Clustering Methods for the Analysis of Spatio-Temporal Distribution of Precipitation and Temperature Data in Northeast India

Submitted

By

Rohan Kumar Ishwar

Under the supervision of

Dr. Sreeja Pekkat

ABSTRACT

This study explores the application of advanced clustering techniques—K-Means, DBSCAN, and Hierarchical Clustering—to analyze the spatio-temporal distribution of precipitation and temperature in Northeast India, with the objective of uncovering regional climate patterns and trends. Each method is evaluated for its efficacy in managing spatial heterogeneity, temporal variability, and noise. Comparative analysis highlights DBSCAN's strength in detecting irregular and non-linear patterns, K-Means' efficiency in identifying structured clusters, and Hierarchical Clustering's capacity to reveal nested relationships. The findings underscore the potential of clustering to identify precipitation hotspots and temperature gradients, providing actionable insights for climate adaptation strategies, resource planning, and disaster mitigation. The report concludes by emphasizing the practical applicability of clustering techniques for analyzing complex spatio-temporal climatic data and outlines plans for further analysis in the next phase.

Understanding the Study Area: Northeast India

Geographic Features

Northeast region of India which consists of the 8 states Assam, Meghalaya, Arunachal Pradesh, Tripura, Nagaland, Mizoram, Manipur Manipur and Sikkim is selected for the study. This region extends from latitude 22.4°N to 28.7°N and longitude 88.2°E to 88.2°E to 96.5°E. Northeast India encompasses with diverse topography, topography, ranging from plains to mountains.

Climate Variability

The region experiences significant climate variability, with extreme extreme weather events such as floods and temperature fluctuations fluctuations posing challenges to its ecosystems and human populations. This region is characterized by heavy rainfall, influenced by influenced by the monsoon season.

Data Preprocessing: Preparing for Analysis

Missing Values

Addressing data gaps caused by sensor sensor malfunctions or poor data transmission was crucial. Temporal interpolation and grid-wise mean imputation were employed to ensure data data integrity.

Temporal Aggregation

Daily data was aggregated into monthly and seasonal averages to highlight seasonal patterns and intra-annual variability, simplifying the analysis of precipitation and temperature trends.

Standardization

Z-score normalization was applied to reduce biases caused by extreme values values and ensure equal weighting of variables across grid points, leading to to more robust clustering results.

Trend Analysis: Identifying Significant Changes

1 Mann-Kendall Test

This test identified significant trends in precipitation, revealing increasing trends in high-altitude northern regions and decreasing trends in southern plains.

2 Sen's Slope Estimator

Quantified the magnitude of these trends, providing a metric for metric for comparing spatial climatic variations and pinpointing pinpointing regions experiencing intensified precipitation or precipitation or prolonged dry spells.

Clustering Techniques: Delineating Climate Zones

K-Means

This method effectively segmented structured regions, identifying six distinct clusters based on precipitation and temperature distributions.

DBSCAN

DBSCAN proved effective in isolating anomalous grid points as noise, highlighting highlighting unique patterns in highly variable variable regions and identifying rare climatic climatic phenomena such as flash floods.

Hierarchical Clustering

This method uncovered nested climatic behaviors, providing granular insights into sub-into sub-regional climate dynamics and multi-multi-scale patterns across the region.

Key Findings: Unveiling Spatial-Temporal Patterns

1 **Precipitation Hotspots**

High rainfall zones like Cherrapunji and Mawsynram were identified as critical areas for hydrological planning, highlighting the highlighting the importance of managing these resources.

2 **Dry Zones**

Valleys and plains exhibited reduced precipitation, emphasizing the need for adaptive strategies in water-scarce regions, particularly regions, particularly for agricultural practices and resource management.

3 **Clustering Insights**

The study demonstrated the value of advanced clustering methods as robust analytical tools for interpreting climate data and climate data and informing strategic planning for climate adaptation and mitigation.

Comparative Analysis: Choosing the Right Tool

1

K-Means

Efficient for structured data, but less adaptive to noise.

2

DBSCAN

Robust for noise and non-linear patterns, but parameter sensitive.

3

Hierarchical Clustering

Ideal for exploration, uncovering nested patterns, but computationally computationally demanding.

4.4 Comparative Analysis of Clustering Methods

The clustering techniques provided complementary insights (Fig. 14):

- **K-Means**: Effective for structured datasets; computationally efficient but less adaptive to noise.
- **DBSCAN**: Robust for noise handling and non-linear patterns; parameter sensitive.
- **Hierarchical Clustering**: Ideal for exploratory analysis and nested patterns; computationally demanding.

Feature	DBSCAN	Hierarchical Clustering	K-Means Clustering
Clusters	3 (+ noise)	5	6
Noise Handling	Yes (noise points)	No	No
Shape Flexibility	Arbitrary	Dependent on linkage criterion	Spherical
Parameter Sensitivity	High (*eps*, *min_samples*)	Moderate (linkage method)	High (k value)
Interpretability	Moderate	High (dendrogram)	High
Computational Cost	Moderate	High	Low

Fig. 14 Key Comparisons Between Clustering Methods

Next Steps: Expanding the Research

1

Temperature Data Analysis

Extend similar analyses to temperature datasets, uncovering patterns and trends in temperature variability across Northeast India. Northeast India.

2

Data Integration

Incorporate additional data sources, such as land use patterns and atmospheric circulation models, to enhance clustering accuracy and provide a more comprehensive understanding of climate dynamics.

3

Real-World Validation

Validate findings through empirical evaluations under real-world climatic conditions, ensuring the practical applicability of the applicability of the research for climate planning and mitigation strategies.