowing identification of nonlinear and lag-dependent relationships relevant to EMP.

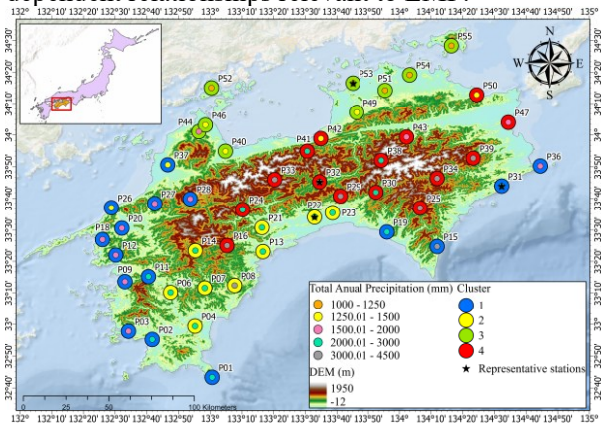

Figure 1. Total annual precipitation across Shikoku Island, together with the WE–SOM clusters and their corresponding representative stations.

3- Results

The WE–SOM clustering identified four hydro-climatic regions across Shikoku Island, and representative stations were selected for each cluster (Figure 1). Among these, Cluster 4 (mountainous interior) and Cluster 3 (northern leeward region facing the Seto Inland Sea) represented the most and least vulnerable regions to extreme EMP, respectively, and were therefore examined in detail. Analysis of MP at the representative stations showed that EMP was strongly concentrated during the wet season (June–September) and pronounced contrasts between the two clusters (Figure 2). Cluster 4 exhibited the highest wet-season medians, the widest interquartile range, and the most frequent extreme values, reflecting strong orographic enhancement under favorable synoptic conditions. In contrast, Cluster 3 showed consistently lower precipitation totals and reduced variability, indicating a persistent rain-shadow effect that limited the occurrence of EMP. Differences between the two clusters were minimal during the dry season, confirming that vulnerability to EMP was primarily governed by warm-season processes.

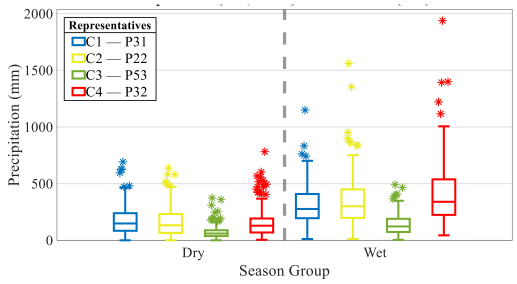

Figure 2. MP at the representative stations for the wet and dry seasons.

Sensitivity analysis indicated clear differences in the large-scale controls of EMP between the two clusters. Correlation-based screening of predictors identified redundancy among several SST and OATI, leading to the retention of Philippine Sea (PS) SST and Sea of Japan (SJ) SST as representative thermal drivers (Figure 3). EMP events, defined using the 90th percentile threshold, were predominantly associated with Baiu frontal activity and typhoon-related rainfall. For example, the 90th percentile threshold corresponded to 521 mm at the Cluster 3 representative station (P32) and 196 mm at the Cluster 4 representative station (P53).

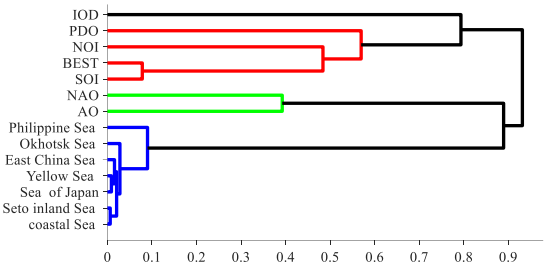

Figure 3. Hierarchical dendrogram based on the dissimilarity metric showing similarity among SST and teleconnection predictors.

Association rule mining further revealed contrasting sensitivity structures. The dominant predictors and their lag-dependent effects for both clusters are summarized in Table 1. For Cluster 4, EMP was strongly associated with low-level convergence at 1000 and 850 hPa (D1000 and D850) combined with PS SST anomalies at short (0–1 month) to intermediate (3–6 months) lags, highlighting the role of moisture-rich onshore flow and orographic lifting. In contrast, EMP in Cluster 3 showed weaker dependence on tropical SST anomalies and stronger associations with SJ SST variability and Indo-Pacific circulation patterns like Indian Ocean Dipole (IOD) and ENSO-related variability represented by the Bivariate ENSO Timeseries (BEST), and the Pacific Decadal Oscillation (PDO), indicating that only large-scale circulation anomalies capable of overcoming the orographic barrier could generate extreme precipitation.

Table 1. Summary of dominant variables for EMP

Lag	C4-P32	C3-P53
Lag 0	D1000, D850, PS SST, SJ SST, NOI, AO	D850, D1000, SJ SST, IOD, PDO, AO
Lag 1	D1000, PS SST, BEST, AO, NOI, IOD	D850, D1000, PS SST
Lag 3	D1000, D850, SJ SST, NOI, BEST	D1000, SJ SST, PS SST, IOD, BEST
Lag 6	PDO, BEST	D850, D1000, IOD

4- Conclusions

This study demonstrated pronounced spatial heterogeneity in EMP over Shikoku Island, with wet-season processes dominating across all regions and large-scale climatic drivers exerting clear region-specific influences. PS SST anomalies primarily modulated EMP in Pacific-facing and windward regions, whereas leeward extremes were more strongly associated with SJ SST variability and large-scale circulation anomalies, and the mountainous interior reflected combined tropical and extratropical influences under strong low-level convergence. Overall, the integration of WE-based clustering with event-focused sensitivity analysis provided a physically consistent framework for resolving regional EMP mechanisms in complex topography. Future work could incorporate fuzzy logic representations to explicitly account for uncertainty and improve EMP prediction.

References

- Lee, S., Nourani, V., Danandeh Mehr, A., Moriasi, D., & Mirchi, A. (2024). Wavelet-Entropy Enhanced Clustering: A Comprehensive Analysis of Drought Patterns in the Southern Plains, United States. Journal of Hydrometeorology, 25(12), 1809-1822.
- Mikami, T. (2023). The Climate of Japan: Present and Past (Vol. 77). Springer Nature.
- Mouri, G., Kanae, S., & Oki, T. (2011). Long-term changes in flood event patterns due to changes in hydrological distribution parameters in a rural–urban catchment, Shikoku, Japan. Atmospheric Research, 101(1), 164-177.
- Nourani, V., Alami, M. T., & Vousoughi, F. D. (2015). Wavelet-entropy data pre-processing approach for ANN-based groundwater level modeling. Journal of Hydrology, 524, 255-269.
- Nourani, V., Sayyah-Fard, M., Kantoush, S. A., Bharambe, K. P., Sumi, T., & Saber, M. (2023). Optimization-Based Prediction Uncertainty Qualification of Climatic Parameters. Journal of Hydrometeorology, 24(10), 1679-1697.