Functional Data Analysis of Weather Data

Aleksandr Jan Smoliakov[1]

[1]Vilnius University, Faculty of Mathematics and Informatics

2025–05–27

Table of Contents

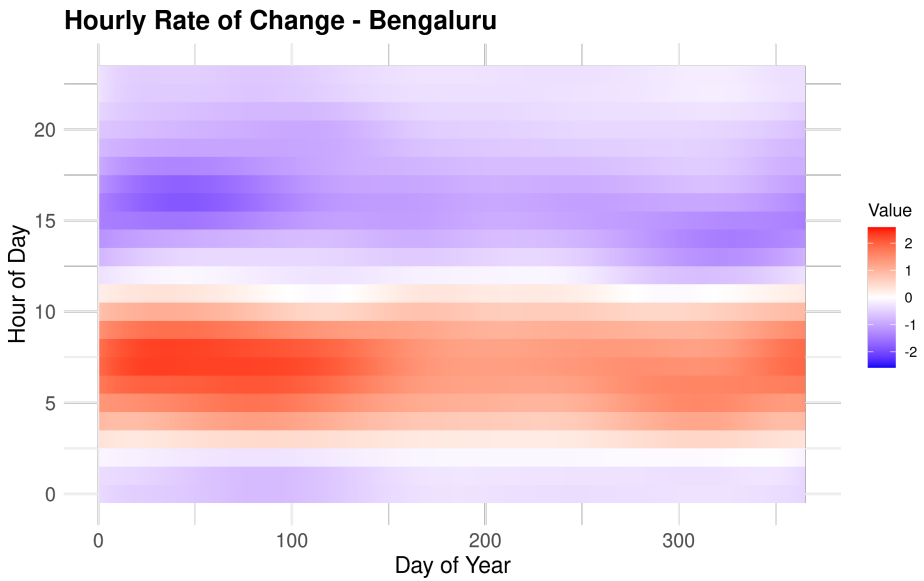
Recap: Data and Smoothing

- **Data:** Hourly temperature data (*tempC*) for 8 Indian cities (2011-2018).
- **Preprocessing:**
  - 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. MAE $\approx 0.51^\circ C$.

Recap: Temperature Derivatives

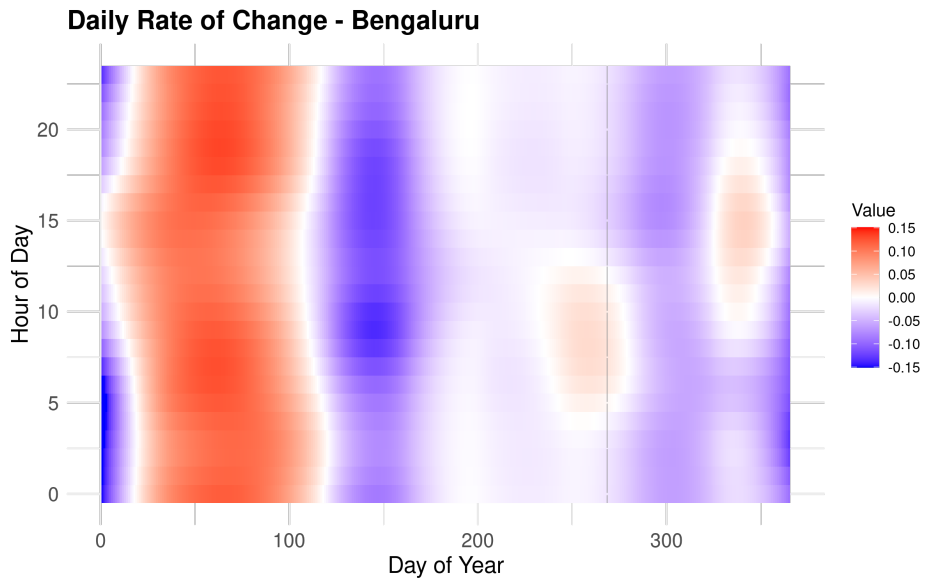
Example: Dynamics for Bengaluru

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.

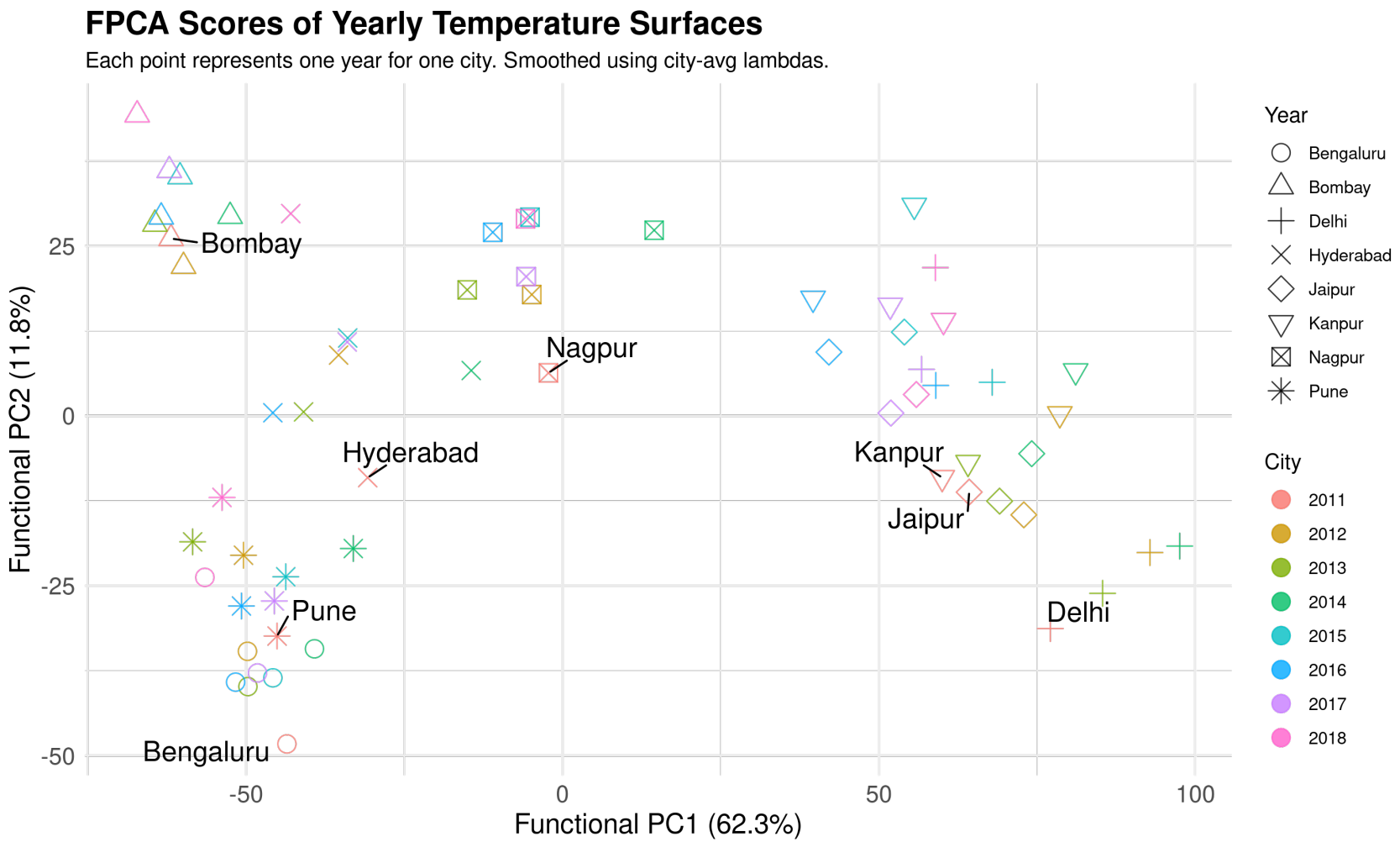
Recap: Initial FPCA

- **Initial FPCA (on City Averages):**
  - Performed on the coefficients of the smoothed city-average surfaces.
  - Revealed primary modes of variation in temperature patterns across cities.
  - PC1 captured 77.0% of the variance, mostly the yearly seasonal cycle.
  - PC2 captured 12.3% of the variance, mostly the inter-day variability.

Today's Focus: Deeper analysis building upon these smoothed functional representations.

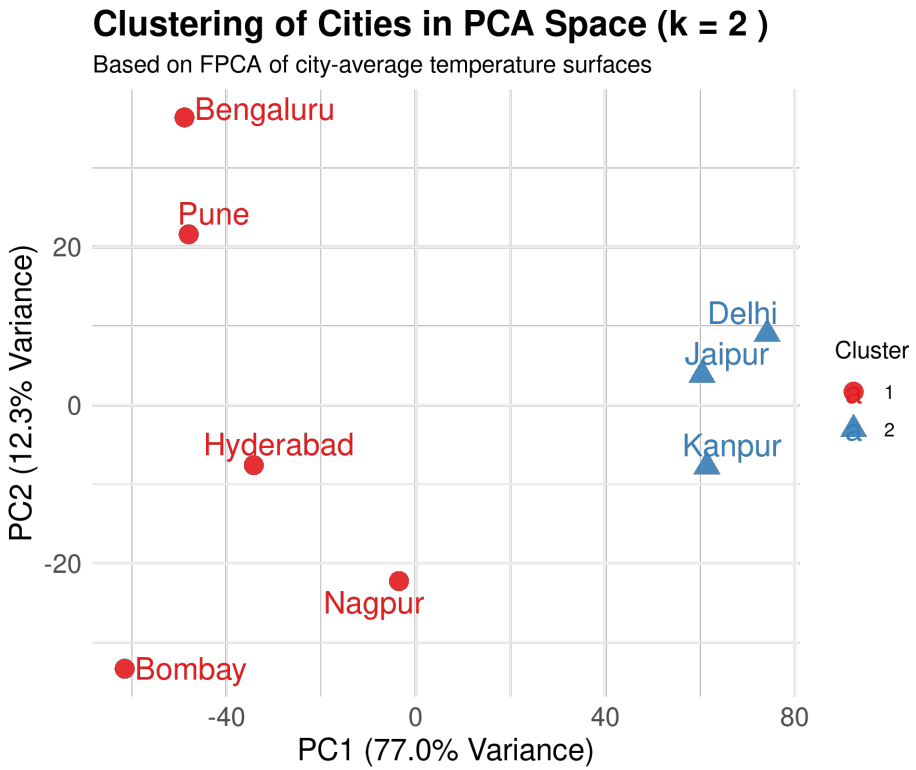
FPCA on Yearly Temperature Surfaces

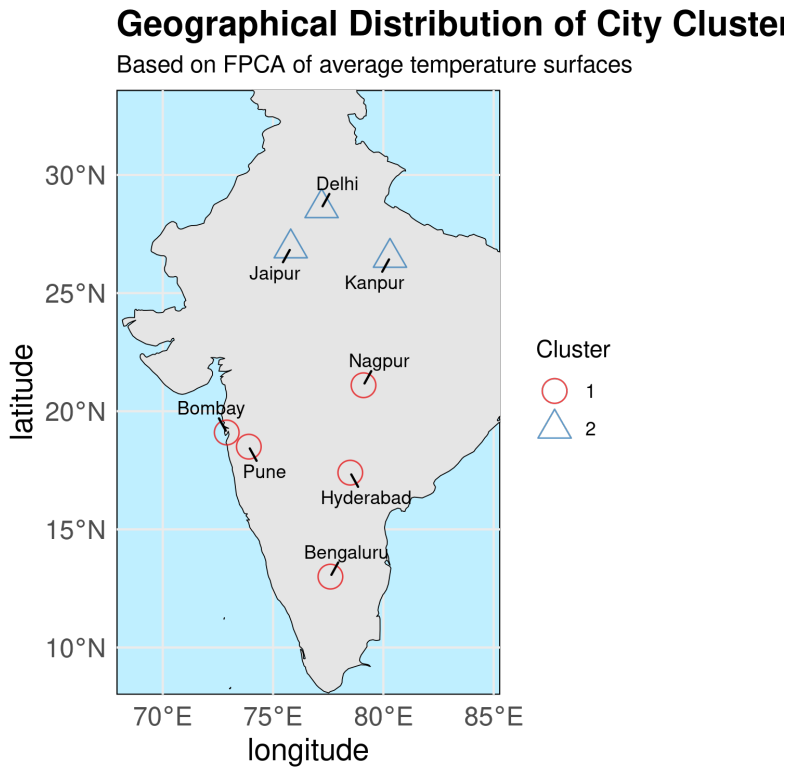
FPCA performed for each city and each year (2011-2018).



Clustering Cities by Average Temperature Profiles

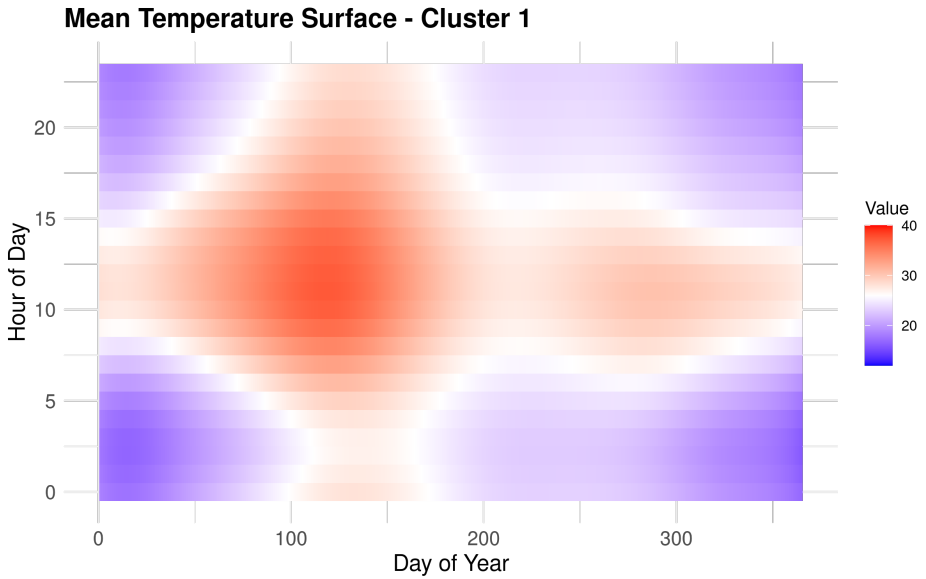
Hierarchical clustering ($k = 2$) on PC1, PC2, PC3 scores of the 8 city-average temperature surfaces.



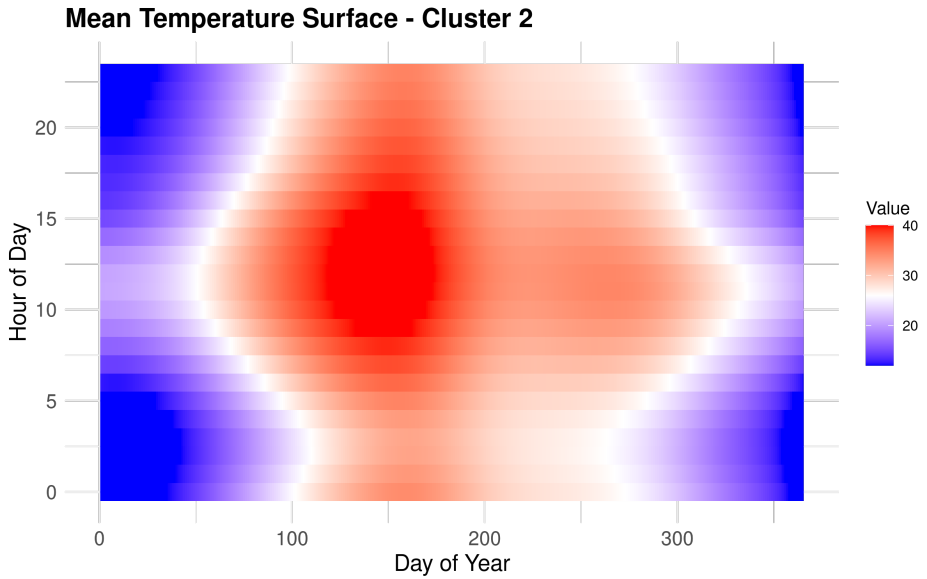


Identified Clusters: Southern/Central vs Northern/Inland cities.

Mean Temperature Surfaces per Cluster



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



Cities: Delhi, Jaipur, Kanpur.

Observations:

- **Cluster 1 (Southern/Central):** Warmer winters, less extreme summer highs, temperature dip during monsoon season.
- **Cluster 2 (Northern/Inland):** More pronounced seasonality with colder winters and hotter summers.

Pointwise FANOVA for *Temperature* ∼ *Cluster*

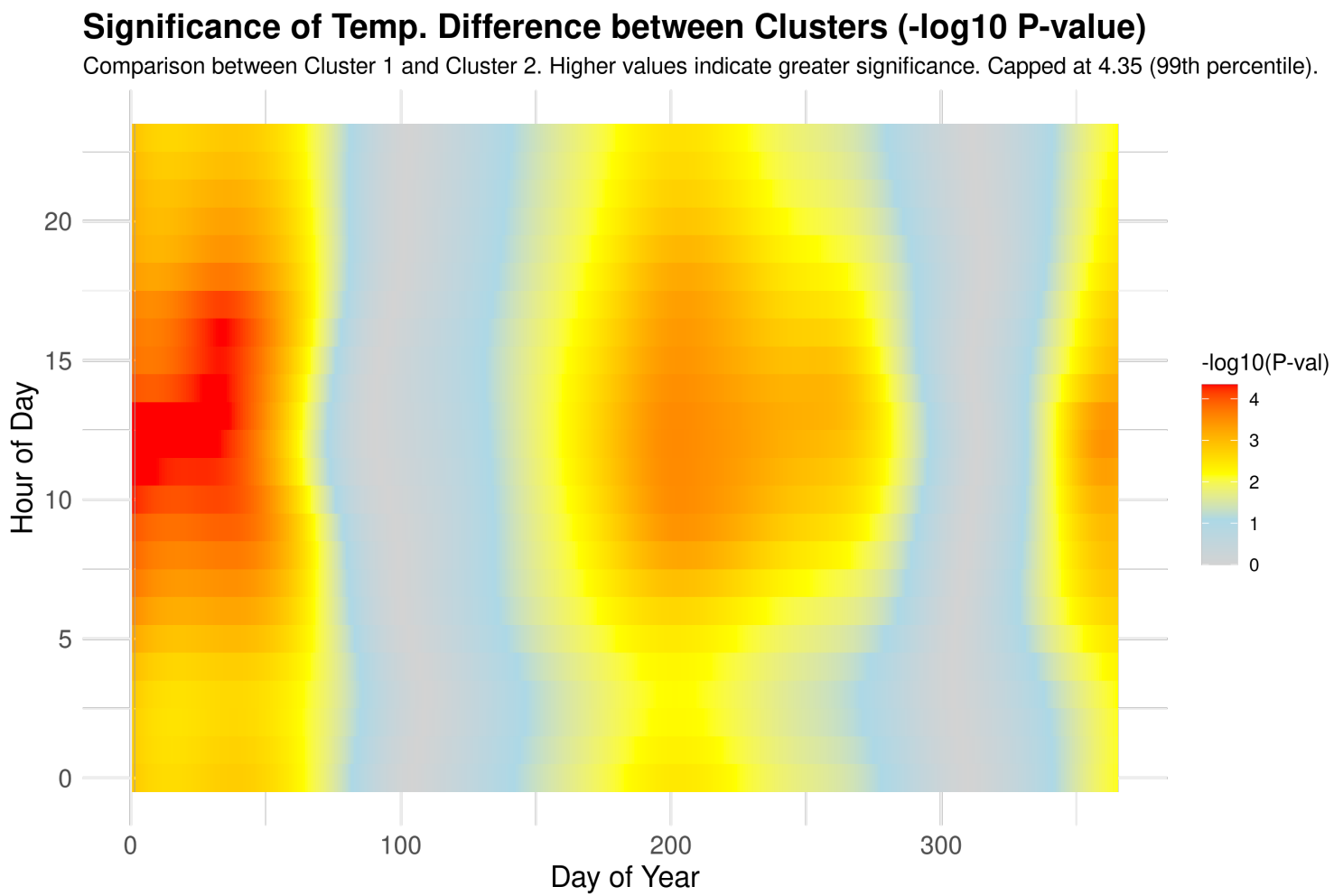


Figure: Higher values indicate stronger statistical significance of temperature difference between clusters.

Conclusions and Future Work

Key Findings:

- **Derivatives & Covariance:** Revealed 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:** Grouped cities into clusters based on their annual temperature surfaces.
  - Southern/Central cities with milder variations.
  - 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:

- **Derivatives & Covariance:** Revealed 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:** Grouped cities into clusters based on their annual temperature surfaces.
  - Southern/Central cities with milder variations.
  - 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:

- Functional regression models (e.g., predicting energy demand).
- Anomaly detection for unusual yearly temperature patterns.

Thank You!

Thank you for your attention!