Neda Mohseni
CS 179 - Group 98

Structure Learning of Spatial Dependencies over Latent Seasonalities

Introduction

I have daily rainfall data collected from 54 weather stations across India spanning a 10-year period from 1985 to 1995, covering all days of the year. To simplify the modeling process and emphasize the occurrence of rainfall events, the raw daily rainfall measurements are discretized into binary labels, where a value of 1 indicates a rainy day (defined as more than 5mm of rainfall), and 0 indicates a non-rainy day. After discretization, the dataset becomes highly sparse, with only 16.04% of all station-day entries labeled as rainy.

The dataset consists of multivariate time series data in form of d-dimensional binary vectors for each discrete time t, where d=54 corresponds to the number of weather stations, and t=3652 Represents the number of discrete daily observations. This project aims To model both temporal dependencies over time and spatial dependencies across stations. Prior work has been done by Kirshner et al., to find spatio-temporal Rainfall dependencies in south-western Australia and Western U.S. [1], This project uses a similar approach to uncover spatio-temporal rainfall patterns across India.

figure 1 - rainfall stations across India

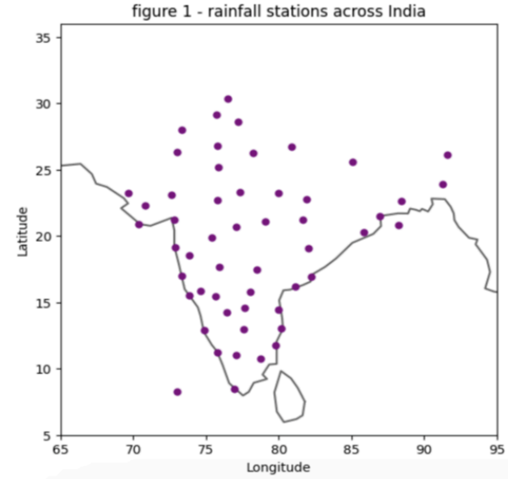


figure 2 - Binary Rainfall Events (>5mm)

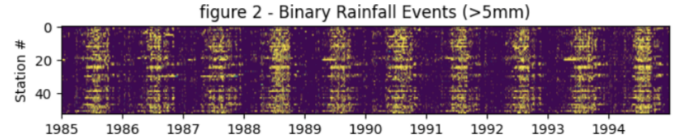
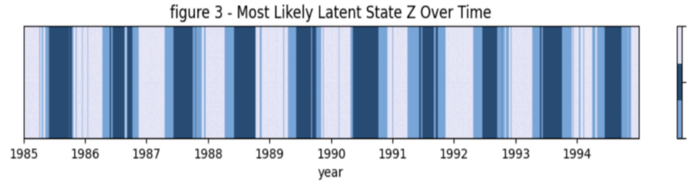


figure 3 - Most Likely Latent State Z Over Time



Approach

Directly modeling temporal dependencies across all stations for each of the 3652 daily observations with markov chains would be highly complex, requiring an exponential number of parameters (Ihler et al. [2]).. Figure 2 shows actual rainfall patterns tend to repeat over the 10-year period. To simplify the temporal modeling process, **Hidden Markov Models (HMM)** are used to reduce time into K latent temporal states, each representing a seasonality (figure 3). This approach captures temporal dependencies using O(K^2) parameters instead of exponential (Kirshner et al. [1]). The resulting lower-dimensional temporal representation enables us to model spatial dependencies across stations within each seasonality, significantly reducing complexity while still capturing both temporal and spatial dependencies.

The HMM is learned via **Baum-Welch Algorithm**, a special type of EM algorithm used to find the optimal parameters of HMM. In our case, the emission distribution $p(x|z=k)$ for each latent state z is modeled using a **Chow-Liu tree**, which captures spatial dependencies across stations with a tree-structured Ising model. During the M-step, each Chow-Liu tree is re-fit using weighted data, where the weights are given by the posterior probabilities of the corresponding state from the E-step. This allows the tree structure for each latent state to adapt over the EM iterations, reflecting changes in the soft assignment of data to latent states and enabling the emission model (Chow-Liu trees) to evolve as the latent temporal structure converges.

Neda Mohseni
CS 179 - Group 98

For the chow-liu emission model, since computing the true log-likelihood for Ising models requires evaluating the partition function, which is computationally expensive, I use the pseudo-log-likelihood (PLL) method from pyGMs library as a surrogate. As a result, both EM training and evaluation rely on emission probabilities derived from PLL rather than the true joint likelihood. All values reported in this report for model evaluation and model selection are based on this PLL approximation.
For initialization, each state's emission is fit to a random 1000-day subset to ensure diverse starting Chow-Liu trees. We use a very strong prior on the transition matrix in order to encourage the model to pick an explanation for the data that has long temporal consistency.

Latent Temporal States

For finding the best number of latent states K, I evaluated the BIC score and performed LOOCV. The 10-year dataset was divided into 10 folds (years), and For each fold, cross-validation was performed by training the HMM on 9 years and testing on the held-out year. The mean PLL per day across all folds was used to compare model performance. The BIC score favored k = 3 while the LOOCV favored K = 4. I chose k = 3 based on qualitative inspection. The K=3 model produced more interpretable and well-separated latent states that aligned better with known seasonal patterns in India. the model with K=4 latent states resulted in one redundant states, making the overall segmentation is less clear and harder to interpret.

	Mean PLL
K = 2	-17.08
K = 3	-16.78
K = 4	-16.71
K = 5	-16.86

LOOCV result

	BIC	Mean PLL
K = 2	123,780	-16.71
K = 3	118,692	**-15.89**
K = 4	118,951	-15.80
K = 5	119,264	-15.71

BIC result

Figure 4 shows the Actual Rainfall Amounts in mm vs The hidden temporal states learned, both over year 3. The rainy/monsoon season in India occurs between June to September, represented by Z = 1 (dark blue), the Transitional Season with mild rain occurs pre & post monsoon, is Represented by z = 0 (blue), and the dry season occurring From December to March is represented by z = 2 (lilac). The full sequence of actual rainfall Vs. Latent states pattern Over all 10 years can be seen in Appendix I.

Figure 4

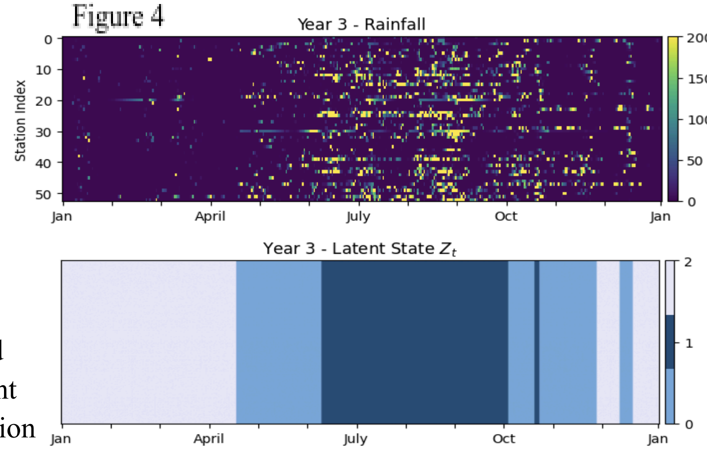
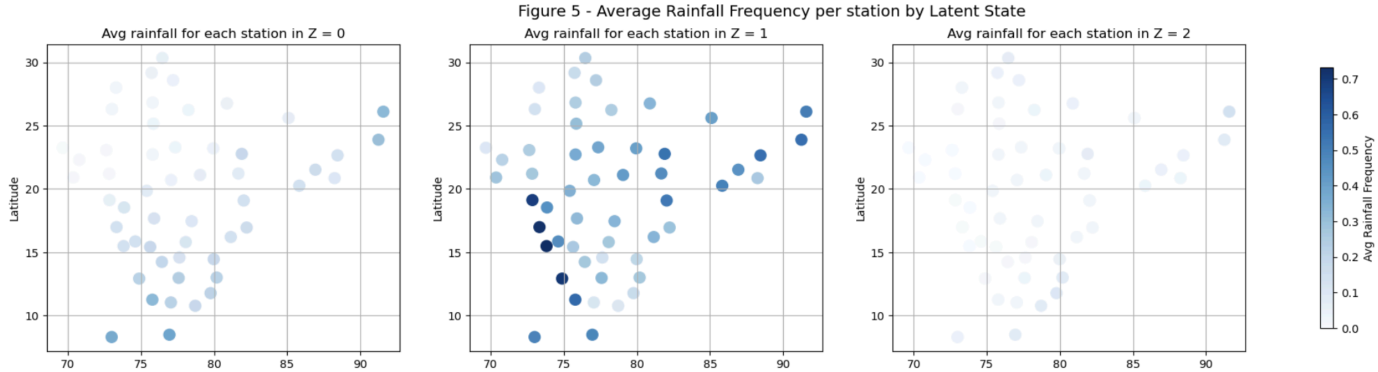


Figure 5 shows the spatial visualization of stations, colored by **average rainfall frequency** across station for each latent state. For each state, the average rainfall frequency per station is computed as the proportion of rainy days over all days assigned to that state. 1008 days have been assigned to z = 0, 1150 days to z = 1 and 1494 days to z =2. The monsoon state (z =1) displays the highest average rainfall concentrated south-west coast, while the dry state (z =2) shows low rainfall frequency across all stations and transitional state (z = 0) shows mild rain concentrated in the south and north-east.

Figure 5 - Average Rainfall Frequency per station by Latent State

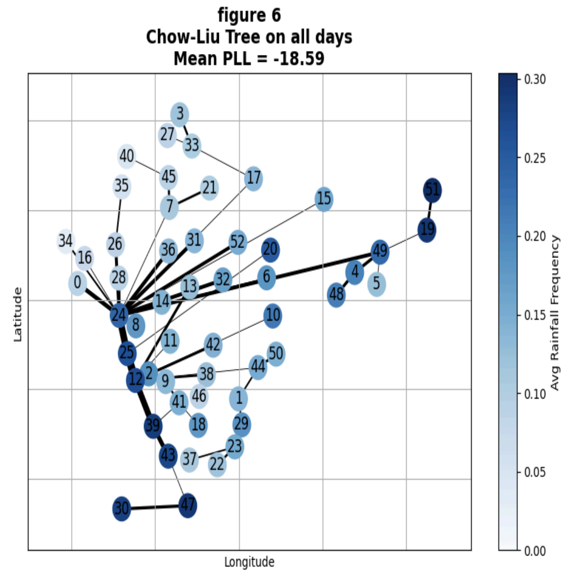


Neda Mohseni
CS 179 - Group 98

Spatial Dependencies

The emission distribution ***p(x|z=k)*** for each latent state z is modeled using a Chow-Liu tree, which captures spatial dependencies across stations via a tree-structured Ising model. The edges of the tree are selected to form a **maximum-weight spanning tree** based on **pairwise mutual information**.

Figure 6 shows a Chow-Liu tree fitted to the entire dataset without accounting for any latent temporal structure. This model has a mean PLL of –18.59, which is calculated as the average PLL per day across all 3652 days. Figure 7 represents the Chow-Liu tree for each latent state learned by HMM, with individual PLLs: –20.62 (z = 0), –21.21 (z = 1), and –30.92 (z = 2). While each model fits fewer data points and thus achieves a lower PLL, the EM-trained HMM returns an improved overall mean PLL of –15.89, slightly better than The mean PLL per day from global Chow-Liu (-18.59), Indicating That incorporating latent temporal structure led to a better overall fit.



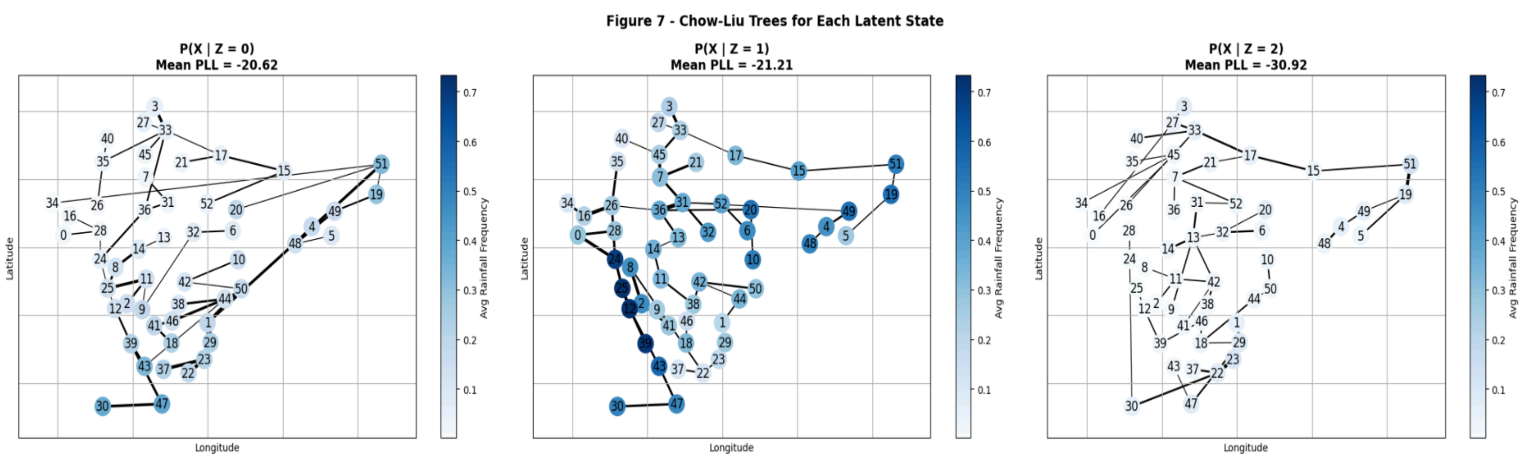


Figure 7- node colors represent the average rainfall frequency at each station across all days assigned to the corresponding latent state, edge width indicates the strength of Mutual Information between stations (Rainfall frequency and mutual information are globally normalized across all states for comparability). Z = 0 corresponds to transitional season, Z = 1 corresponds to Rainy/Monsoon season, Z=2 corresponds to dry season.

Despite each latent state conditioned Chow-Liu model achieving a lower PLL than the global model, meaningful spatial correlations across Stations within each state can be observed. The global Chow-Liu Tree includes edges with strong mutual information connecting distant stations, while each state specific tree, especially the one for rainy/monsoon states (Z=1), tends to capture more local Dependencies between neighboring stations with strong mutual information edges within the southwest coast of India, which is known for heavy rain. The pre/post-monsoon Transitional state (Z=0) includes longer edges connecting the stations from south to the northeast, another region with significant rainfall. Figure 8 shows the most rainy regions Across India. [3]

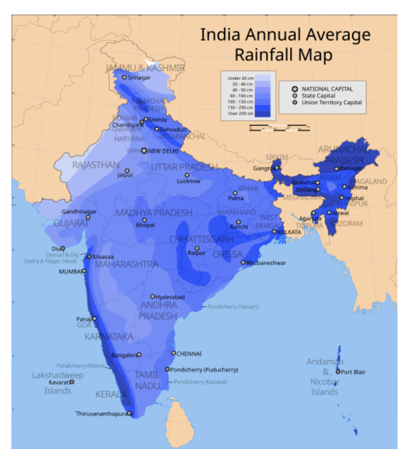


Figure 8[3]

Adding BIC penalty to Chow-Liu Trees

Chow-Liu inevitably chooses a fully connected tree. However, we can add a BIC complexity penalty to our chow liu function that allows it to select a forest, i.e., select fewer edges when they do not improve the log likelihood sufficiently(Ihler [4]). Figure 9 and 10 show the Chow-Liu Tree's for each latent state with a penalty of 0.015 and 0.025, respectively.

Figure 9 - Chow-Liu Trees for Each Latent State. penalty = 0.015

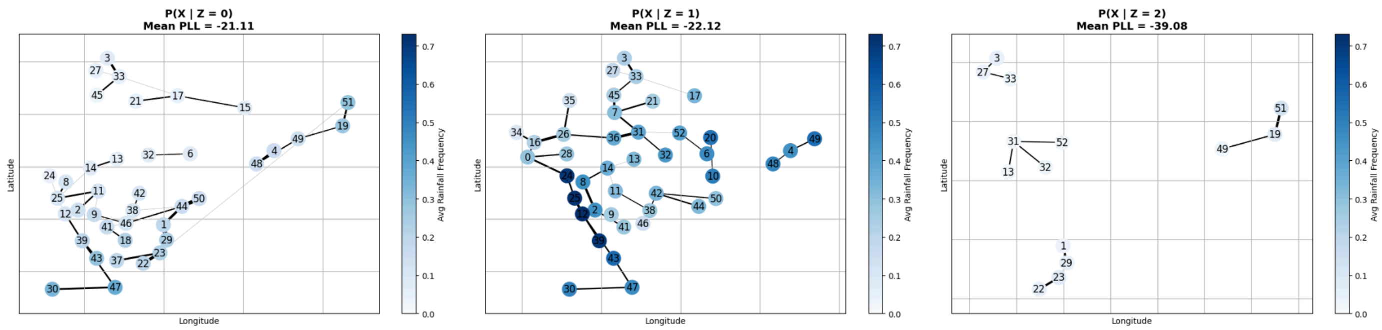
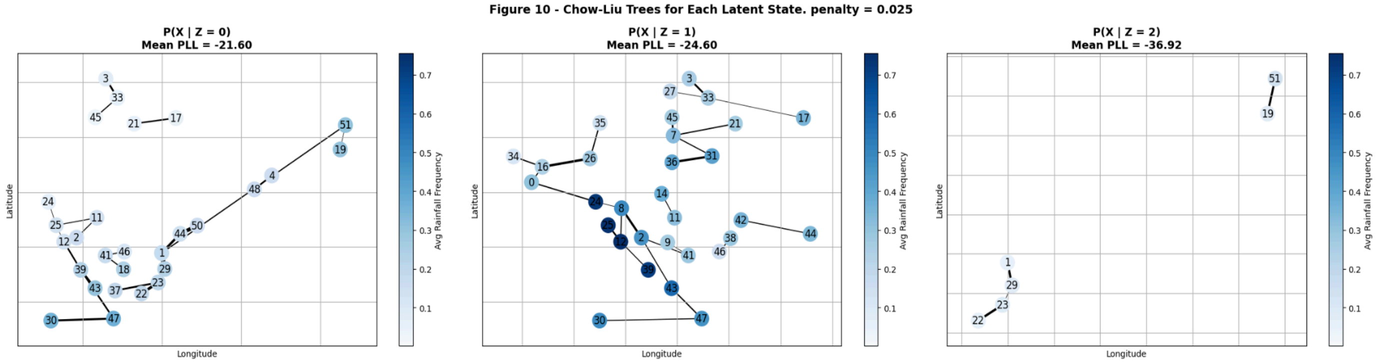


Figure 10 - Chow-Liu Trees for Each Latent State. penalty = 0.025



After applying the BIC penalty, the Chow-Liu trees became noticeably sparser. The tree for the monsoon state (z=1) still contains edges between neighboring stations in the southwest coast, the main rainfall region in India. In the transitional state (z=0), the resulting tree primarily captured connections among stations in the southwest and northeast, India's main rainy regions. The edges align with the map shown in figure 8, despite the model having no access to latitude/longtitude info. The dry state (z=2), was most affected by the BIC penalty, with only a few edges remaining, showing low correlation between stations during the dry period.

Future work:

As future work, I plan to try other pyGMs Ising models as emission distribution such as sparse logistic regression neighborhood selection or use structure learning methods from other packages. I also plan to use deep learning approaches, such as GNNs and LSTMs, to model spatio-temporal dependencies or predict rainfall.

Code resource:

HMM implementation: code adapted from [Activity HMM Demo] with changes to emission class.
Chow-Liu implementation: code borrowed from professor Ihler's [pyGMs Ising class] and adapted to get edge Mutual Information values for visualization purposes.
code for loading data / some visualization borrowed from [Ising Rainfall Demo].

References

[1] Kirshner, S., Smyth, P., & Robertson, A. W. (2004). Conditional Chow–Liu tree structures for modeling discrete-valued vector time series. *Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence (UAI'04)*, pp. 317–324. AUAI Press.
Also available at: https://arxiv.org/abs/1207.4142

[2] Ihler, Alexander T, et al. "*Graphical Models for Statistical Inference and Data Assimilation"* , www.ics.uci.edu/~ihler/papers/physd07.pdf. Accessed 11 June 2025.

[3] "Monsoon of South Asia." *Wikipedia*, Wikimedia Foundation, 2 June 2025, en.wikipedia.org/wiki/Monsoon_of_South_Asia.

[4] Ihler, Alexander T. " Learning from Data Demo." *GitHub*, github.com/ihler/pyGMs/blob/master/notebooks/06%20Learning%20from%20Data.ipynb.

Appendix 1

Latent temporal states over 10 years