Functional Data Analysis of Weather Data

Aleksandr Jan Smoliakov[1]

[1]Vilnius University, Faculty of Mathematics and Informatics

2025–04–29

Table of Contents

Data Source

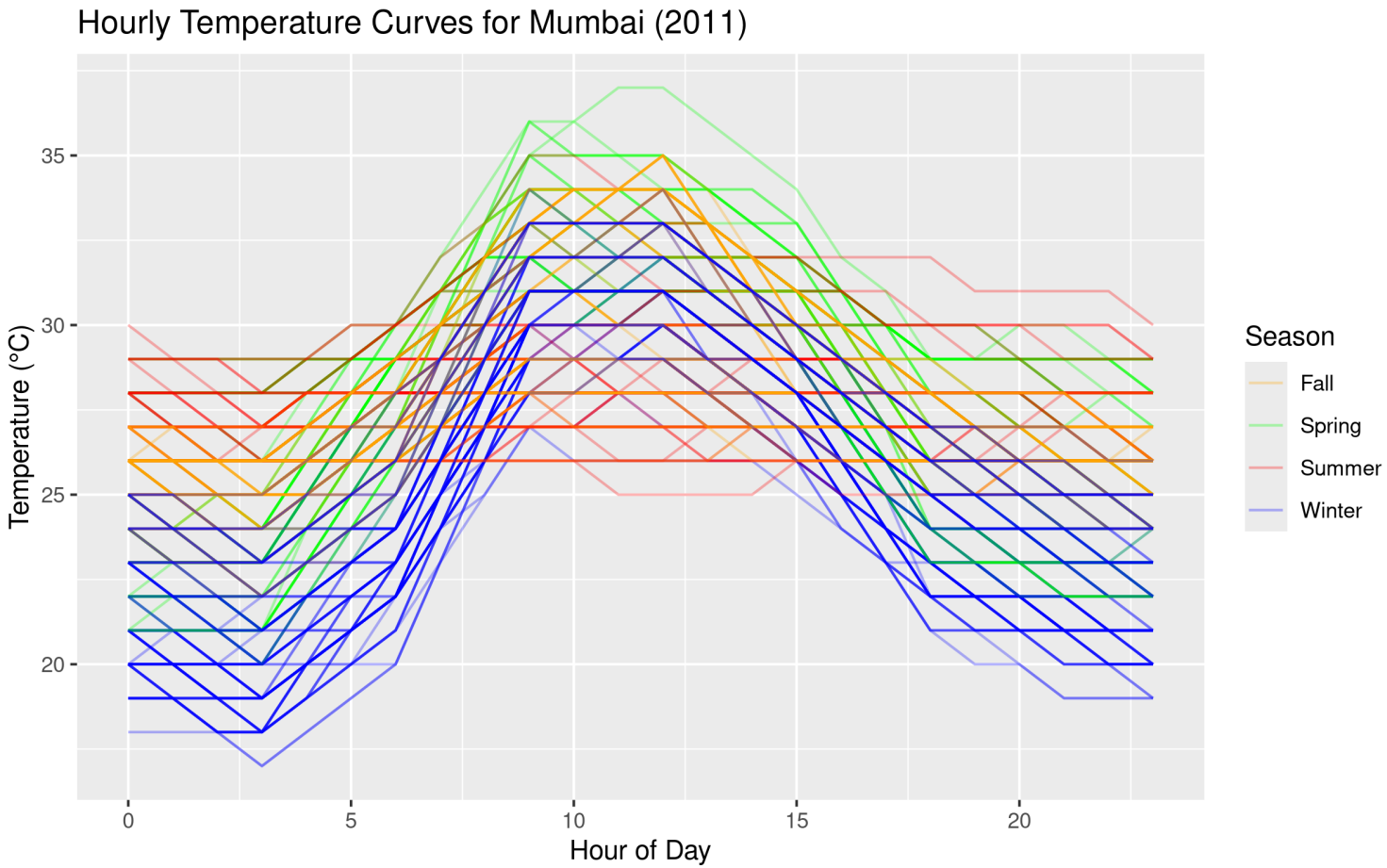
- **Historical Weather Data in India**
  - Hourly observations (2006–2019)
  - 8 major Indian cities
  - Over 20 meteorological variables (focus on Temperature)

Research Questions

- **Main goal:** Explore seasonal patterns in temperature data.
- **Questions:**
  - How do daily temperature curves differ by season?
  - Can yearly and intra-day temperature patterns be identified?
  - Are there consistent patterns in the derivatives of the curves?

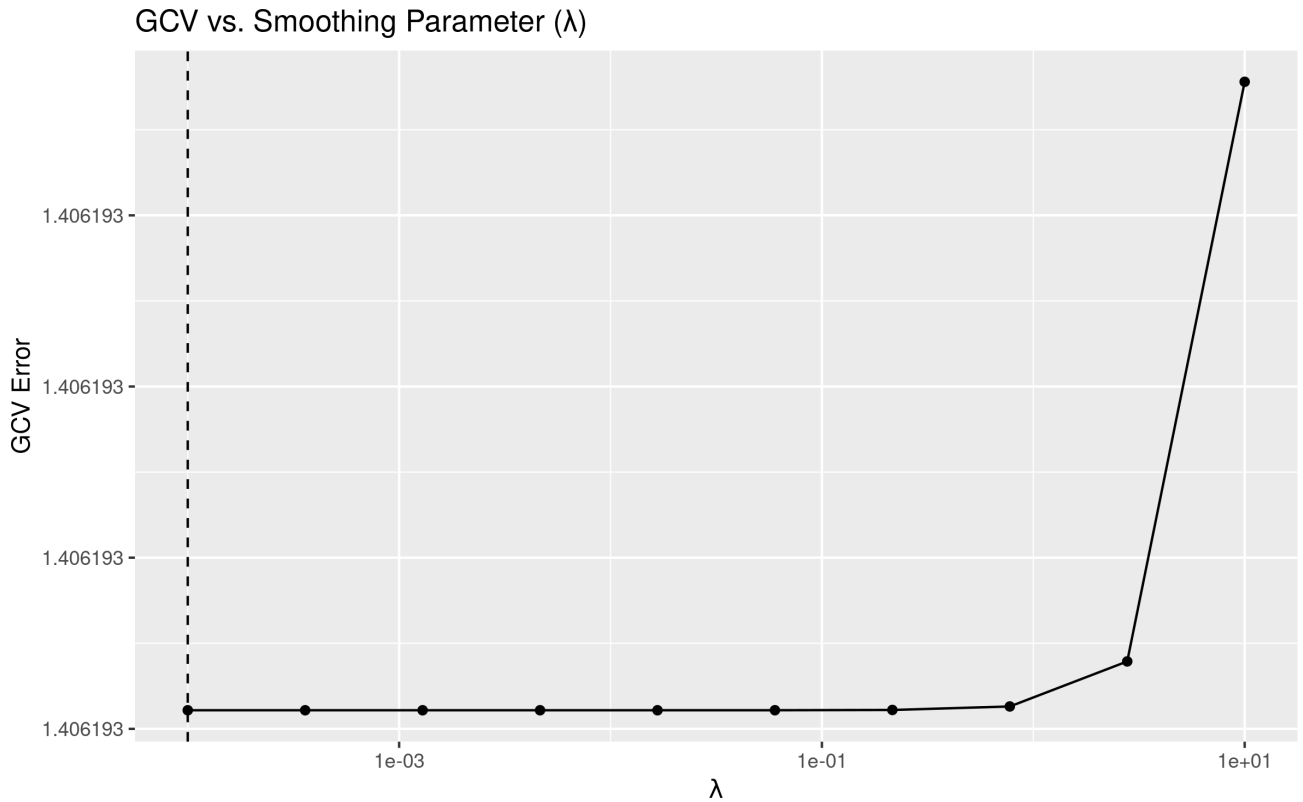
Unsmoothed Temperature Curves

- Daily temperature profiles from Mumbai in 2011.
- Lines colored by season (winter, spring, summer, fall).
- "Stepwise" appearance due to hourly observations.



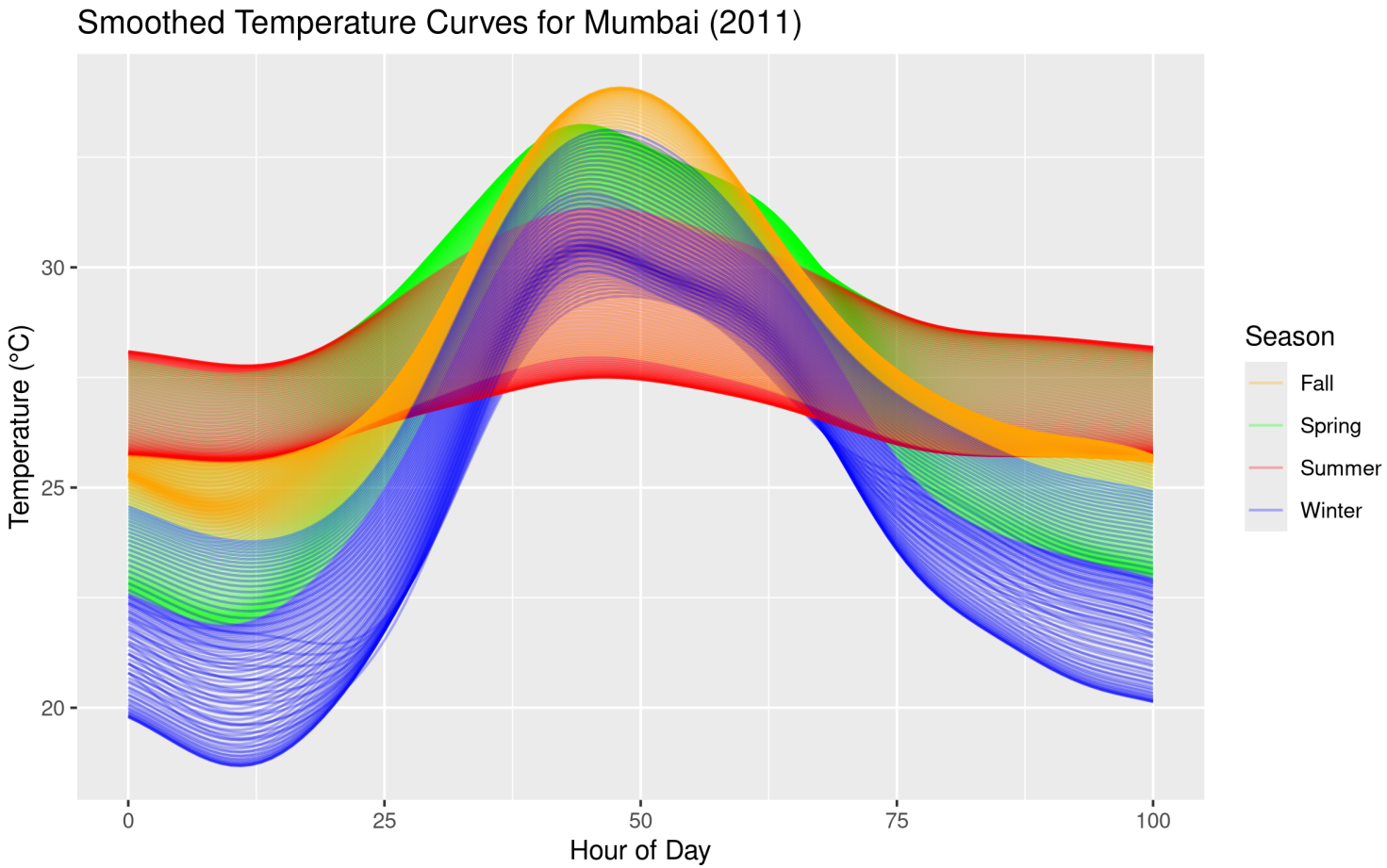
Lambda Selection via GCV

- Bivariate smoothing: B-splines in *day* (nbasis=12), Fourier in *hour* (harmonics=5)
- Search $\lambda \in [10^{-4}, 10^{1}]$ to minimize GCV
- Optimal λ_{opt} indicated by dashed line (selected $\lambda = \epsilon = 10^{-7}$)

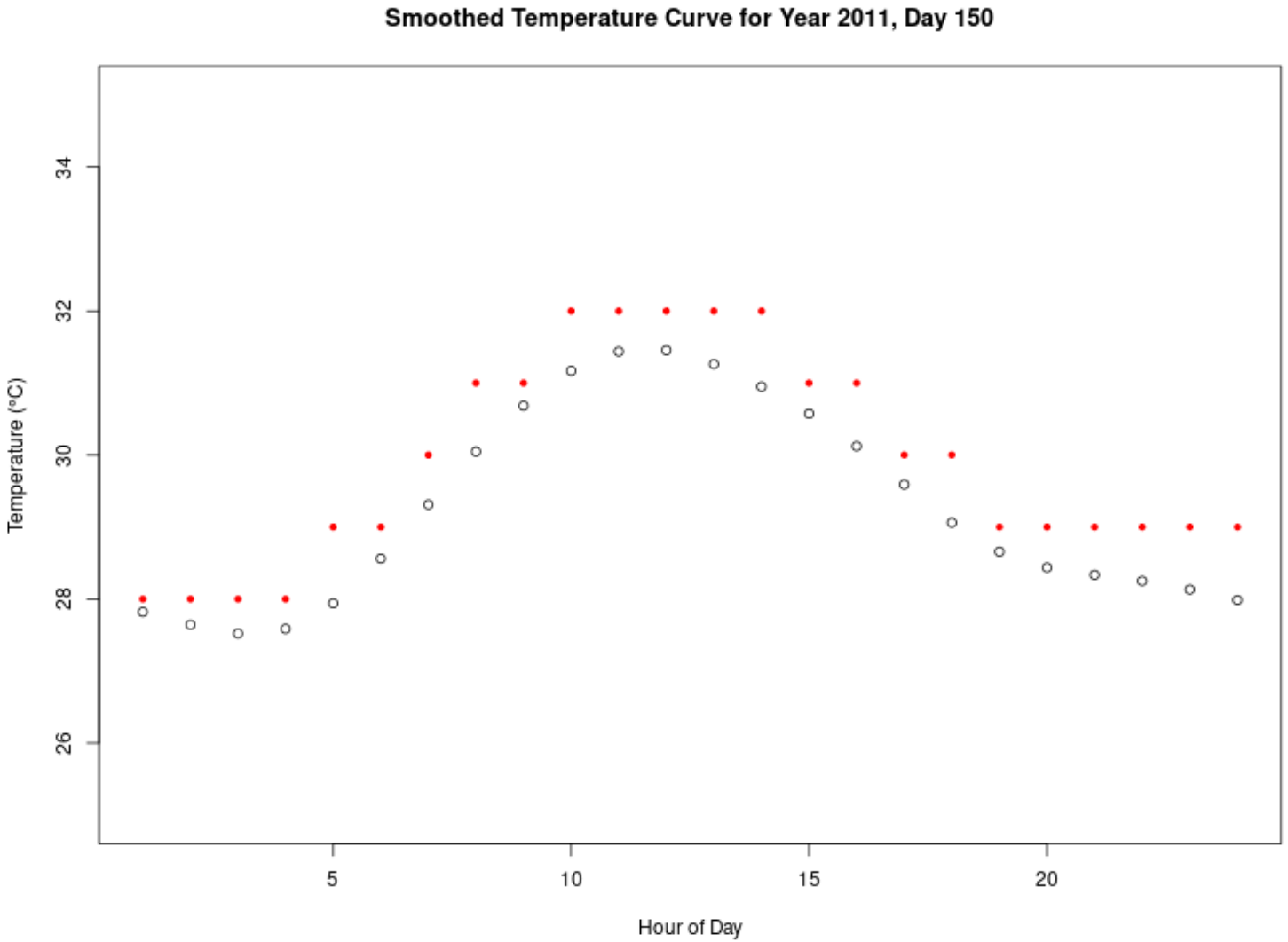


Smoothed Temperature Curves

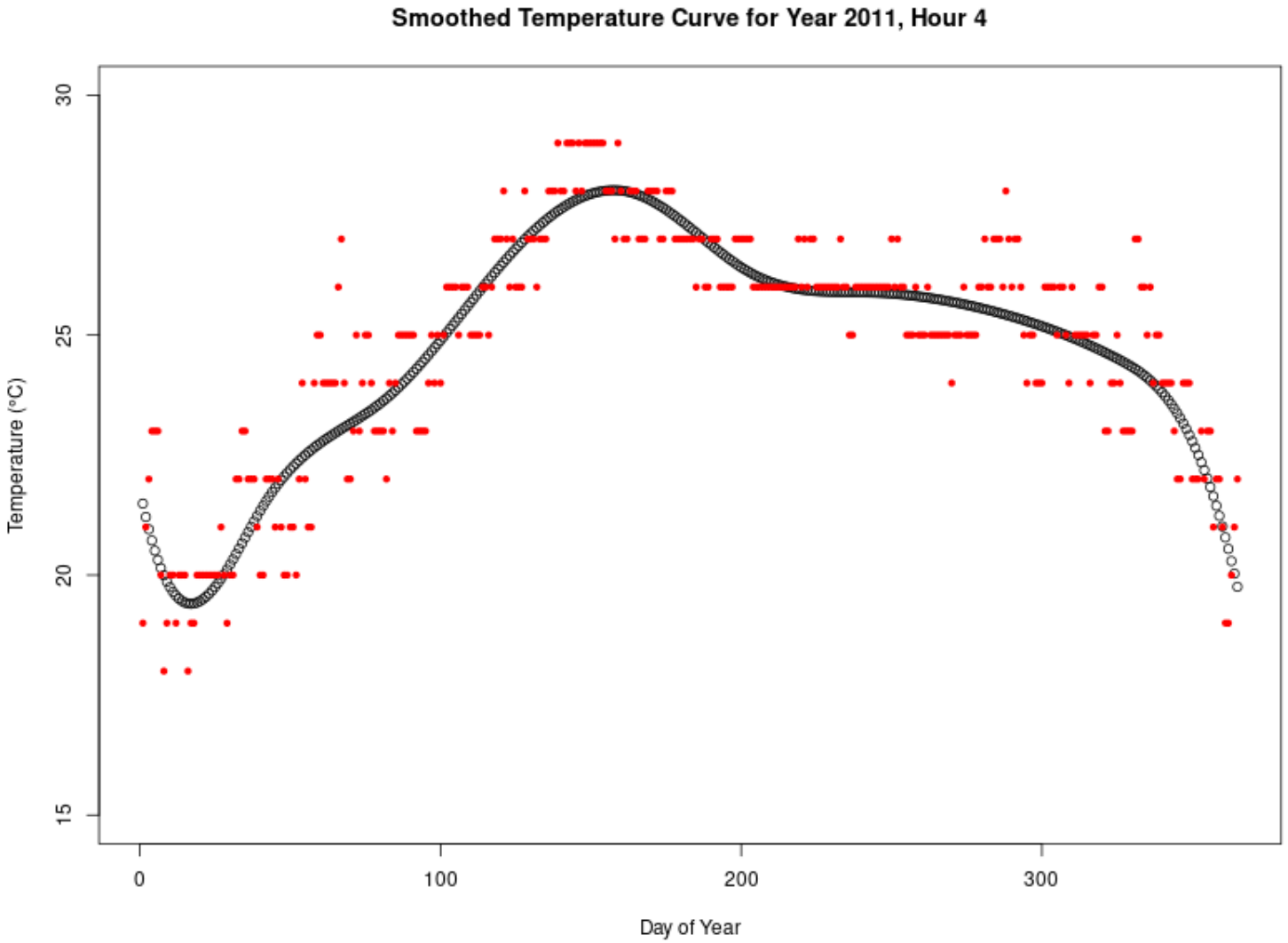
- Applied bivariate smoothing with λ_{opt}
- Reveals underlying daily patterns by season



Smoothed Temperature Curves *Day*



Smoothed Temperature Curves *Year*



First and Second Derivatives (Days)

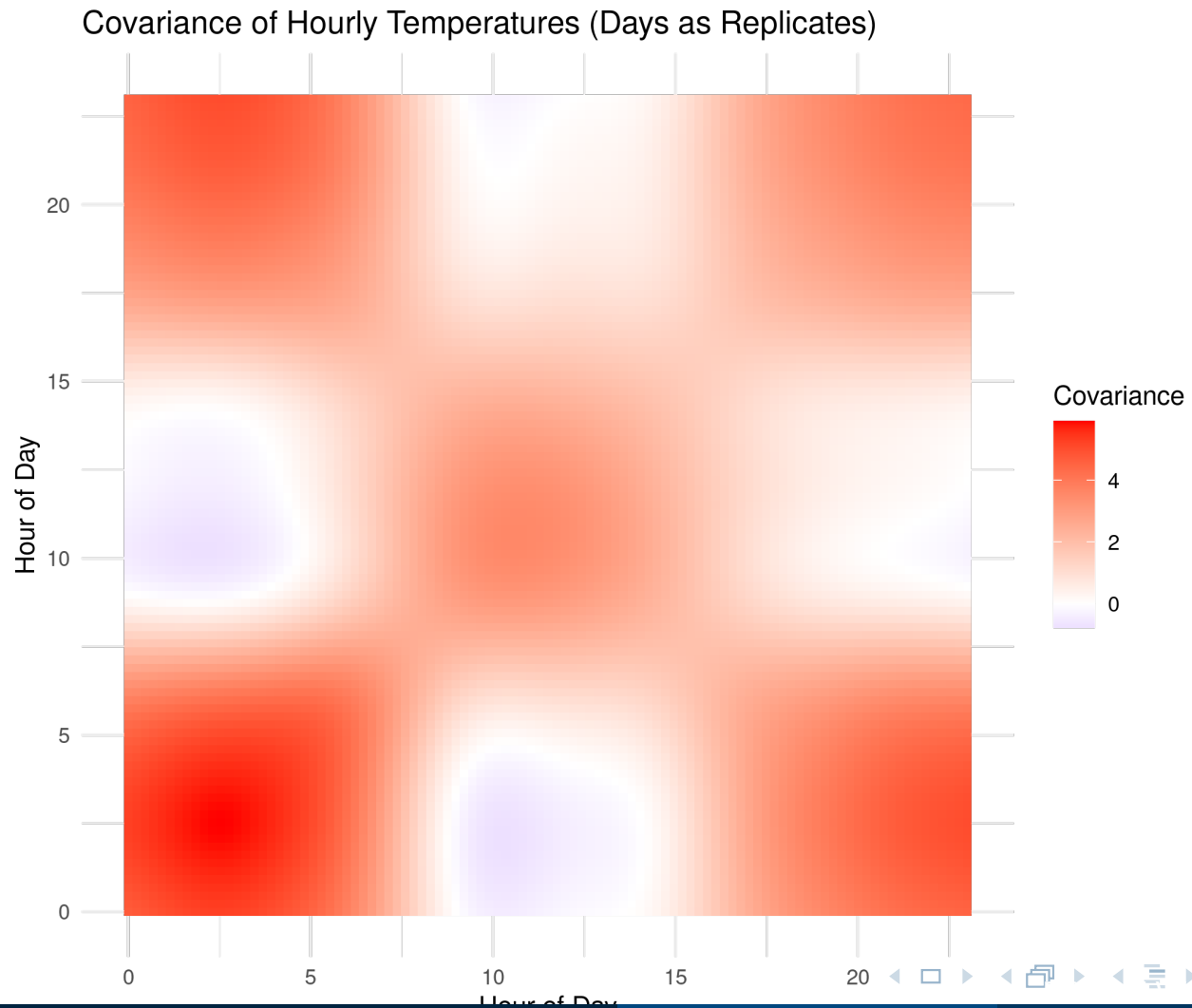
- First derivative: rate of change across the year
- Second derivative: acceleration highlights seasonal transitions

First and Second Derivatives (Hours)

- Intra-day temperature dynamics
- Captures typical morning/evening temperature dynamics

Covariance Heatmap of Hourly Temperatures

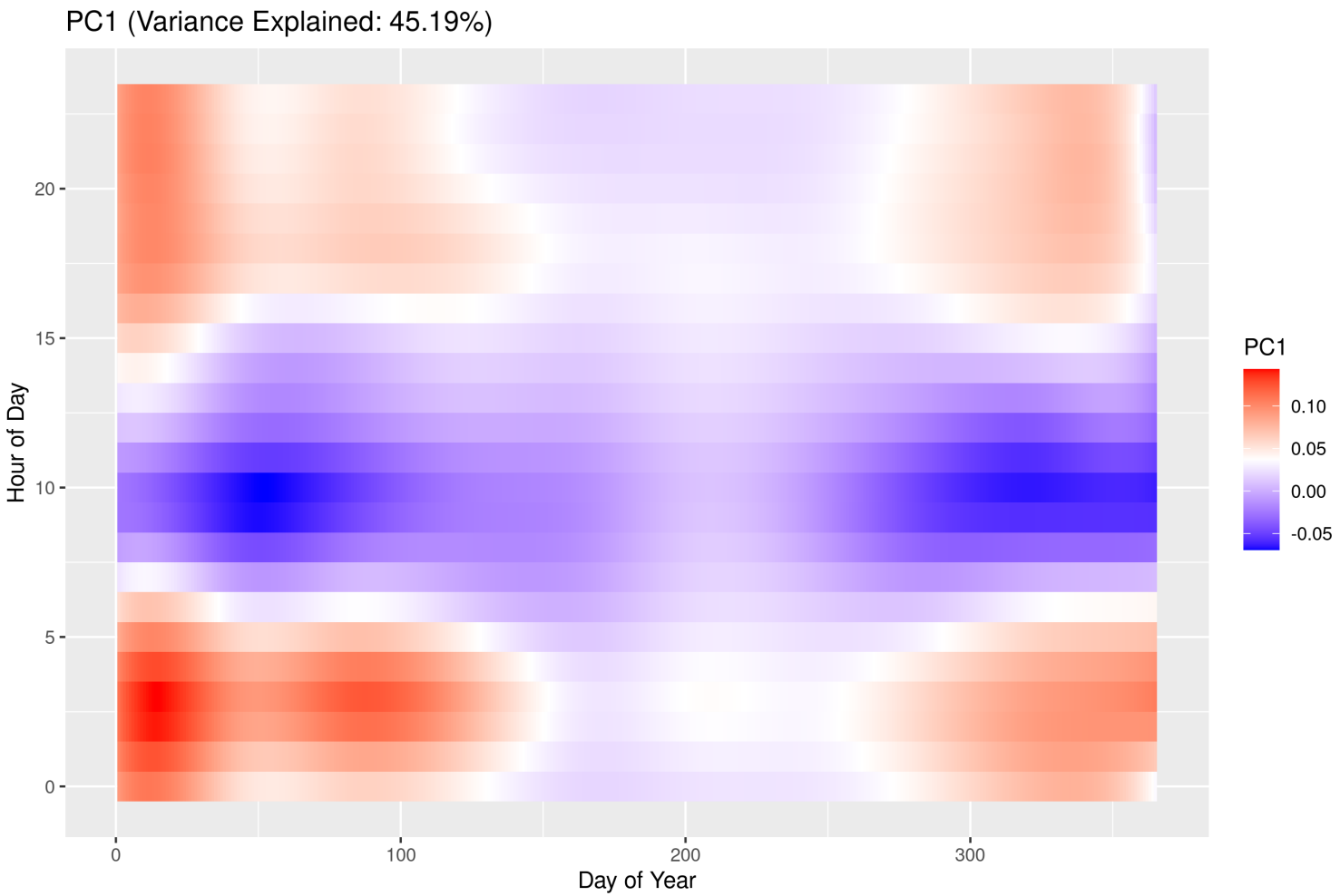
- Treat days as replicates over hours
- Covariance reveals which times of day co-vary



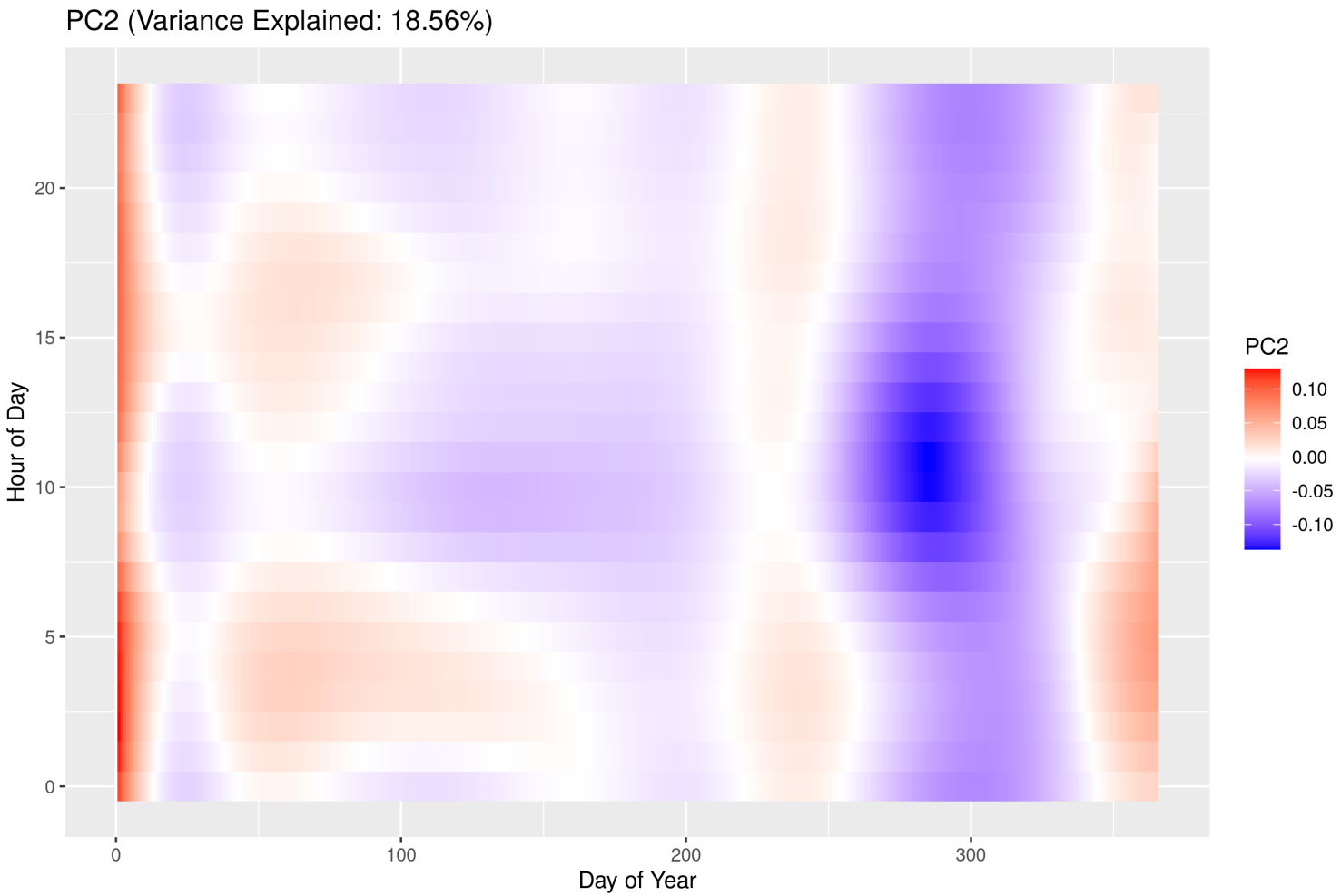
FPCA Results

- Principal components on smoothed bivariate functions
- PC1–PC3 explain over 81% of variance
- Heatmaps of PC surfaces expose key modes:

First Principal Component

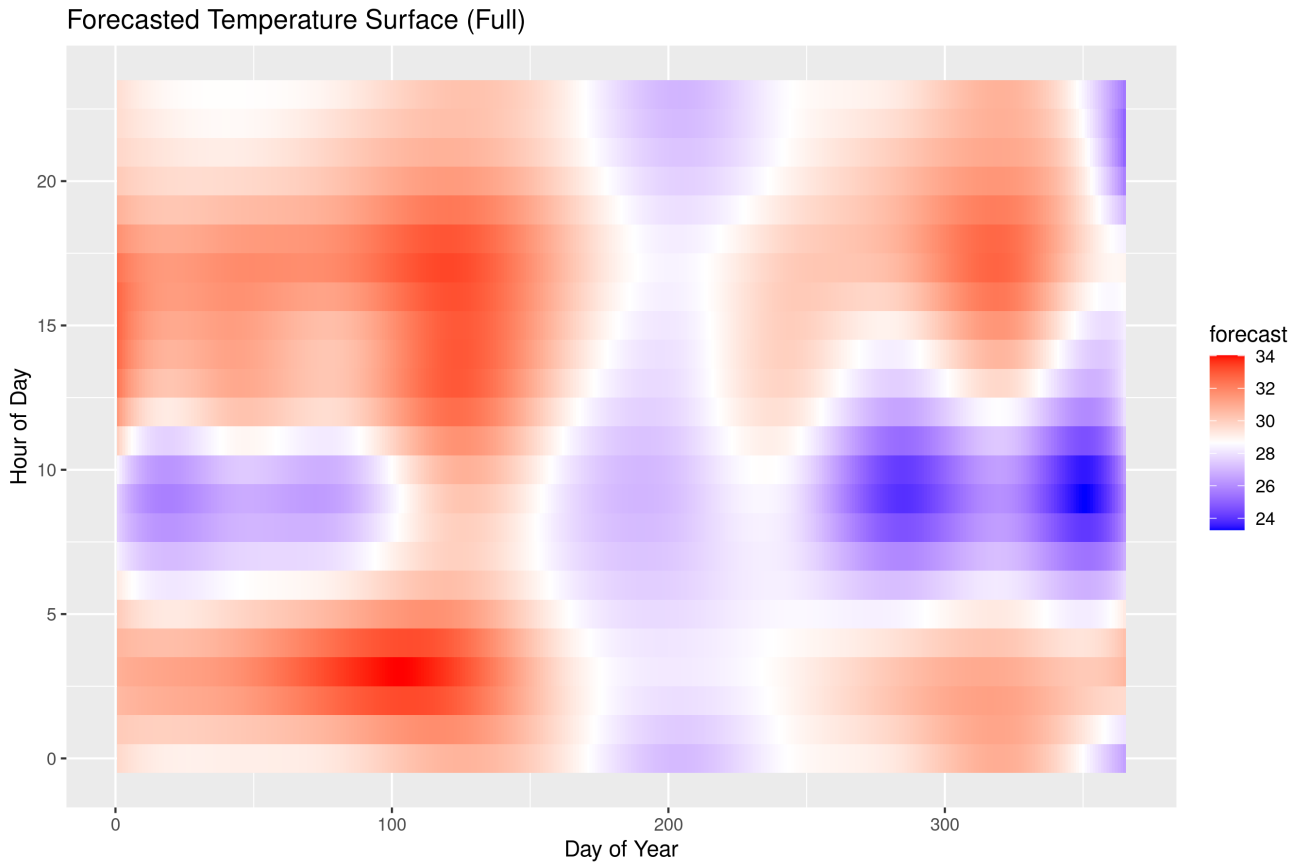


Second Principal Component



Annual Surface Forecast

- FAR model on FPCA scores (VAR(1) on PC1–PC3)
- Forecast the next year's temperature surface



Summary & Next Steps

- FDA effectively identifies seasonal patterns in weather data
- Even relatively 'rough' bases with capture a good part of variability
- Next steps:
  - Hypothesis testing

Thank You!

Thank you for your attention!