Functional Data Analysis of Weather Data

Aleksandr Jan Smoliakov[1]

[1]Vilnius University, Faculty of Mathematics and Informatics

2025–05–27

Table of Contents

Data Source

- **Historical Weather Data in India**
  - Hourly observations (2006–2019)
  - 8 major Indian cities
  - Over 20 meteorological variables (focus on Temperature)

Recap: Data, Smoothing, and Initial FPCA

- **Data:** Hourly temperature data (*tempC*) for 8 Indian cities (2011-2018).
- **Preprocessing:**
  - Data cleaned and augmented (day of year, hour of day).
  - Averaged multi-year temperatures for each city, day of year, and hour of day, creating an average annual temperature surface (365 days $\times$ 24 hours) per city.
- **Bivariate Smoothing:**
  - Day dimension: B-spline basis (12 basis functions).
  - Hour dimension: Fourier basis (11 basis functions, 5 harmonics + intercept).
  - Optimal smoothing parameters ($\lambda_s = 0.001$, $\lambda_t = 10^{-5}$) selected via Generalized Cross-Validation (GCV).
  - Result: Smoothed bivariate functional data object for each city's average temperature surface ($Y_{smoothed\_city\_avg}$). MAE $\approx 0.51^\circ C$.
- **Initial FPCA (on City Averages):**

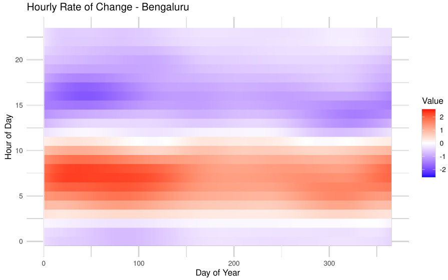
Recap: Data, Smoothing, and Initial FPCA

- **Data:** Hourly temperature data (*tempC*) for 8 Indian cities (2011-2018).
- **Preprocessing:**
  - Data cleaned and augmented (day of year, hour of day).
  - Averaged multi-year temperatures for each city, day of year, and hour of day, creating an average annual temperature surface (365 days $\times$ 24 hours) per city.
- **Bivariate Smoothing:**
  - Day dimension: B-spline basis (12 basis functions).
  - Hour dimension: Fourier basis (11 basis functions, 5 harmonics + intercept).
  - Optimal smoothing parameters ($\lambda_s = 0.001$, $\lambda_t = 10^{-5}$) selected via Generalized Cross-Validation (GCV).
  - Result: Smoothed bivariate functional data object for each city's average temperature surface ($Y_{smoothed\_city\_avg}$). MAE $\approx 0.51^\circ C$.
- **Initial FPCA (on City Averages):**

Derivatives: Rates of Temperature Change

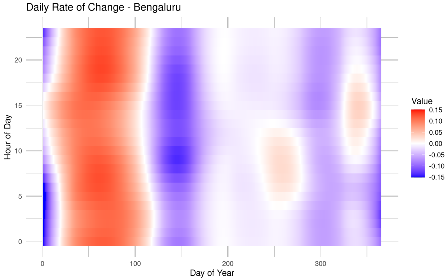
Understanding Dynamics for Bengaluru (Example City)

Hourly Rate of Change ($\frac{\partial T}{\partial \text{hour}}$)



Rate of temperature change throughout the day, across the year. Red: warming, Blue: cooling.

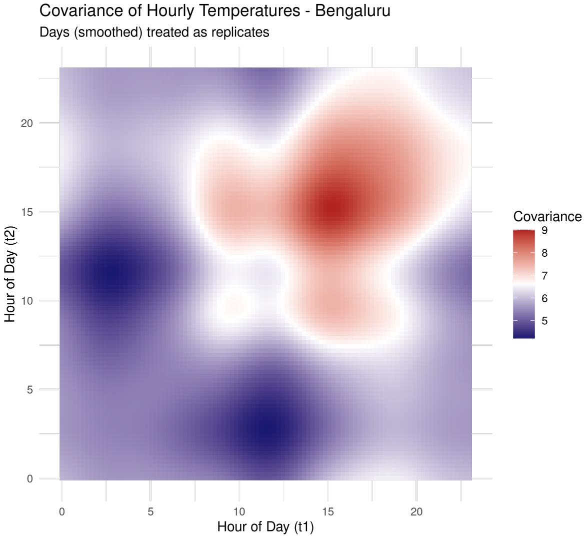
Daily Rate of Change ($\frac{\partial T}{\partial \text{day}}$)



Rate of temperature change from one day to the next, at different hours. Red: inter-day warming, Blue: inter-day cooling.

Covariance of Hourly Temperatures

For Bengaluru (Smoothed Daily Profiles as Replicates)

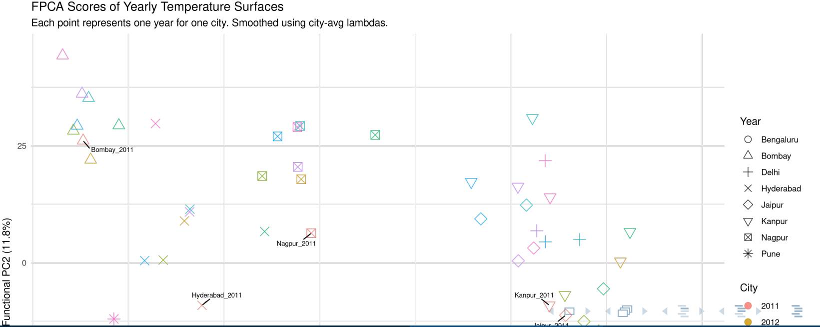


FPCA on Yearly Temperature Surfaces

Each City-Year Combination as a Replicate (64 surfaces)

Objective: Explore inter-annual variability and city-specific temporal trends.

- Yearly temperature surfaces (365 days $\times$ 24 hours) for each city and each year (2011-2018) were smoothed.
- FPCA performed on the coefficients of these 64 smoothed surfaces.

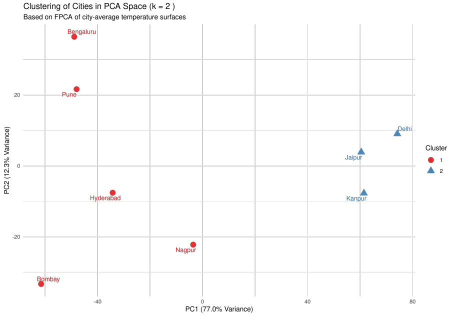


Clustering Cities by Average Temperature Profiles

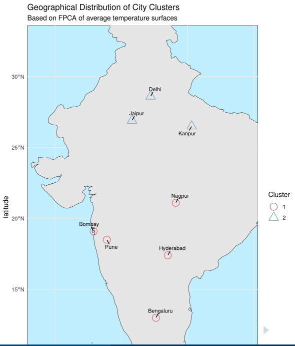
Based on FPCA Scores from City-Average Surfaces

- Hierarchical clustering (Ward.D2 method) on PC1, PC2, PC3 scores of the 8 city-average temperature surfaces.
- A chosen number of $k = 2$ clusters was selected.

Clusters in PCA Space (PC1 vs PC2)



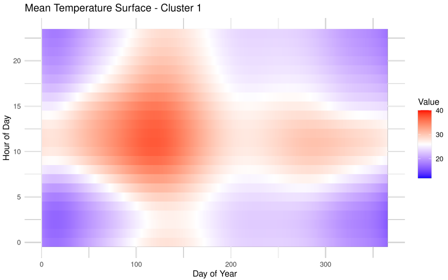
Geographical Distribution of Clusters



Comparing Cluster Characteristics

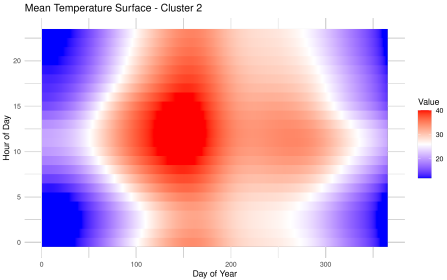
Mean Temperature Surfaces per Cluster

Mean Surface - Cluster 1



Cities: Bengaluru, Bombay, Hyderabad, Nagpur, Pune.

Mean Surface - Cluster 2



Cities: Delhi, Jaipur, Kanpur.

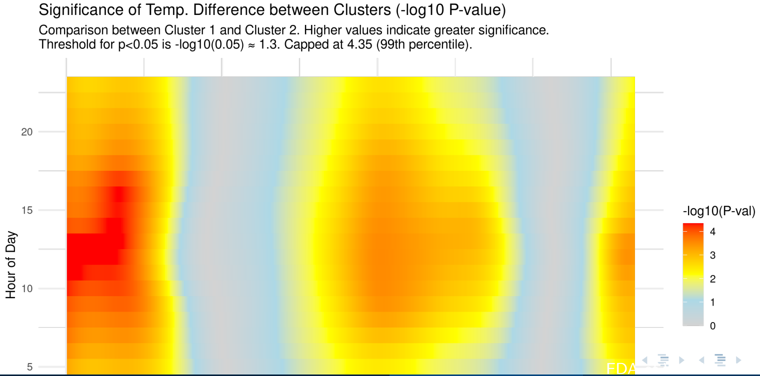
Observations:

- **Cluster 1 (Southern/Central):** Generally warmer winters, less extreme summer highs, possibly more influence from monsoonal patterns affecting daily temperature ranges.

Pointwise FANOVA: Significance of Temperature Differences

Testing H_0: Mean Temp Surface is Same for Both Clusters

- For each day of year and hour of day, an ANOVA was performed: $\mathit{Temperature} \sim \mathit{Cluster}$.
- The heatmap shows the $-\log_{10}$(p-value) for the cluster effect.



Conclusions and Future Work

Key Findings from Advanced FDA:

- **Derivatives & Covariance:** Revealed detailed intra-day and inter-day temperature dynamics and relationships.
- **Yearly FPCA:** Quantified inter-annual variability and highlighted distinct city-level climatic trajectories over the years.
- **Clustering:** Successfully grouped cities into geographically and climatically meaningful clusters based on their annual temperature surfaces.
  - Cluster 1: Southern/Central cities with milder variations.
  - Cluster 2: Northern cities with more extreme seasonal variations.
- **Pointwise FANOVA:** Confirmed statistically significant differences in temperature patterns between clusters, particularly during winter and peak summer daytimes.

Conclusions and Future Work

Key Findings from Advanced FDA:

- **Derivatives & Covariance:** Revealed detailed intra-day and inter-day temperature dynamics and relationships.
- **Yearly FPCA:** Quantified inter-annual variability and highlighted distinct city-level climatic trajectories over the years.
- **Clustering:** Successfully grouped cities into geographically and climatically meaningful clusters based on their annual temperature surfaces.
  - Cluster 1: Southern/Central cities with milder variations.
  - Cluster 2: Northern cities with more extreme seasonal variations.
- **Pointwise FANOVA:** Confirmed statistically significant differences in temperature patterns between clusters, particularly during winter and peak summer daytimes.

Potential Future Work:

- Incorporate other meteorological variables (humidity,

Thank You!

Thank you for your attention!