Functional Median Polish with Climate Applications

Ying Sun, SAMSI, sunwards@samsi.info Marc G. Genton, Texas A&M University, genton@stat.tamu.edu

Motivation

Scientific Questions: how do we compare sources of variability in observations or model outputs?

- Observations: a corroborating source of information about physical processes.
- Climate Models: numerically solve systems of differential equations representing physical relationships in the climate system.

Analysis of Variance

A two-way additive ANOVA model:

$$y_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij}.$$

- Fitted by arithmetic means (no outliers): $\hat{\mu} = \bar{y}$, $\hat{\alpha}_i = \bar{y}_{i\cdot} - \bar{y}$, $\hat{\beta}_j = \bar{y}_{\cdot j} - \bar{y}$.
- Fitted by medians (robust): median polish.

Median Polish

Original table (Tukey, 1970): find row medians.

1. 1st iteration: subtract row medians, find column medians.
2. 2nd iteration: subtract column medians, find row medians,

$$\begin{array}{ccc|c} 6 & 3 & 11 & 6 \\ 3 & 2 & 4 & 3 \\ 9 & 0 & 0 & 0 \end{array} \rightarrow \begin{array}{ccc|c} 0 & -3 & 5 & 6 \\ 0 & -1 & 1 & 3 \\ 9 & 0 & 0 & 0 \\ \hline 0 & -1 & 1 & 3 \end{array} \rightarrow$$

3. subtract new row medians, add their medians to grand median, find column medians.

$$\begin{array}{ccc|c||c} 0 & -2 & 4 & 0 & 3 \\ 0 & 0 & 0 & 0 & 0 \\ 9 & 1 & -1 & 1 & -3 \\ \hline 0 & -1 & 1 & 0 & 3 \end{array} \rightarrow \begin{array}{ccc||c} 0 & -2 & 4 & 3 \\ 0 & 0 & 0 & 0 \\ 8 & 0 & -2 & -3 \\ \hline 0 & 0 & 0 & 0 \\ 0 & -1 & 1 & 3 \end{array}$$

Polished table: new row and column medians are zero after two iterations.

$$\rightarrow \begin{array}{ccc||c} 0 & -2 & 4 & 3 \\ 0 & 0 & 0 & 0 \\ 8 & 0 & -2 & -3 \\ \hline 0 & -1 & 1 & 3 \end{array}$$

Functional Median Polish

Model: $y_{ijk}(x) = \mu(x) + \alpha_i(x) + \beta_j(x) + \epsilon_{ijk}(x)$.

- x can be time for curves or spatial index for surfaces/images.
- Iterative procedure sweeping out functional column and row medians.

Band Depth for Functional Data

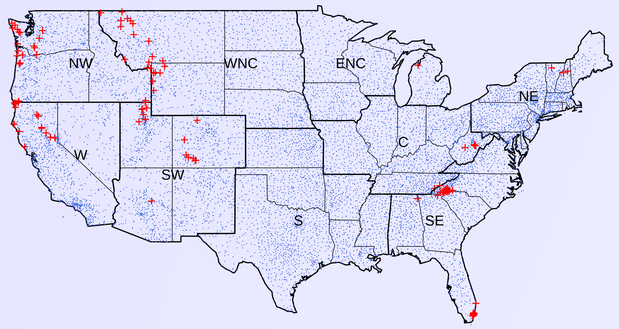
Functional median maximizes band depth (BD) or modified band depth (MBD):

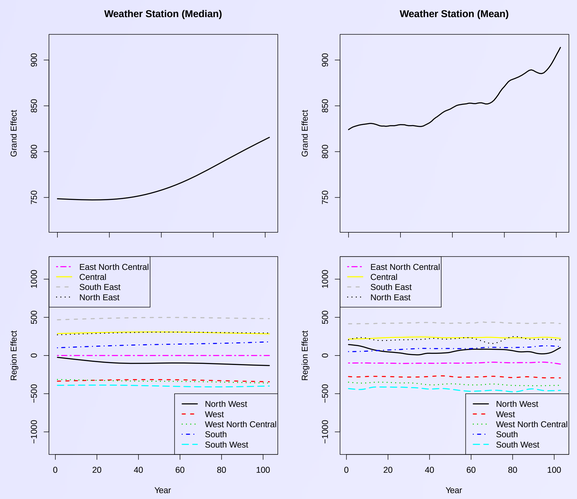
$$BD_n(y) = \binom{n}{2}^{-1} \sum_{1 \le i_1 < i_2 \le n} I\{G(y) \subseteq B(y_{i_1}, y_{i_2})\}.$$

$$MBD_n(y) = \binom{n}{2}^{-1} \sum_{1 \le i_1 < i_2 \le n} \lambda_r\{A(y; y_{i_1}, y_{i_2})\}.$$

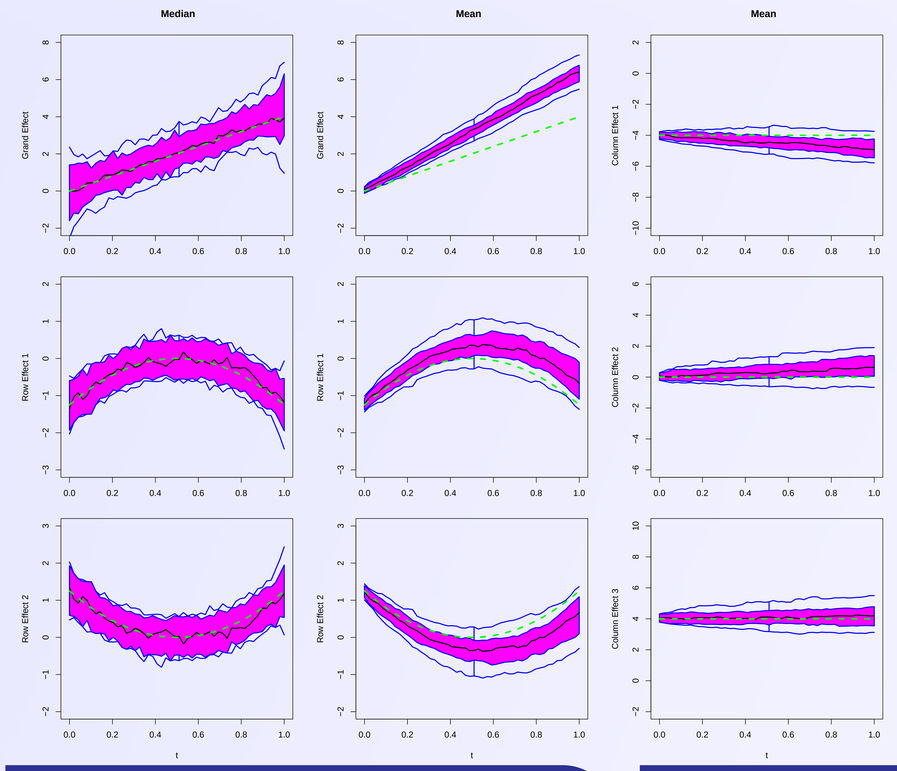
Climate Region Effect

Annual spatio-temporal precipitation: four areas of outliers detected by Sun & Genton (2011, 2012).



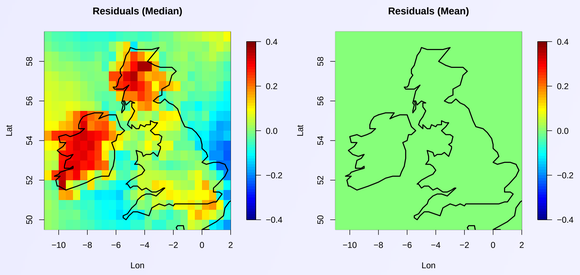


Functional Boxplots for Simulation Studies



UK and Ireland Temperatures

- Regional Climate Model (RCM): has higher resolution, covers a limited area of the globe.
- The boundaries of RCM are driven by variables output from a global climate model (GCM).
- Question: how much variability in the model output is from RCM and how much is due to the boundary conditions from GCM.
- Data: PRUDENCE project (Christensen, Carter, and Giorgi 2002), consists of control runs (1961-1990).
- Two factors: RCM (HIRHAM and RCAO), GCM (ECHAM4 and HadAm3H).



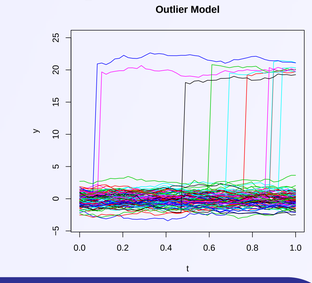
Outlier Model

Generate data from a true model with 100 curves in each cell of the 2×3 table at 50 time points.

$$\epsilon_{ijk}(t) = e_{ijk}(t) + c_{ijk} K I\{t \ge T_{ijk}\},$$

$$e_{ijk}(t) \sim GP(0, \gamma),$$
$$\gamma(t_1, t_2) = \exp\{-|t_2 - t_1|\}.$$

- $T_{ijk} \sim U(0,1)$ and c_{ijk} is 1 with probability q_{ij} and 0 with probability $1 - q_{ij}$, q_{ij} is different for each cell.
- Replication: 1,000.



GCM and RCM Effects