THE OHIO STATE UNIVERSITY

Variograms for Kriging and Clustering of Spatial Functional Data with Phase Variation

Sebastian Kurtek

Department of Statistics

The Ohio State University

This is joint work with my former Ph.D. student, Dr. Xiaohan Guo, and Prof. Karthik Bharath, School of Mathematical Sciences, University of Nottingham.

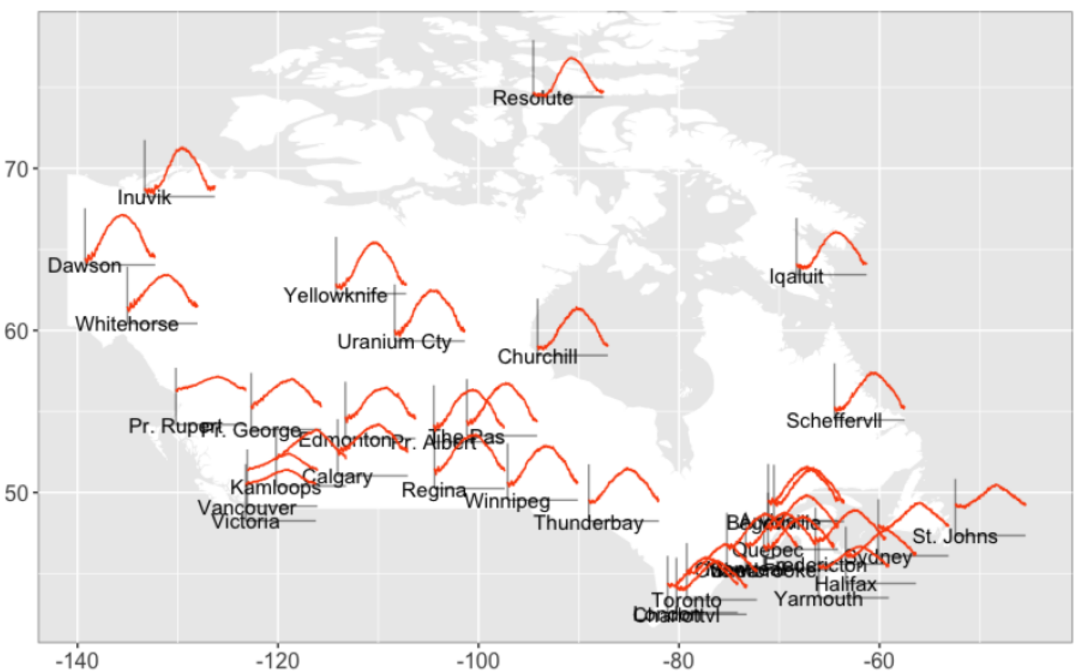
We thank the NSF, NIH and EPSRC for supporting this research.

MOTIVATION

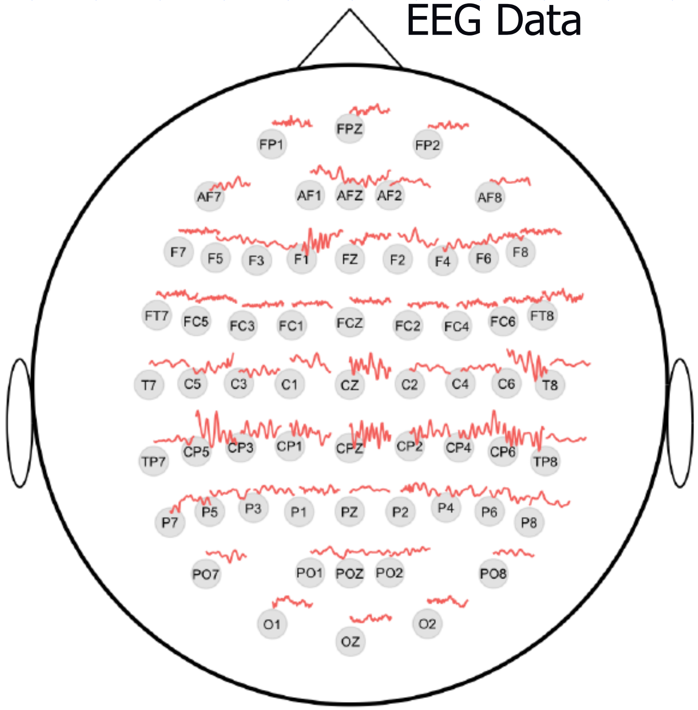
THE OHIO STATE UNIVERSITY

- Spatial functional data: functional data with complex spatial dependencies.
- Applications: environmental science, medicine, biology, geology, econometrics, etc.
- Two important spatial functional data analysis tasks: kriging (spatial prediction) and spatial clustering.
- Useful in other tasks: spatially penalized registration of multivariate functional data.

Canadian Weather Data

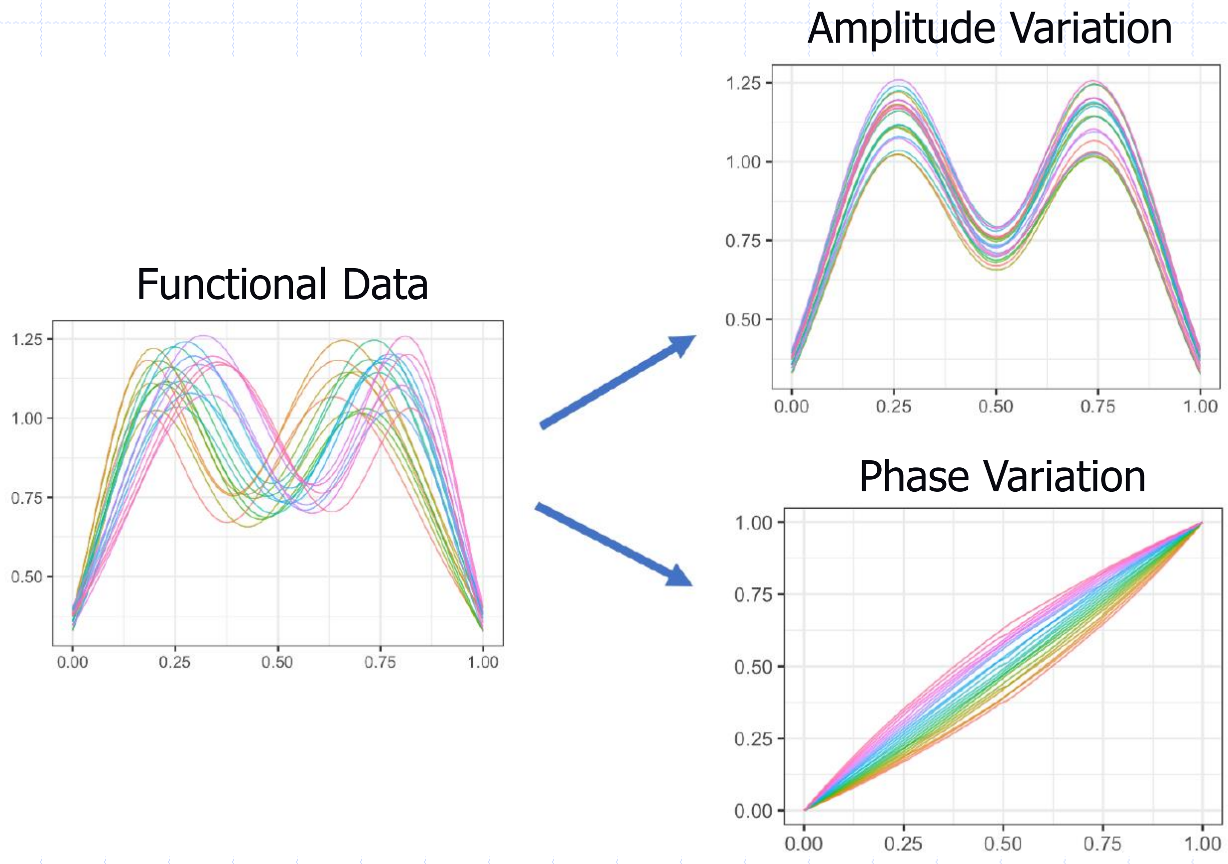


EEG Data



MOTIVATION

- Amplitude (y-axis) variation: shape and scale differences among functional data.
- Phase (x-axis) variation: timing differences in amplitude features, e.g., local extrema.
- Spatial, amplitude and phase variations in spatial functional data are confounded.



PREVIOUS WORK AND CHALLENGES

- Let $\{f_s,\ s \in \mathcal{D}\} \subseteq \mathbb{L}^2$ denote a second-order stationary and isotropic functional random field on a spatial domain $\mathcal{D} \subset \mathbb{R}^d$.
- Trace-variogram (Giraldo et. al, 2011): popular approach for modeling spatial dependence in functional data. $\|\cdot\|$ is the $\mathbb{L}^2$ norm and $|\cdot|$ is the Euclidean norm in $\mathbb{R}^d$.

$$V(h) = \frac{1}{2}E(\|f_s - f_{s'}\|^2), \quad h = |s - s'|$$

- Implicit assumption: temporal correspondence across spatial functional observations is fixed.
  - Functions are perfectly aligned.
  - Phase variation is treated as negligible noise.
  - Phase variation may impact stationarity assumption.

The Ohio State University

PRELIMINARIES

- Function space: $\mathcal{F} = \{f : [0,1] \to \mathbb{R} \mid f \text{ is absolutely continuous}\}$
- Set of warping functions (phase):

$$\Gamma = \{\gamma : [0,1] \to [0,1] \mid \gamma(0) = 0,\ \gamma(1) = 1,\ 0 < \dot{\gamma} < \infty\}$$

- Function warping: $f \circ \gamma$
- Need a warping-invariant metric for statistical analysis. $\mathbb{L}^2$ is not invariant.
- Extended Fisher-Rao metric: origins in statistics for analyzing probability density functions.
- Invariance to warping (Cencov, 1982): $d_{FR}(f_1, f_2) = d_{FR}(f_1 \circ \gamma, f_2 \circ \gamma)$
- Limitation: difficult to use in practice.

THE OHIO STATE UNIVERSITY

PRELIMINARIES

- Square-Root Slope Function (SRSF): $Q : \mathcal{F} \to \mathbb{L}^2([0,1], \mathbb{R})$

$$Q(f) = q = \text{sign}(\dot{f})\sqrt{|\dot{f}|}$$

- Inverse mapping: $Q^{-1}(q, f(0))(t) = f(0) + \int_0^t q(s)|q(s)|ds$
- Extended Fisher-Rao metric becomes the $\mathbb{L}^2$ metric: $d_{FR}(f_1, f_2) = \|q_1 - q_2\|$
- Function warping through SRSF: $q \odot \gamma = (q \circ \gamma)\sqrt{\dot{\gamma}}$
- Pairwise alignment: $\hat{\gamma} = \arg\min_{\gamma \in \Gamma} \|q_1 - q_2 \odot \gamma\|$
- Amplitude distance: $d_a(q_1, q_2) = \|q_1 - q_2 \odot \hat{\gamma}\|$
- Let $\psi = Q(\gamma) = \sqrt{\dot{\gamma}}$.
- Relative (extrinsic) phase distance: $d_p(q_1, q_2) = \|Q(\hat{\gamma}) - Q(\gamma_{id})\| = \|\hat{\psi} - 1\|$

AMPLITUDE AND PHASE VARIOGRAMS

- Decomposition of transformed functional random field $\{q_s, \ s \in \mathcal{D}\}$:
  - Amplitude random field: $\{q_s \odot \gamma_s, \ s \in \mathcal{D}\}$
  - Phase random field: $\{\psi_s = Q(\gamma_s), \ s \in \mathcal{D}\}$
- Amplitude trace-variogram: $V_a(h) = \frac{1}{2}E(\|q_s \odot \gamma_s - q_{s'} \odot \gamma_{s'}\|^2), \quad h = |s - s'|$
- Elements of the phase random field are only relative and depend on shapes of functions in the random field $\{q_s\}$.
- Let $[\bar{q}] = \{(q \odot \gamma)/\|q\| \mid \gamma \in \Gamma\}$ represent the shape of q, and define the shape random field as $\mathcal{S} = \{[\bar{q}_s], \ s \in \mathcal{D}\}$.
- Conditional phase trace-variogram (Schmidt et al. 2011):

$$V_p(h_\omega) = \frac{1}{2}E(\|\psi_s - \psi_{s'}\| \mid \mathcal{S})$$

$$h_\omega\left((s, [\bar{q}_s]), (s', [\bar{q}_{s'}])\right) = \sqrt{|s - s'|^2 + \omega d_a(\bar{q}_s, \bar{q}_{s'})^2}, \quad \omega \geq 0$$

SPATIAL HIERARCHICAL CLUSTERING

- Spatial functional data: $q_i, \ s_i \in \mathcal{D}, \ i = 1, \ldots, n$
- Empirical amplitude and phase trace-variograms:

$$\hat{V}_a(h) = \frac{1}{2|N_a(h)|} \sum_{i,j \in N_a(h)} d_a(q_i, q_j)^2 \qquad N_a(h) = \{(s_i, s_j) \mid |s_i - s_j| \in (h - \epsilon, h + \epsilon)\}$$

$$\hat{V}_p(h_\omega) = \frac{1}{2|N_p(h_\omega)|} \sum_{i,j \in N_p(h_\omega)} d_p(q_i, q_j)^2$$

$$N_p(h_\omega) = \{(s_i, [\bar{q}_i]), (s_j, [\bar{q}_j]) \mid h_\omega((s_i, [\bar{q}_i]), (s_j, [\bar{q}_j])) \in (h_\omega - \epsilon, h_\omega + \epsilon)\}$$

- We use Matern variogram model fits to the empirical variograms, estimated using ordinary least squares.
- Weighted distance matrices for hierarchical clustering:

$$d_{A,ij} = d_a(q_i, q_j) V_a(h_{ij}), \quad d_{P,ij} = d_p(q_i, q_j) V_p(h_{\omega,ij}), \quad i, j = 1, \ldots, n$$

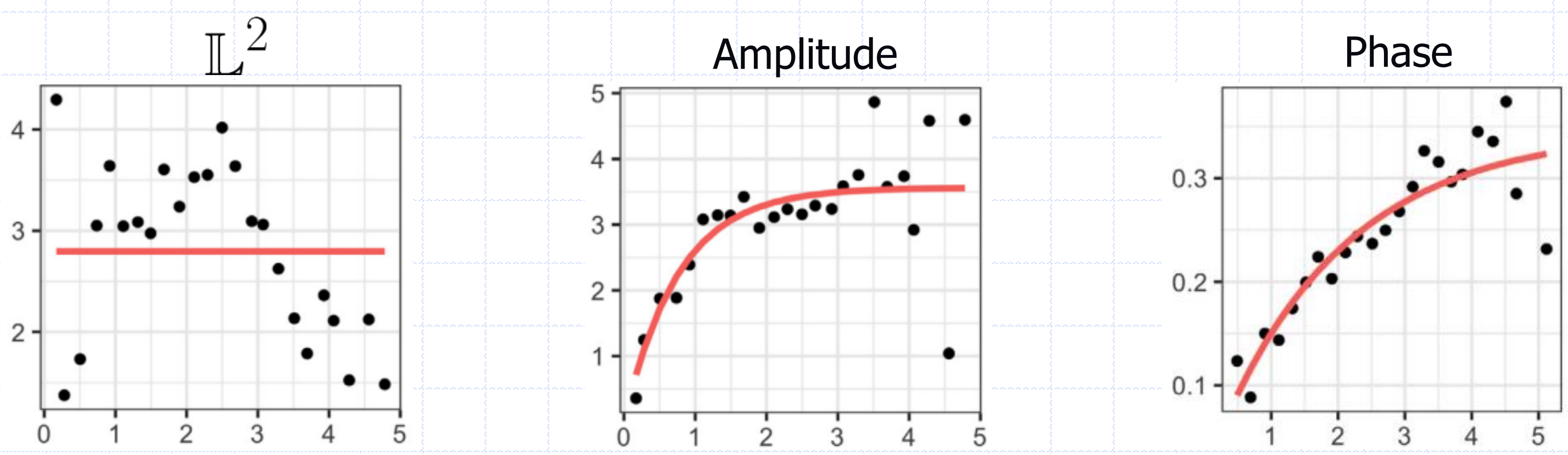
- Tuning parameter ω is chosen to minimize squared error of Matern fit to empirical estimate.

AMPLITUDE AND PHASE VARIOGRAMS

- Decomposition of the $\mathbb{L}^2$ trace-variogram into amplitude and phase components.

- The functional data was simulated such that the amplitude and phase are correlated in space.

- The $\mathbb{L}^2$ variogram suggests a quadratic pattern for spatial dependency, which is the truth for neither amplitude or phase. The resulting fitted Matern variogram model fails to capture the spatial correlation in the data.



THE OHIO STATE UNIVERSITY

SIMULATION

number of clusters

number of observations per cluster

- Simulated data:

$$f_{ij}(t) = \{(a_{ij}\mu + e_{ij}) \circ \gamma_{ij}\}(t), \; t \in [0,1] \qquad i = 1, \ldots, I; j = 1, \ldots, n_i$$

- Deterministic mean function: $\mu(t) = -\cos(2\pi t)$

- Error process: e_{ij} is a zero-mean Gaussian process with diagonal covariance

- Spatial dependence in amplitude: $a_{ij} = i\delta_a + \epsilon_{a,ij}$

controls magnitude of amplitude cluster differences

spatial dependence in amplitude

- Spatial dependence in phase: γ_{ij} is the cumulative distribution function of $\mathrm{Beta}(1, e^{b_{ij}})$, where $b_{ij} = i\delta_b + \epsilon_{b,ij}$.

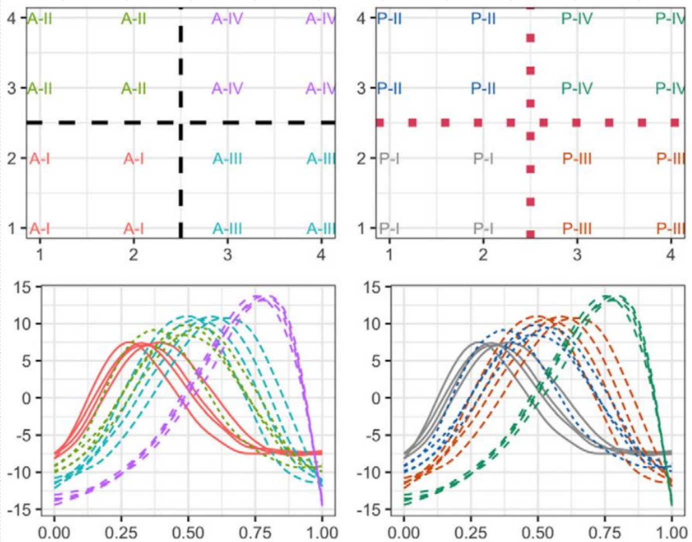
controls magnitude of phase cluster differences

spatial dependence in phase

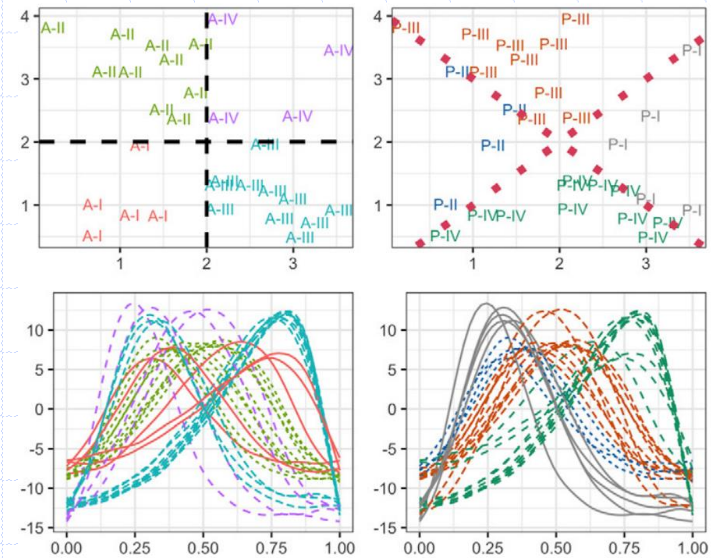
The Ohio State University

SIMULATION

Agree (regular grid)



Disagree (uniformly sampled)



Clustering evaluation criterion: average (standard deviation) of rand index over 100 replications.

δ_a	δ_b	Method	Agree		Disagree	
			Amplitude	Phase	Amplitude	Phase
1	0.1	APC	**0.828 (0.106)**	0.751 (0.101)	**0.808 (0.107)**	**0.711 (0.088)**
		L^2C	0.772 (0.079)	**0.772 (0.079)**	0.731 (0.085)	**0.711 (0.067)**
	0.5	APC	0.870 (0.091)	**0.958 (0.052)**	**0.752 (0.086)**	**0.887 (0.083)**
		L^2C	**0.910 (0.072)**	0.910 (0.072)	0.701 (0.046)	0.877 (0.07)
2	0.1	APC	**0.945 (0.067)**	0.767 (0.104)	**0.916 (0.076)**	0.708 (0.087)
		L^2C	0.808 (0.082)	**0.808 (0.082)**	0.779 (0.085)	**0.742 (0.068)**
	0.5	APC	**0.949 (0.071)**	**0.955 (0.059)**	**0.835 (0.085)**	**0.879 (0.087)**
		L^2C	0.908 (0.080)	0.908 (0.080)	0.707 (0.055)	0.873 (0.071)

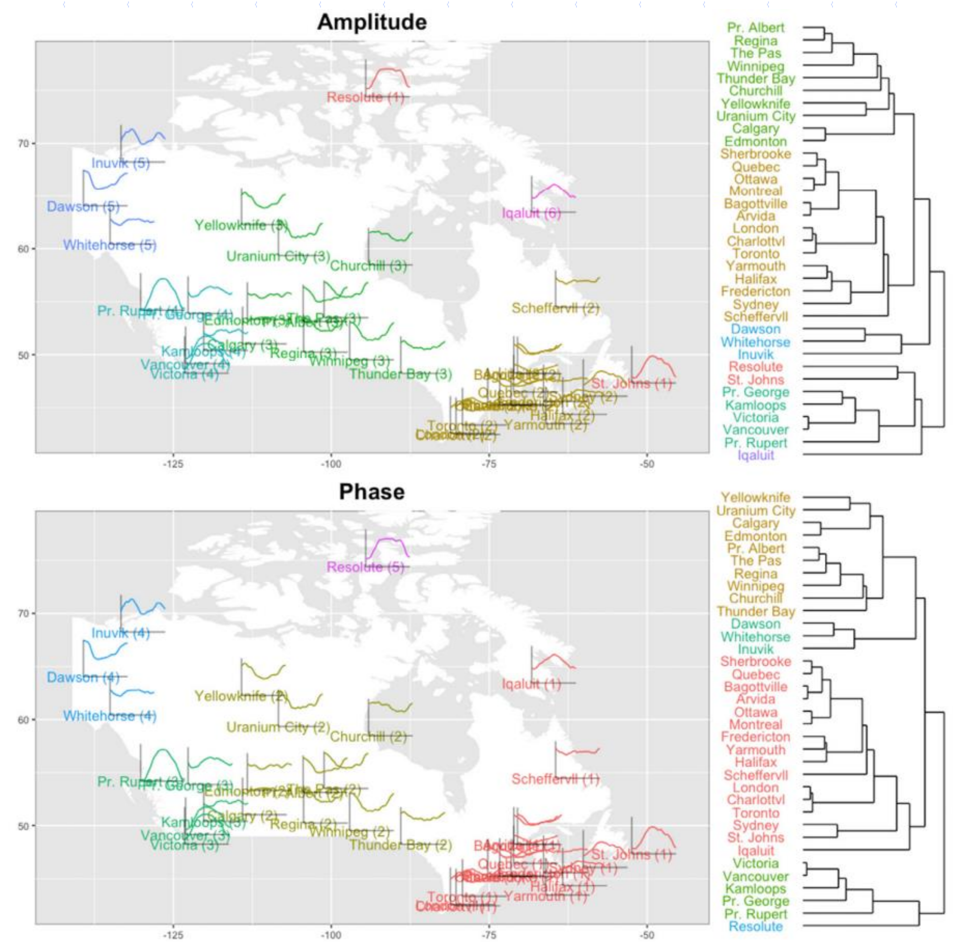
THE OHIO STATE UNIVERSITY

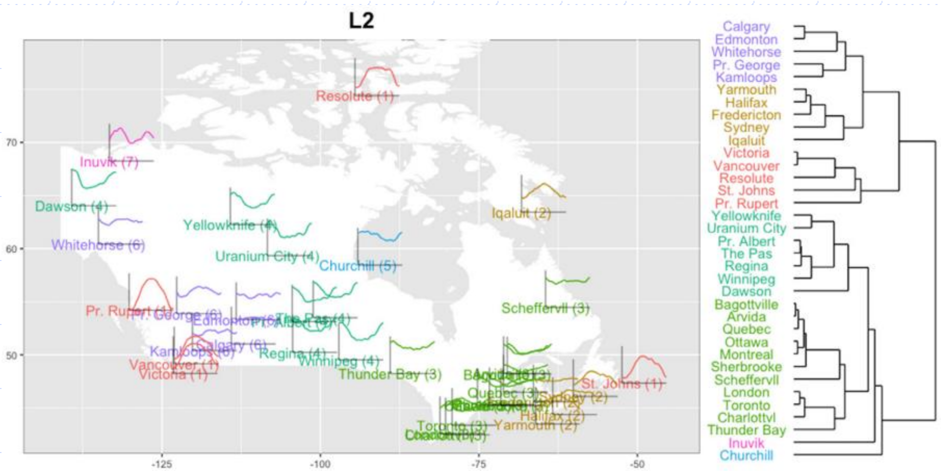
CANADIAN WEATHER DATA

- Daily temperature data averaged over 1960-1994 collected at 35 stations in Canada.
- The 35 stations cover a large area: we first filter out longitudinal and latitudinal trends via functional linear regression.
- Functional residuals are smoothed using splines (with low smoothing parameter) and used as data for clustering.
- We use hierarchical clustering with average linkage.
- The number of clusters is determined by minimizing the average silhouette (Rousseeuw, 1987), which quantifies the difference in similarity of an object to its own cluster versus other clusters.

CANADIAN WEATHER DATA

Comparison of spatial amplitude-phase clustering to spatial $\mathbb{L}^2$ - based clustering.





The Ohio State University

KRIGING

- Empirical amplitude and phase trace-variograms:

$$\boxed{\hat{V}_a(h) = \frac{1}{2|N_a(h)|} \sum_{i,j \in N_a(h)} \|q_i \odot \hat{\gamma}_i - q_j \odot \hat{\gamma}_j\|^2}$$

$$\boxed{\hat{V}_p(h_\omega) = \frac{1}{2|N_p(h_\omega)|} \sum_{i,j \in N_p(h_\omega)} \|\hat{\psi}_i - \hat{\psi}_j\|^2}$$

- Goal: predict function f_0 at spatial location s_0 via separate predictions of its amplitude $\tilde{q}_0$ and phase $\tilde{\gamma}_0$ (and translation).
- Amplitude/phase predictions: weights minimize the expected amplitude/phase prediction errors; estimation requires trace-variogram estimates only.

$$\tilde{q}_0 = \sum_{i=1}^{n} \eta_i \left[q_i \odot \hat{\gamma}_i\right], \ \sum_{i=1}^{n} \eta_i = 1 \qquad \tilde{\psi}_0 = \sum_{i=1}^{n} \xi_i \hat{\psi}_i, \ \xi_i > 0, \ \sum_{i=1}^{n} \xi_i = 1$$

KRIGING

- Issue: the phase components in the data are not observed and must be estimated via alignment to an appropriate template.

- Solution: spatially-weighted template that serves the dual role of the amplitude kriging estimator and a local template for alignment.

Algorithm overview:

1. Template initialization.
2. Alignment to current estimate of template.
3. Estimation of trace-variogram using aligned functions.
4. Template update via amplitude kriging using estimated trace-variogram and aligned functions.

- We use the function starting points for kriging of the translation component.

- The combined kriging prediction for the function f_0 utilizes the separate kriging predictions of amplitude, phase and translation:

$$Q^{-1}(\tilde{q}_0 \odot \tilde{\gamma}_0^{-1}, \tilde{T}_0)$$

THE OHIO STATE UNIVERSITY

SIMULATION

Simulated data: $f_i(\gamma_i(t)) = \sum_{j=1}^{K} a_{ij}\phi_j(t) + e_i(t), \ t \in [a,b]$

Different simulation scenarios:

1. Bimodal: single basis function ($\phi_1(t) = -\cos(2\pi t), \ t \in [-1,1]$); amplitude spatial correlation induced via basis coefficient; spatially correlated phase simulated in same manner as for clustering; spatial phase correlation independent of function shapes.
2. B-spline: 10 cubic B-spline basis functions on [0,1]; amplitude spatial correlation induced via coefficients; spatially correlated phase simulated in same manner as for clustering.
   a. Spatial phase correlation independent of function shapes.
   b. Spatial phase correlation dependent on function shapes.

Kriging methods:

1. APK: amplitude-phase kriging
2. TSK: align to estimated amplitude mean then ordinary kriging
3. OK: ordinary kriging, no alignment (Giraldo et al., 2011)
4. UK: universal kriging, no alignment (Menafoglio et al., 2013)

Evaluation metrics:

1. Amplitude MSE (AMSE) via amplitude distance between true function and prediction
2. Phase MSE (PMSE) via relative phase distance between true function and prediction

SIMULATION

THE OHIO STATE UNIVERSITY

(PMSE multiplied by 10)

1. Bimodal: amplitude and phase independent

Error	APK	TSK	OK	UK
AMSE	**0.81 (0.24)**	0.82 (0.23)	4.24 (1.93)	3.72 (1.14)
PMSE	0.55 (0.17)	**0.54 (0.15)**	0.88 (0.37)	0.83 (0.24)

2.a. B-spline: amplitude and phase independent

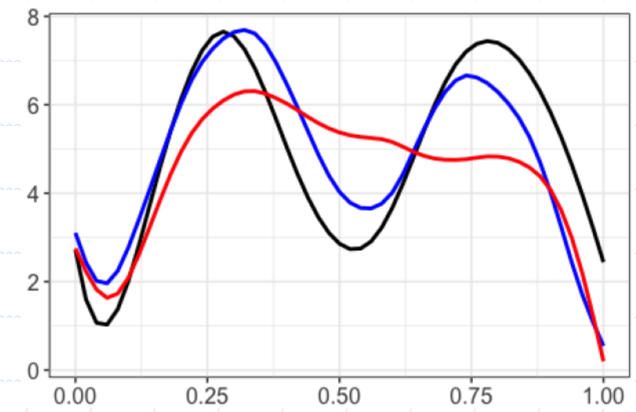
Error	APK	TSK	OK	UK
AMSE	**2.52 (0.53)**	2.61 (0.55)	3.12 (0.68)	3.17 (0.67)
PMSE	**1.23 (0.28)**	1.39 (0.34)	1.40 (0.22)	1.43 (0.26)

2.b. B-spline: spatial phase correlation dependent on function shapes

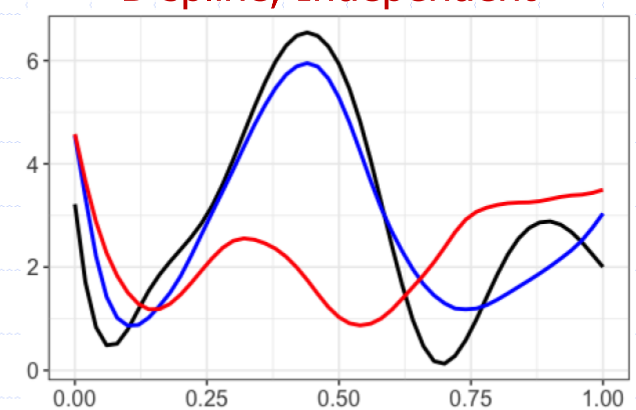
Error	APK	TSK	OK	UK
AMSE	**2.55 (0.49)**	2.66 (0.52)	3.64 (1.10)	3.46 (0.78)
PMSE	**1.34 (0.27)**	1.44 (0.34)	1.66 (0.36)	1.61 (0.32)

True Function
Amplitude-phase Kriging
Ordinary Kriging

Bimodal

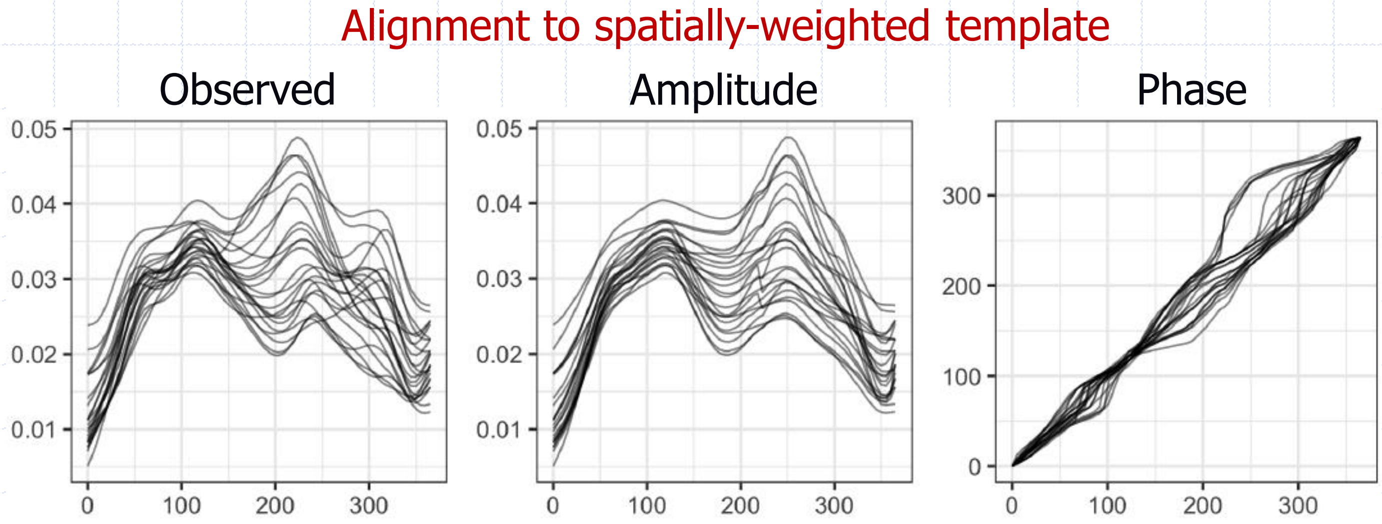


B-spline, Independent



OZONE CONCENTRATION

Data: daily ozone concentration functions from 24 stations in northern California in 2018.



Leave-one-out cross-validation average prediction errors (all values multiplied by 1000):

	APK	TSK	OK
AMSE	**1.94**	2.21	2.23
PMSE	49.02	**45.90**	53.64

THE OHIO STATE UNIVERSITY

DISCUSSION AND FUTURE WORK

Discussion:

- Difficult to verify assumptions of stationarity and isotropy for spatial functional data, especially after amplitude-phase separation.
- Need for further study of the conditional phase-trace variogram: functional random field over an infinite-dimensional domain.

Future work:

- Extension to universal amplitude-phase kriging.
- Setting of noisy and/or sparse spatial functional data.
- Extensions to other, more complex, spatial functional data, e.g., shapes of curves and surfaces.
- Marked point processes with functional or shape marks.

THE OHIO STATE UNIVERSITY

THANK YOU!