Random Fields on the Sphere of Planet Earth

Emilio Porcu

UTFSM at Valparaiso

Nitteroi, November 2016

Valparaiso

Valparaiso

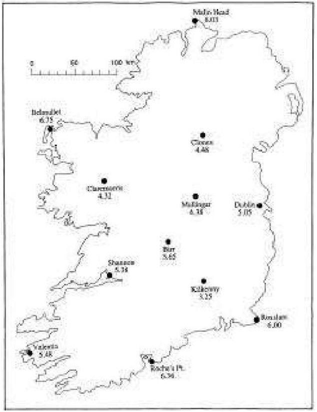
Valparaiso

UTFSM

UTFSM

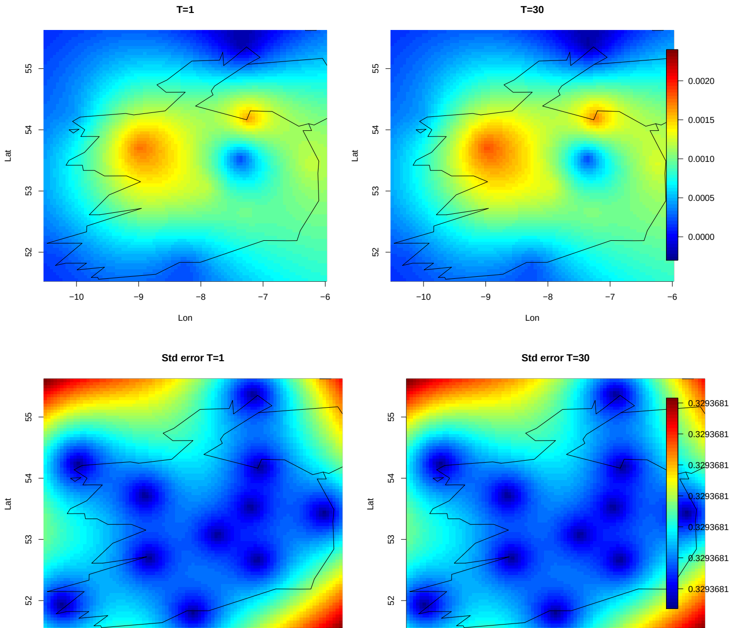

PLANO UBICACION
CASA CENTRAL
PORTERIA VALDES
PORTERIA PORTALES
CIMA
TALLER METAL MECANICO
BIOTECNOLOGIA
BIBLIOTECA
OBRAS CIVILES
TERRAZA EXPOSICIONES
CANCHAS
QUIMICA
METALURGIA
PROMOCION UNIVERSITARIA
LAB. SISMOLOGIA
LAB. CAD
PORTERIA PLACERES
GIMNASIO 2
MUSEO
DISEÑO DE PRODUCTOS
TUMBA FEDERICO SANTA MARIA
PORTERIA AV. ESPAÑA
PROCESOS QUIMICOS
MECANICA
CIAC
ELECTRICIDAD
DEFIDER / GIMNASIO 1
ELECTRONICA
AULA MAGNA
PATIO CENTRAL
ARQUITECTURA
RADIO
ESTUDIOS HUMANISTICOS
INDUSTRIAS / COMERCIAL
FISICA
PATIO CEIBOS
INFORMATICA
MATEMATICA
3IE
M
R
C
C
B
B
G
A
D
E
F
T
UNIVERSIDAD TECNICA FEDERICO SANTA MARIA
www.usm.cl
AVENIDA ESPAÑA 1680 · VALPARAISO · FONO 32-654259
info@usm.cl
LINEA 800·321·504

Ireland wind speed



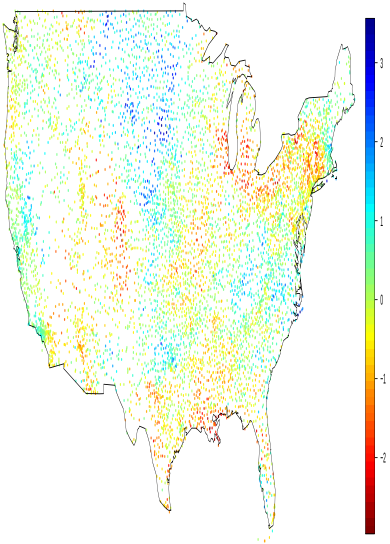
- Data available: 11 locations sites , 18 years for each day that is 72270 observations.
- Main goal: building a dynamic map of the wind speed over Ireland.
- A possible model? Space time Gaussian random field.
- Estimation? ML unfeasible.
- We need estimation methods with a good balance between computational complexity and statistical efficiency.

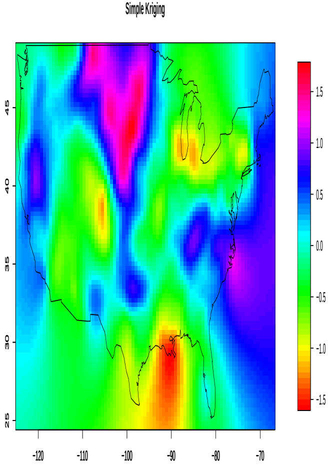
Irish wind temporal maps ($T = 1$, $T = 30$)



USA precipitation map

Simple kriging prediction using $pl_M(d)$ estimation.





State of the Art: Space-time Gaussian Fields

- $\{Z(\mathbf{s}, t), \mathbf{s} \in D \subset \mathbb{R}^d, \quad t \in \mathbb{R}\}$ Gaussian random fields (GRF).
- Covariance function

$$Cov(Z(\mathbf{s}_1, t_1), Z(\mathbf{s}_2, t_2))$$

- Weak Stationarity

$$E(Z(\mathbf{s})) = 0$$

$$\mathrm{Cov}(Z(\mathbf{s}_1, t_1), Z(\mathbf{s}_2, t_2)) = C(\mathbf{s}_1 - \mathbf{s}_2, t_1 - t_2) =: C(\mathbf{h}, u)$$

($\mathbf{h} = \mathbf{s}_1 - \mathbf{s}_2$, spatial lag, and $u := t_1 - t_2$ temporal lag).

Probabilistic Intermezzo

- Let $\mathbf{Z} = (Z(\mathbf{s}_1, t_1), Z(\mathbf{s}_2, t_2), \ldots, Z(\mathbf{s}_n, t_n))^T$ the sample data. Assume it is a (partial) realization of a space-time Gaussian field $Z(\mathbf{s}, t)$.
- $$\mathbf{Z} \sim N(\mathbf{0}, \Sigma)$$
- $\Sigma = \{K(||\mathbf{s}_i - \mathbf{s}_j||, |t_i - t_j|)\}_{ij=1,lk=1}^{n}$ a $n \times n$ matrix.
- $\Sigma \geq 0$ **iff** $K(\cdot)$ is a positive definite function.
- Then practical estimation generally requires first the selection of some parametric class of positive definite functions , *i.e.* $C(h) = C(h, \theta)$ and $\Sigma = \Sigma(\theta)$.

Covariance Functions: any tips?

$C : \mathbb{R}^d \to \mathbb{R}$ is a continuous covariance function

$$\Updownarrow$$

C is positive definite on $\mathbb{R}^d$.

$$\Updownarrow$$

C is the characteristic function of a random vector in $\mathbb{R}^d$

$$\Updownarrow$$

C is the Fourier transform of a positive and bounded measure μ

Same wine, different bottles

Mathematical Analysis Complex Analysis Harmonic Analysis Approximation Theory	Positive Definite
Probability Theory	Characteristic Function
Geostatistics Spatial Statistics Stochastic Processes	Covariance functions
Numerical Analysis	Radial Basis Functions
Machine Learning, Image Analysis	Kernels

Geostatistical prediction (Kriging)

- Main goal: prediction of Z at an unknown space-time of location $\mathbf{s}_0$ using information available from the sample data $\mathbf{Z}$.

$$\widehat{Z}(\mathbf{s}_0) = \mathbf{c}(\theta)^T \Sigma(\theta)^{-1} \mathbf{Z}$$

where $\mathbf{c}(\theta) = \mathrm{Cov}(\mathbf{Z}, Z(\mathbf{s}_0); \theta)$

- We need "good" estimation of θ to make "good" prediction.

Same wine, different bottles (2)

Mathematical Analysis Complex Analysis Harmonic Analysis Approximation Theory	Linear Projection Operators
Probability Theory	Best Linear Unbiased Prediction
Geostatistics Spatial Statistics Stochastic Processes	Kriging
Numerical Analysis	Radial Basis Functions Interpolator
Machine Learning, Image Analysis	Machine Learning Techniques

Stationarity and Isotropy

- Spatial Isotropy (Daley and Porcu, PAMS, 2013)

$$C(\mathbf{h}, 0) = K(\|\mathbf{h}\|, 0)$$

Examples

$$K(r) = \sigma^2 \exp(-r/b) \qquad r \geq 0,$$

$$K(r) = \sigma^2 (r/b)^{\nu} \mathcal{K}_{\nu}(r/b)$$

- So, K comes from some parametric family $\mathcal{C}_{\theta}$

A popular class

- The Gneiting class (Gneiting, 2002; Zastavnyi and Porcu, 2011)

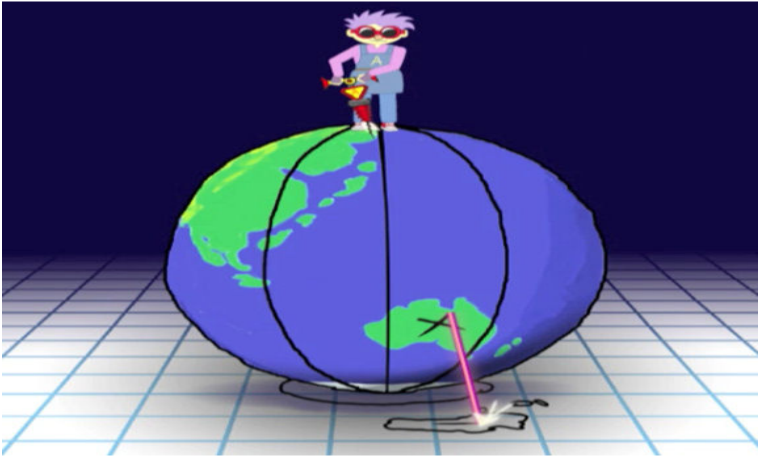
$$C(\mathbf{h}, u) = \frac{\sigma^2}{\psi(u^2)^{d/2}} \varphi\left(\frac{\|\boldsymbol{h}\|^2}{\psi(u^2)}\right)$$

- Example: $\varphi(t) = \mathrm{e}^{-t}$, $\psi(t) = (1+t)^{-1}$

$$C(\mathbf{h}, u) = \frac{\sigma^2}{\left(1 + \left(\frac{u}{c_t}\right)^2\right)^{d/2}} \exp\left(\frac{\|\boldsymbol{h}\|^2}{c_s\left(1 + \left(\frac{u}{c_t}\right)^2\right)}\right)$$

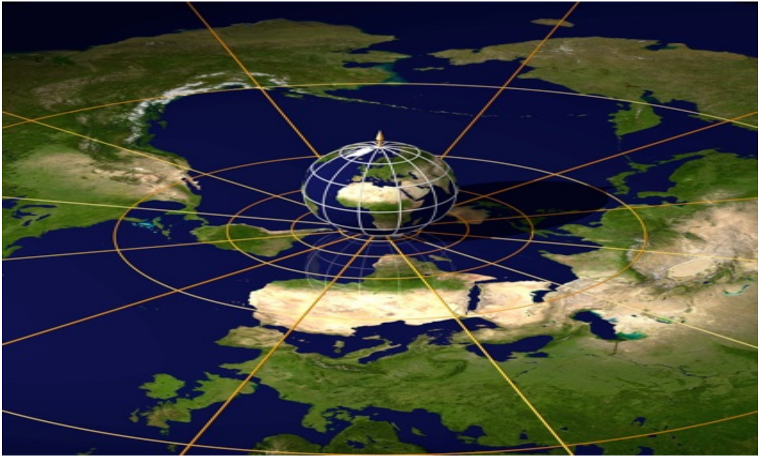
Spheres

Spatial Stats and the Euclidean Paradigm



Spheres

Spatial Stats and the Euclidean Paradigm

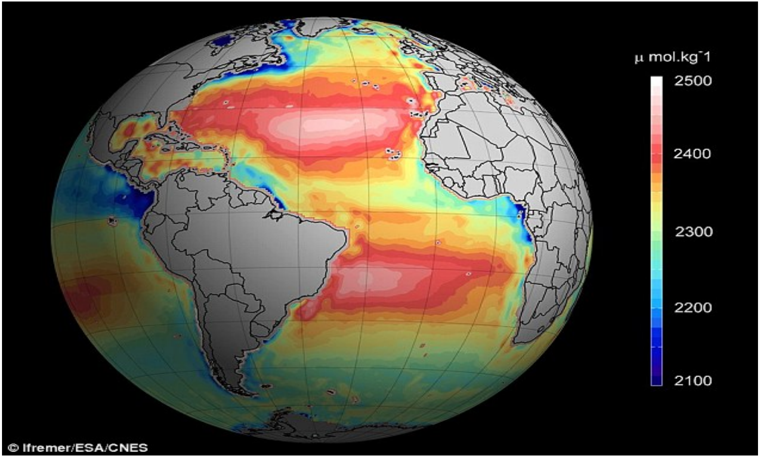


Spheres

Spatial Stats and the Euclidean Paradigm

Spheres

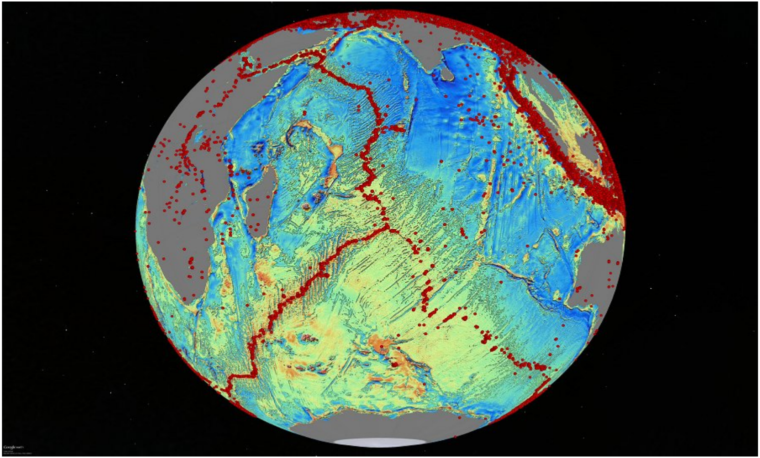
Spatial Stats and the Euclidean Paradigm



Spheres

Spatial Stats and the Euclidean Paradigm

copia.jpg



The Mercatore Projection

Spatial Stats and the Euclidean Paradigm

The Mercatore Projection

Spatial Stats and the Euclidean Paradigm

The Mercatore Projection

Spatial Stats and the Euclidean Paradigm

The Mercatore Projection

Spatial Stats and the Euclidean Paradigm

The Mercatore Projection

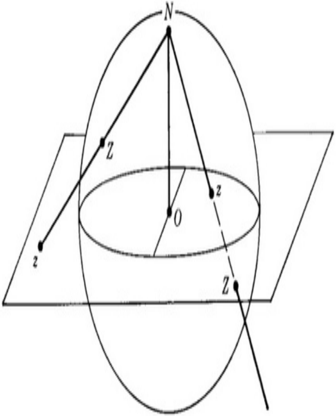
Spatial Stats and the Euclidean Paradigm

Spheres

Chordal and Great Circle

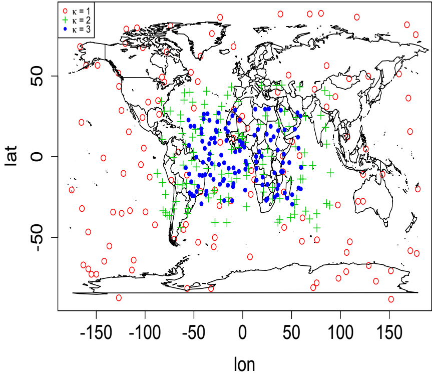
Chordal distance: criticism on

- Negative Correlations
- Counter to Spherical geometry



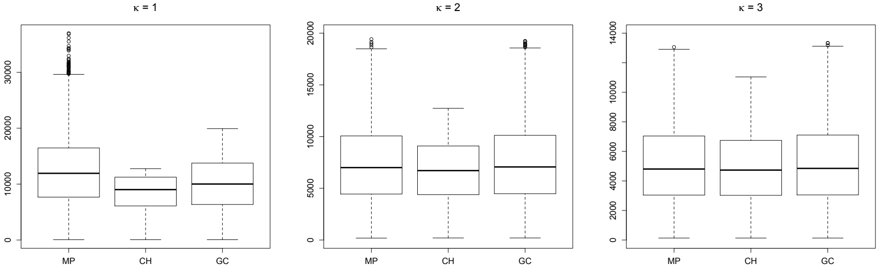
Spheres

A toy Example



Spheres

A toy example (2)



Word of Caution

- Suppose K is a isotropic covariance function. Can we replace the Euclidean with the geodesic distance?
- The answer is NOT.
- Example:

$$C(\mathbf{h}) = (\alpha\|\mathbf{h}\|)^{\nu}\mathcal{K}_{\nu}(\alpha\|\mathbf{h}\|), \qquad \mathbf{h} \in \mathbb{R}^{d}$$

Spheres

Space-Time Challenges

Frequently, the temporal development of a process observed on a sphere is also of interest, so that the process needs to be modeled on the sphere cross time. Nevertheless, the literature on the corresponding correlation structures is sparse [...]

Tilmann Gneiting, *Problem 16* of Online supplement to *Bernoullli* (2013).

Outline

1. motivations
2. Schoenberg coefficients and functions
3. The class $\Psi_{d,T}$
4. Generalizations
5. Construction Principles
6. Assessing discrepancies between the great circle distance and other metrics
7. Analysis of TOMS data
8. Computational Challenges
9. Methods of estimations
   - Covariance Tapering
   - Likelihood Approximations
   - Computational Challenges for simulation

Based on

1. **Porcu, E.**, Bevilacqua, M. & Genton, M.G. (2016). *Journal of the American Statistical Association*. To appear.
2. Berg, C. & **Porcu, E.** (2016). *Constructive Approximation*. To appear.

Spheres

The sphere

- d-dimensional unit sphere of $\mathbb{R}^{d+1}$, given as

$$\mathbb{S}^d = \{x \in \mathbb{R}^{d+1} \mid ||x|| = 1\}, \; d \geq 1.$$

- Great Circle distance: $\theta : \mathbb{S}^d \times \mathbb{S}^d \to [0, \pi]$,

$$\theta(\xi, \eta) = \arccos(\xi \cdot \eta),$$

- Chordal Distance: $\mathrm{ch}(\xi, \eta) = 2 \sin\left(\frac{\theta}{2}\right)$.

Motivations

Gaussian fields on $\mathbb{S}^d \times \mathbb{R}$

- Stationary Gaussian fields $\{Z(\xi, t),\ (\xi, t) \in \mathbb{S}^d \times \mathbb{R}\}$,
- Covariance functions $C : \mathbb{S}^d \times \mathbb{S}^d \times \mathbb{R}$, so that

$$\boxed{\operatorname{cov}\left(Z(\xi, t), Z(\eta, t')\right) = C(\theta(\xi, \eta), t - t')}, \qquad (\xi, t), (\eta, t') \in \mathbb{S}^d \times \mathbb{R}.$$

- Class $\Psi_{d,T}$ of continuous functions $f : [-1, 1] \times \mathbb{R} \to \mathbb{C}$ such that C can be written as

$$C(\theta(\xi, \eta), t - t') = \boxed{f(\cos\theta(\xi, \eta), t - t')}, \quad \xi, \eta \in \mathbb{S}^d,\ t, t' \in \mathbb{R}. \tag{1}$$

Spheres and Schoenberg Class

An Intermezzo: The Class Ψ_d

- We also consider

$$\Psi_{\infty,T} := \bigcap_{d=1}^{\infty} \Psi_{d,T},$$

$$\mathbb{S}^{\infty} = \{(x_k)_{k\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}} | \sum_{k=1}^{\infty} x_k^2 = 1\},$$

  which is the unit sphere in the Hilbert sequence space ℓ_2 of square summable real sequences.

- Inclusion Relation

$$\Psi_{1,T} \supset \Psi_{2,T} \supset \ldots \supset \Psi_{\infty,T}.$$

Spheres and Schoenberg Class

An Intermezzo: The Class Ψ_d

- The class $\Psi_{d,T}$ is parenthetical to the class Ψ_d of continuous functions $\psi : [0, \pi] \to \mathbb{R}$ such that $\psi(\theta)$ is a covariance function on $\mathbb{S}^d \times \mathbb{S}^d$.

Schoenberg's Class: an Intermezzo

Iso Schoenberg

21-04-1903 in Galati, 21-04-1990

- Son of a medical doctor
- 1922: M.A. at Jessy University
- 1922: Goettingen (Schur)
- 1925: Edmund Landau

1930: Harvard and Princeton

- *the isometric imbedding of metric spaces into Hilbert space and positive definite functions.*
- 1950: Pólya
- 1966: University of Pennsylvania

He was...[by Richard Askey]

a man of broad culture, fluent in several languages, addicted to art, music and world literature, sensitive, gracious and giving in all ways. The working desk at his home where he engages in research is actually a draftsman's bench complete with T-square, etc. and a tall stool. Mobiles, artistic works, models of ruled surfaces, icosahedrons and other objects are strewn throughout the room. English, French and German novels, numerous paintings and artefacts are scattered on all the nearby easy chairs.

Spheres: how to build the Class Ψ_d

The Class Ψ_d: How to build it

- Gegenbauer polynomials

$$(1 - 2xr + r^2)^{-\lambda} = \sum_{n=0}^{\infty} C_n^{(\lambda)}(x) r^n, \quad |r| < 1, x \in \mathbb{C}.$$

- For $\lambda > 0$,

$$\int_{-1}^{1} (1 - x^2)^{\lambda - 1/2} C_n^{(\lambda)}(x) C_m^{(\lambda)}(x) \, \mathrm{d}x = \frac{\pi \Gamma(n + 2\lambda) 2^{1-2\lambda}}{\Gamma^2(\lambda)(n + \lambda) n!} \delta_{m,n}.$$

Spheres: how to build the Class Ψ_d

The Class Ψ_d

- Important!

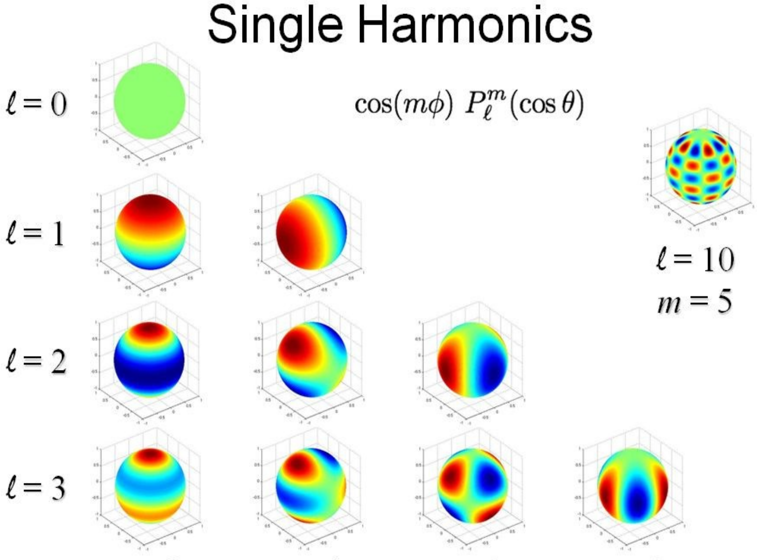
$$|C_n^{(\lambda)}(x)| \leq C_n^{(\lambda)}(1), \quad x \in [-1,1].$$

- $\lambda = (d-1)/2$ and its connection with spherical harmonics.
- $n \in \mathbb{N}$. Vector Space $\boxed{\mathcal{H}_n(d) \subset \mathcal{C}(\mathbb{S}^d)}$, dimension

$$N_n(d) := \dim \mathcal{H}_n(d) = \frac{(d)_{n-1}}{n!}(2n+d-1), \; n \geq 1, \quad N_0(d) = 1,$$

Spheres: how to build the Class Ψ_d

The Class $\Psi_{d,T}$



Spheres: how to build the Class Ψ_d

The Class Ψ_d

- Orthogonality relation:

$$\int_{-1}^{1}(1-x^2)^{d/2-1}c_n(d,x)c_m(d,x)\,\mathrm{d}x = \frac{||\omega_d||}{||\omega_{d-1}||N_n(d)}\delta_{m,n}.$$

- $\boxed{\textit{Theorem}(\textit{Schoenberg}, 1942)}$ A continuous function $f:[-1,1]\to\mathbb{R}$ belongs to the class Ψ_d, $d=1,2,\ldots$, if and only if

$$f(\cos\theta)=\sum_{n=0}^{\infty} b_{n,d}c_n(d,\cos\theta), \quad b_{n,d}\geq 0, \theta\in[0,\pi],$$

for a summable sequence $(b_{n,d})_{n=0}^{\infty}$ given as

$$b_{n,d}=\frac{||\omega_{d-1}||N_n(d)}{||\omega_d||}\int_{-1}^{1} f(x)c_n(d,x)(1-x^2)^{d/2-1}\,\mathrm{d}x.$$

Spheres: how to build the Class Ψ_d

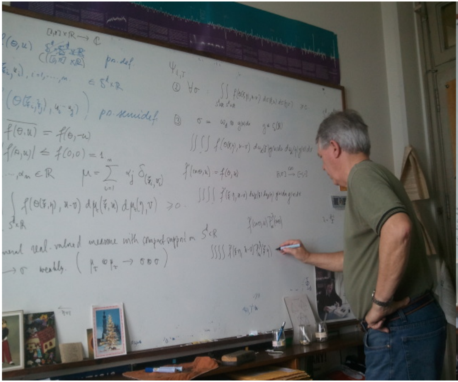
The Class $\Psi_{d,T}$

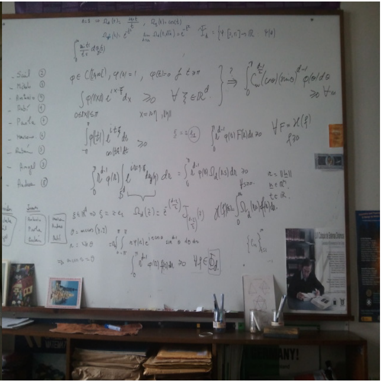
- If $f(0) = 1$, $(b_{n,d})$ is a probability sequence
- Daley and Porcu (2013) call $b_{n,d}$ d-Schoenberg coefficients and the sequence $(b_{n,d})$ a d-Schoenberg sequence
- In $d = 1$,

$$f(\cos\theta) = \sum_{n=0}^{\infty} b_{n,1} \cos(n\theta), \quad b_{n,1} \geq 0, \theta \in [0, \pi],$$

Spheres

Characterization of $\Psi_{d,T}$





Results

The Class $\Psi_{d,T}$: Characterization Theorems

Theorem 1 Let $d \in \mathbb{N}$ and let $f : [-1,1] \times \mathbb{R} \to \mathbb{C}$ be a continuous function. Then f belongs to $\Psi_{d,T}$ if and only if there exists a sequence $\varphi_{n,d} : \mathbb{R} \to \mathbb{R}$ of p.d. functions with $\sum \varphi_{n,d}(0) < \infty$ such that

$$\boxed{f(\cos\theta, t) = \sum_{n=0}^{\infty} \varphi_{n,d}(t) c_n(d, \cos\theta),}$$

and the above expansion is uniformly convergent for $(\theta, t) \in [0, \pi] \times \mathbb{R}$. We have

$$\varphi_{n,d}(t) = \frac{N_n(d)||\omega_{d-1}||}{||\omega_d||} \int_{-1}^{1} f(x,t) c_n(d,x)(1-x^2)^{d/2-1}\,\mathrm{d}x.$$

Results

The Class $\Psi_{d,T}$

Theorem 2. Let $d \in \mathbb{N}$ and suppose that $f \in \Psi_{d+2,T} \subset \Psi_{d,T}$. Then we have

(a) For $d = 1$,

$$\varphi_{0,3} = \varphi_{0,1} - \frac{1}{2}\varphi_{2,1}$$

and

$$\varphi_{n,3} = \frac{1}{2}(n+1)(\varphi_{n,1} - \varphi_{n+2,1}), \quad n \geq 1.$$

(b) For $d \geq 2$,

$$\varphi_{n,d+2} = \frac{(n+d-1)(n+d)}{d(2n+d-1)}\varphi_{n,d} - \frac{(n+1)(n+2)}{d(2n+d+3)}\varphi_{n+2,d}, \quad n \geq 0.$$

Results

The Class $\Psi_{\infty,T}$

Theorem 3. Let $f : [-1,1] \times \mathbb{R} \to \mathbb{C}$ be a continuous function. Then f belongs to $\Psi_{\infty,T}$ if and only if there exists a sequence $\varphi_n : \mathbb{R} \to \mathbb{R}$ of p.d. functions with $\sum \varphi_n(0) < \infty$ such that

$$\boxed{f(\cos\theta, t) = \sum_{n=0}^{\infty} \varphi_n(u) \cos^n \theta,}$$

and the above expansion is uniformly convergent for $(\theta, u) \in [0,\pi] \times \mathbb{R}$.
Moreover,

$$\lim_{d\to\infty} \varphi_{n,d}(t) = \varphi_n(t) \text{ for all } (n,t) \in \mathbb{N}_0 \times \mathbb{R}. \tag{2}$$

Menegatto's School

Collaborations with Ana Peron, Rafaela Neves Bonfim, Thaís Jordão, Jean Carlo Guella, Mario de Castro (the short one), Victor Barbosa, Jose C. Ferreira.
MANY PAPERS including the very difficult case of strict positive definiteness.

Construction principles

with Moreno Bevilacqua and Marc Genton

Results

The Class $\Psi_{d,T}$

A natural construction:

$$\boxed{Z(\eta, t) = \sum_{k=0}^{\infty} \sum_{\nu \in \Upsilon_{k,d}} \xi_{k,\nu}(t) Y_{k,\nu,d}(\eta), \qquad \eta \in \mathbb{S}^d, t \in \mathbb{R},}$$

- $Y_{k,\nu,d} : \mathbb{S}^d \to \mathbb{C}$: normalized hyperspherical harmonics;
- $\xi_{k,\nu}(t)$: Gaussian processes, with zero mean and $\mathbb{E}\xi_{k,\nu}(t)\xi_{k',\nu'}(t') = \delta_{k,k'}\delta_{\nu,\nu'} g_k(t - t')$, $t, t' \in \mathbb{R}$.

Construction Principles

An easy construction principle

Let $\{g_k(\cdot)\}_{k=0}^{\infty}$ be an absolutely convergent sequence of continuous and positive definite functions on the real line, such that $g_k(0) = b_k$ for all $k = 0, 1, \ldots$, with $\{b_k\}_{k=0}^{\infty}$ being a probability mass sequence. Then,

$$\boxed{C(\theta, u) = \sum_{k=0}^{\infty} g_k(u)\,(\cos\theta)^k\,, \qquad (\theta, u) \in [0, \pi] \times \mathbb{R},}$$

is a representation for members of the class $\Psi_{\infty,T}$.

Results

The Class $\Psi_{d,T}$

Family	Analytic expression	Parameters range
Negative Binomial	$C(\theta,u)=\left\{\frac{1-\varepsilon}{1-\varepsilon g(u)\cos\theta}\right\}^{\tau}$	$\varepsilon\in(0,1),\ \tau>0$
Multiquadric	$C(\theta,u)=\frac{(1-\varepsilon)^{2\tau}}{\{1+\varepsilon^2-2\varepsilon g(u)\cos\theta\}^{\tau}}$	$\varepsilon\in(0,1),\ \tau>0$
Sine Series	$C(\theta,u)=e^{g(u)\cos\theta-1}\{1+g(u)\cos\theta\}/2$	
Sine Power	$C(\theta,u)=1-2^{-\alpha}\{1-g(u)\cos\theta\}^{\alpha/2}$	$\alpha\in(0,2]$
Adapted Multiquadric	$C(\theta,u)=\left[\frac{\{1+g^2(u)\}(1-\varepsilon)}{1+g^2(u)-2\varepsilon g(u)\cos\theta}\right]^{\tau}$	$\varepsilon\in(0,1),\tau>0$ $2g(\cdot)/\{1+g^2(\cdot)\}$ corr. function on $\mathbb{R}$
Poisson	$C(\theta,u)=\exp\left[\lambda\{\cos\theta g(u)-1\}\right]$	$\lambda>0$

Results

Spatial Adapting from the Gneiting class

- Gaussian process Z on $\mathbb{R}^d \times \mathbb{R}$, points $(\boldsymbol{x}, t)$ and $(\boldsymbol{y}, t')$ such that $\|\boldsymbol{y} - \boldsymbol{x}\| = \boldsymbol{h}$ (with $\|\cdot\|$ denoting the Euclidean distance) and $t - t' = u$,

$$C(\boldsymbol{h}, u) := \frac{\sigma^2}{\psi(\|\boldsymbol{h}\|^2)^{d/2}} \varphi \left\{ \frac{u^2}{\psi(\|\boldsymbol{h}\|^2)} \right\}, \qquad (\boldsymbol{h}, u) \in \mathbb{R}^d \times \mathbb{R},$$

- φ is completely monotone on the positive real line such that $\varphi(0) = 1$, ψ is a positive-valued Bernstein function, and σ^2 is a variance parameter.

Results

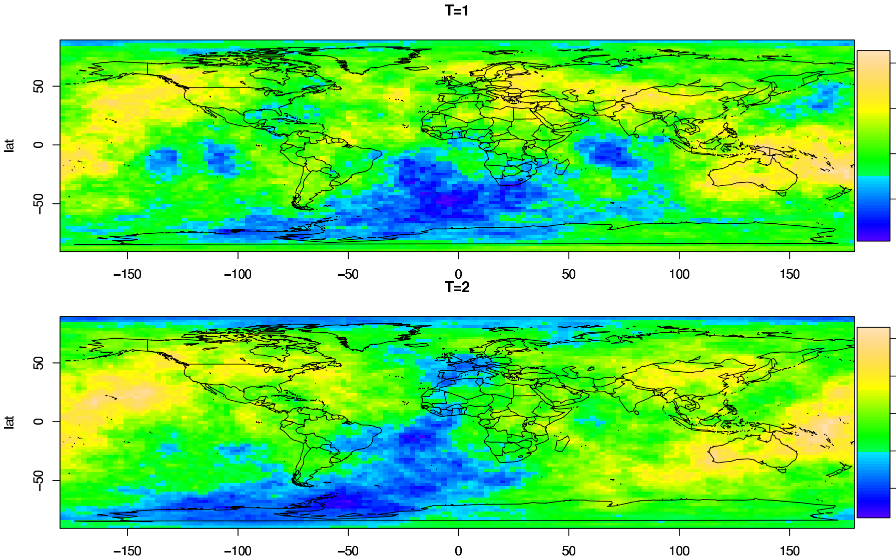
Spatial Adapting from the Gneiting class

Let $\theta : \mathbb{S}^d \times \mathbb{S}^d \to [0, \pi]$ be the great circle distance. Let $\varphi : [0, \infty) \to \mathbb{R}_+$ be a completely monotone function on the positive real line, with $\varphi(0) = 1$, and let ψ be a positive-valued Bernstein function. Denote by $\psi_{[0,\pi]}$ the restriction of ψ to the interval $[0, \pi]$. Then, the function

$$C(\theta, u) := \frac{\sigma^2}{\psi_{[0,\pi]}(\theta)^{1/2}} \varphi \left\{ \frac{u^2}{\psi_{[0,\pi]}(\theta)} \right\}, \qquad (\theta, u) \in [0, \pi] \times \mathbb{R}, \qquad (3)$$

belongs to the class $\Psi_{\infty,T}$.

Gneiting class on the sphere



Results

Relevant Comments

- Mean square differentiability for processes on spheres. Attempts in Jeong and Jun (2015).
- Exception being the Sine Power model.
- Adapted construction has the same problem: Matérn is valid only for $\nu \in (0, 1/2]$.
- Adapted construction allows for rescaling the spatial component, direct construction not
- Direct construction allow for any type of temporal margin, provided g is a temporal correlation function.

Results

Examples from The Adapted Gneiting Class

- $$C(\theta, u) = \frac{\sigma^2}{\left\{1 + \left(\frac{R\theta}{c_S}\right)^{\alpha}\right\}^{\delta+\beta/2}} \exp\left[-\frac{\left(\frac{|u|}{c_T}\right)^{2\gamma}}{\left\{1 + \left(\frac{R\theta}{c_S}\right)^{\alpha}\right\}^{\beta\gamma}}\right],$$

- $$C(\theta, u) = \frac{\sigma^2}{\left\{1 + \left(\frac{R\theta}{c_S}\right)^{\alpha}\right\}^{\delta+\beta/2}} \left[1 + \frac{\left(\frac{|u|}{c_T}\right)^{2\tau}}{\left\{1 + \left(\frac{R\theta}{c_S}\right)^{\alpha}\right\}^{\tau\beta}}\right]^{-\lambda},$$

Results

Examples from The Adapted Gneiting Class

- Take the negative binomial family and $u \mapsto g(u;\alpha) := (1+|u|^{\alpha})^{-1}$, $\alpha \in (0,2]$,

$$C(\theta,u) = \sigma^2 \left[\frac{1-\varepsilon}{1-\varepsilon\left\{1+\left(\frac{|u|}{c_T}\right)^{\alpha}\right\}^{-1}\cos\theta} \right]^{\tau}, \qquad (\theta,u) \in [0,\pi]\times\mathbb{R},$$

- From the multiquadric,

$$C(\theta,u) = \frac{\sigma^2(1-\varepsilon)^{2\tau}}{\left[1+\varepsilon^2-2\varepsilon\left\{1+\left(\frac{|u|}{c_T}\right)^{\alpha}\right\}^{-1}\cos\theta\right]^{\tau}}, \qquad (\theta,u) \in [0,\pi]\times\mathbb{R}, \tag{4}$$

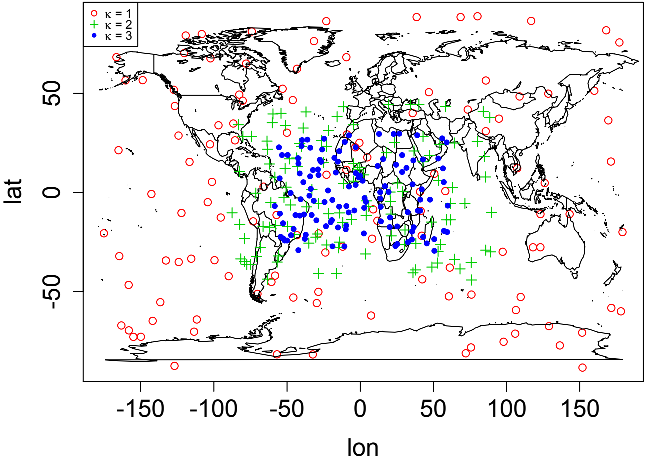
with the same restriction on the parameters as in the previous model.

-

$$C(\theta,u) = \sigma^2 \upsilon \left[1 + \frac{\cos\theta}{c_S\left\{1+\left(\frac{|u|}{c_T}\right)^{\alpha}\right\}^{-1}} \right] \exp\left[\frac{\cos\theta}{c_S\left\{1+\left(\frac{|u|}{c_T}\right)^{\alpha}\right\}^{-1}} \right], \qquad (\theta,u) \in [0,\pi]\times\mathbb{R},$$

Simulation Studies

Scenarios



Simulation Studies

Scenarios

- Three scenarios;
- Estimate $\boldsymbol{\lambda}$ using ML;
- Under $C(\theta, u; \boldsymbol{\lambda}) = \frac{\sigma^2}{\left\{1+\left(\frac{R\theta}{c_S}\right)\right\}} \exp\left[-\frac{|u|}{c_T\left\{1+\left(\frac{R\theta}{c_S}\right)\right\}^{1/4}}\right]$,
- Using either the GC, CH or MP distances. Notation $\widehat{\boldsymbol{\lambda}}_{\mathcal{X}}^{(k)}$, with $\mathcal{X} =$ GC, CH or MP
- $k = 1, \ldots, 1000$ Simulations.

Simulation Studies

Scenarios

Discrepancy between the ML estimates using either GC, CH and MP distances. Given $\widehat{\boldsymbol{\lambda}}_{\mathcal{X}}^{(k)}$, we call $M(\cdot)$ the measure

$$M^{\mathcal{X}}({}_i\widehat{\lambda}) = \sqrt{\frac{\sum_{k=1}^{1000}({}_i\widehat{\lambda}_{\mathrm{GC}}^{(k)} - {}_i\widehat{\lambda}_{\mathcal{X}}^{(k)})^2}{1000}}, \qquad i = 1,2,3, \qquad \mathcal{X} = \mathrm{CH}, \mathrm{MP}. \tag{5}$$

We also define another measure $A(\cdot)$ by

$$A^{\mathcal{X}}({}_i\widehat{\lambda}) = \sqrt{\frac{\sum_{k=1}^{1000}({}_i\widehat{\lambda}_{\mathcal{X}}^{(k)} - {}_i\lambda)^2}{1000}}, \qquad i = 1,2,3, \qquad \mathcal{X} = \mathrm{GC}, \mathrm{CH}, \mathrm{MP}, \tag{6}$$

where ${}_i\lambda$ denotes the nominal value chosen under one of the proposed scenarios.

Simulation Studies

Scenarios

		$\kappa = 1$		$\kappa = 2$		$\kappa = 3$
	(I)	(II)	(I)	(II)	(I)	(II)
$M^{CH}(\hat{c}_S)$	6.25	24.07	0.97	4.06	0.42	2.19
$M^{MP}(\hat{c}_S)$	93.03	201.99	9.86	17.45	3.17	5.91
$M^{CH}(\hat{c}_T)$	0.013	0.014	0.004	0.008	0.004	0.007
$M^{MP}(\hat{c}_T)$	0.058	0.080	0.015	0.022	0.007	0.012
$M^{CH}(\hat{\sigma}^2)$	0.004	0.011	0.001	0.003	0.001	0.002
$M^{MP}(\hat{\sigma}^2)$	0.026	0.045	0.005	0.008	0.002	0.004
$A^{GC}(\hat{c}_S)$	116.47	206.77	67.74	111.52	45.67	74.90
$A^{CH}(\hat{c}_S)$	117.73	211.78	68.00	112.34	45.73	74.98
$A^{MP}(\hat{c}_S)$	133.73	261.39	68.59	112.96	45.89	74.16
$A^{GC}(\hat{c}_T)$	0.212	0.211	0.212	0.212	0.212	0.211
$A^{CH}(\hat{c}_T)$	0.212	0.212	0.212	0.212	0.212	0.212
$A^{MP}(\hat{c}_T)$	0.220	0.226	0.213	0.213	0.213	0.212
$A^{GC}(\hat{\sigma}^2)$	0.088	0.101	0.084	0.094	0.083	0.093
$A^{CH}(\hat{\sigma}^2)$	0.088	0.104	0.084	0.094	0.083	0.093
$A^{MP}(\hat{\sigma}^2)$	0.089	0.105	0.084	0.095	0.083	0.093

TOMS DATA

Scenario

- Level-3 Total Ozone Mapping Spectrometer (TOMS): daily total column ozone levels.
- Spatially irregular grid (1^o latitude by 1.25^o longitude away from the poles)
- Original data: Latitude interval $[-89.5, 89.5]$ and longitudes $[-180, 180]$
- Jun and Stein (2008): spatial dataset
- Here: 15 obs. in time, for a total of of $20,160$ points (288 longitudinal and 70 latitudinal) observed during 15 days, for a total of $302,400$ observations.

TOMS DATA

Scenario

- For the missing data: follow Jun and Stein (2008): local averaging (24 observations) for each local averaging.
- Likelihood estimation unfeasible: select a subgrid of 336 spatial points and all temporal observations, for a total of $5,040$ observations.
- Detrend the data using spatio-temporal splines
- Residuals as a realization from a zero mean space-time Gaussian random field.

TOMS DATA

Scenario

A. Two models based on the adapted Gneiting classes

B. Three models based on direct construction, hence valid with GC only.

C. A model based on the Gneiting class valid using CH and MP distances

TOMS DATA

Scenario

Distance	GC	CH	MP	GC	CH	MP	GC	CH	MP
Model		A.1			A.2			C.1	
c_S	742.7	743.9.3	734.2	681.3	672.2	733.3	-	450.8	417.5
c_T	2.54	2.54	2.17	1.76	1.67	1.71	-	1.56	1.29
β	1	1	0.95	1	1	1	-	0.95	0.89
σ^2	102.6	102.1	102.7	106.1	103.0	111.4	-	97.9	98.5
Nugget	9.81	9.82	6.35	6.30	5.60	4.97	-	12.26	5.72
Likelihood	-17233.8	-17234.0	-17296.5	-17257.1	-17258.5	-17317.3	-	-17156.3	-17223.9

Model		B.1			B.2			B.3	
τ	0.01	-	-	144.09	-	-	4.04	-	-
c_T	51.05	-	-	50.58	-	-	36.85	-	-
α	1.49	-	-	1.49	-	-	1.61	-	-
σ^2	85.8	-	-	86.0	-	-	89.9	-	-
Nugget	18.60	-	-	18.54	-	-	16.84	-	-
Likelihood	-17168.3	-	-	-17167.9	-	-	-17162.8	-	-

Computational Challenges for Space-Time Data

Maximum likelihood estimation

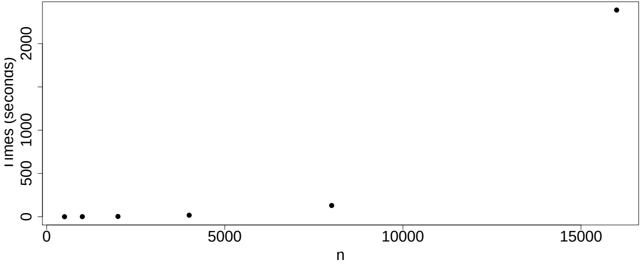
- Require the distribution of the underlying RF to be known.
- If $\mathbf{Z} \sim N(0, \Sigma(\theta))$. Then

$$l(\theta) = -\frac{1}{2}\log\det\Sigma(\theta) - \frac{1}{2}\mathbf{Z}'\Sigma(\theta)^{-1}\mathbf{Z} \tag{7}$$

- The most critical part of the likelihood calculation is to evaluate the determinant and inverse of the covariance matrix. Each calculation of the likelihood requires $O((nm)^3)$ operations.
- Under increasing domain the maximizer of $l(\theta)$ is consistent and asymptotically Gaussian with covariance matrix (Fisher information):

$$[F(\theta)]_{ij} = \frac{1}{2}\mathrm{tr}\left([\Sigma(\theta)]^{-1}\frac{d\Sigma}{d\theta_i}[\Sigma(\theta)]^{-1}\frac{d\Sigma}{d\theta_j}\right). \tag{8}$$

Computing $I(\theta)$



Time in seconds needed to evaluate the Gaussian likelihood as a function of the data using package GeoR of R software.

Computational problem

We need estimation methods with a good trade off between computational and statistical efficiency.
Some papers addressing the computational problem:

- Considering a different objective function (Vecchia, 1988 Stein et al. 2004), Bevilacqua, Porcu, Gaetan, Mateu (JASA 2012).
- Considering approximation of the covariance matrix (Kaufmann et al 2008), Daley, Porcu and Bevilacqua (SERRA 2013), Porcu, Bevilacqua and Genton (JRSSB 2015).
- Considering a slighty different model: Gaussian Markov Random Field approximations (Lindgren et al 2011)

The tapering approach (spatial case)

- Idea: correlations between pairs of distant sampling locations are often nearly zero
- Taper: A correlation model with compact support.
  An example from the Wendland class:

$$Tap(h; d) = \left(1 - \frac{\|h\|}{d}\right)_+^4 \left(1 + 4\frac{\|h\|}{d}\right)$$

- Tapering a covariance model. An example:

$$C_{Tap}(h, \theta, d) = \left(\sigma^2 e^{-\frac{\|h\|}{b}}\right) Tap(h; d)$$

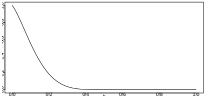
- Tapering a covariance matrix:

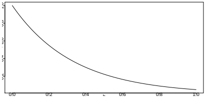
$$\Sigma_{Tap}(\theta, d) = \Sigma(\theta) \circ T(d)$$

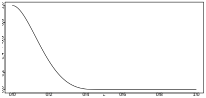
- $\Sigma_{Tap}(\theta, d)$ is a sparse matrix.

The tapering approach

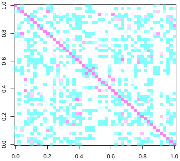
$$C_{Tap}(h,\theta,d) \quad = \quad \left(\sigma^2 e^{-\frac{h}{b}}\right) \quad \times \quad Tap(h,d)$$

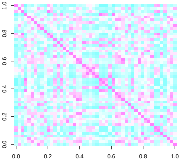


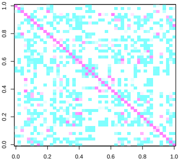




$$\Sigma_{Tap}(\theta,d) \quad = \quad \Sigma(\theta) \quad \circ \quad T(d)$$







Covariance tapering estimation

Kaufmann et al (2008) proposed to maximize the tapered likelihood:

$$l_{Tap}(\theta, d) = -\frac{1}{2}\log|\Sigma_{Tap}(\theta, d)|, -\frac{1}{2}\mathbf{Z}'(\Sigma_{Tap}(\theta, d)^{-1} \circ T(d))\mathbf{Z} \quad (9)$$

Features:

- Unbiased estimating equation.
- Sparse matrix algorithms can be used to compute the inverse and the determinant of the tapered matrix.
- Computational gains depends on the percentage of zero in $\Sigma_{Tap}(\theta, d)$ *i.e.* on the choice of d.
- $\lim_{d\to\infty} l_{Tap}(\theta, d) = l(\theta)$.

Composite likelihoods

General idea

1. Let $\mathbf{Z} = (Z_1, \ldots, Z_n)'$ be a n-dimensional vector random variable with density $f(\mathbf{Z}; \theta)$ for some unknown parameter $\theta \in \Theta \subseteq \mathbb{R}^p$.
2. Suppose that $f(\mathbf{Z}; \theta)$ is difficult to evaluate or to specify, but that it is possible to compute or specify distribution for some subsets of $\mathbf{Z}$.
3. It may be expedient to consider instead a pseudolikelihood compounding such likelihood objects.

Composite likelihood definition

- Let A_k be a marginal or conditional set of the data, the composite likelihood (CL) (Lindsay 1988) is an objective function defined as a product of K sub-likelihoods

$$CL(\theta) = \sum_{k=1}^{K} l(\theta; A_k) w_k,$$

- $l(\theta; A_k)$ is the likelihood generated from $log(f(z;\theta))$ by considering only the random variables in A_k
- w_k are suitable non negative weights that do not depend on θ.
- The maximum CL estimate is given by $\hat{\theta} = \mathrm{argmax}_\theta \, CL(\theta)$.

Composite likelihoods drawbacks and benefits

1. **Drawback I**: general loss of statistical efficiency is expected from CL estimation with respect to ML methods.
2. **Drawback II**: set of estimating methods very large. How one can choose in this set?
3. **Benefit I**: computational tractability. (Our case)
4. **Benefit II**: it requires only model assumptions on lower dimensional marginal densities, and not detailed specification of the full joint

Composite likelihoods based on pairs . pl_X, $X = M, C, D$

- Setting $A_k = (Z(s_i), Z(s_j))$, we obtain the pairwise marginal Gaussian likelihood l_{ij} and the function

$$pl_M(\theta) = \sum_{i=1}^{n} \sum_{j>i}^{n} l_{ij}(\theta) w_{ij}$$

- Setting $A_k = (Z(s_i)|Z(s_j))$ we obtain the pairwise conditional Gaussian likelihood $l_{i|j}$ and the function

$$pl_C(\theta) = \sum_{i=1}^{n} \sum_{j\neq i}^{n} l_{i|j}(\theta) w_{ij} = \sum_{i=1}^{n} \sum_{j>i}^{n} (2l_{ij}(\theta) - l_i(\beta) - l_j(\beta)) w_{ij}$$

where $l_i(\beta)$ is the marginal likelihood and $\beta \subseteq \theta$.

- Setting $A_k = (Z(s_i) - Z(s_j))$ we obtain the pairwise difference Gaussian likelihood l_{i-j}.

$$pl_D(\theta) = \sum_{i=1}^{n} \sum_{j>i}^{n} l_{i-j}(\theta) w_{ij}$$

pl_X methods: Comments

- The number of operations requested $O(N^2)$ (no matrix inversion is requested).
- Unbiased estimating equations.
- Which kind of weights?
- Which is the best method among $pl_X, X = M, D, C$?

pl_X methods: improving the efficiency

- Specific choice of the weights w_{ij} allows to improve the statistical efficiency of pl_X methods.
- We show that a convenient choice for the weights is:
  $$w_{ij} = \begin{cases} 1 & ||s_i - s_j|| \leq d \\ 0 & otherwise \end{cases}$$
- Then pl_X depends on the choice of d i.e. $pl_X(d)$.
- This kind of weights allows further computational benefits.
- An alternative method to improve the efficiency can be found in Bevilaqua, Gaetan, Mateu and Porcu (JASA 2012).

Asymptotic results for pl_X estimator, $X = P, D, C$

- Let θ_0 the true value and $\hat{\theta}_X^n$ the (sequence of) estimators obtained expanding the domain in space and/or in time in an increasing domain fashion
- Let ∇pl_X^n the estimating equation vector associated to pl_X^n Then it can be shown:
- $\hat{\theta}_X^n \xrightarrow{p} \theta_0, \quad X = P, D, C$
- Under specific conditions on the underliyng RF

$$(J_X^n(\theta_0))^{-1/2} H_X^n(\theta_0)(\hat{\theta}_n - \theta_0) \xrightarrow{d} \mathcal{N}(0, I_p) \tag{10}$$

where

$$H_X^n(\theta) = -\mathbb{E}[\nabla^2 pl_X^n(\theta)], \qquad J_X^n(\theta) = Var[\nabla pl_X^n(\theta)] \tag{11}$$

pl_X: comparing statistical efficiency

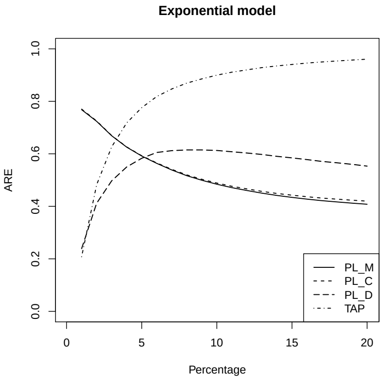
- We find closed form expression for the asymptotic variance of pl_X estimators.
- Comparing them we can say that there is no clear evidence that a method outperforms the others.

An example:

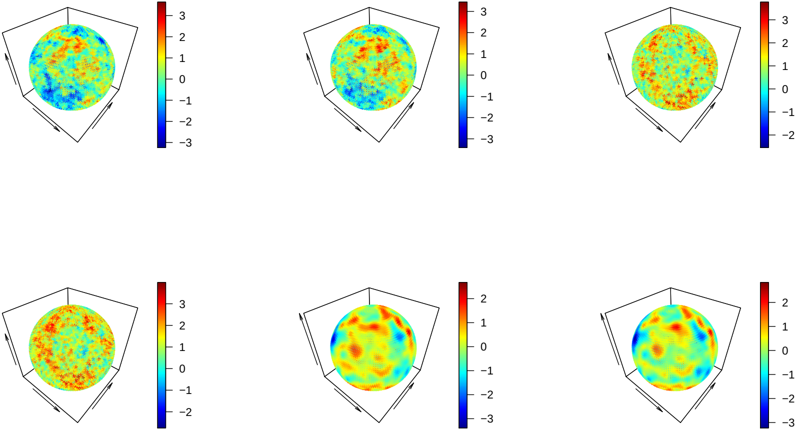
- Estimating a GRF with exponential model ($\sigma^2 = 1$, $b = 0.4$) observed on 500 sites on the square $[0, 1]^2$.
- Comparing $pl_X(d)$ and $TAP(d)$ asymptotic variances using

$$ARE_a(d) = \left(\frac{|G_a(\theta; d)|}{|F_{ML}(\theta)|} \right)^{1/p}, \quad a = C, D, M, T, \quad p = 2$$

.



Simulation Challenges



Hope to re-visit Brasil very soon!

Thank you for your Attention!