BREAKOUT SESSION D: NON-GAUSSIAN PROCESSES

Alexandra M. Schmidt

EBOH, McGill University, Canada

Homepage: http://alex-schmidt.research.mcgill.ca/

Challenges in the Statistical Modeling of Stochastic Processes for the Natural Sciences in honor of Peter Guttorp

Banff International Research Station for Mathematical Innovation and Discovery, Canada

July 2017

Peter has had a great influence on my life

- 2000 : Two PhD students working on the Bayesian approach of the Sampson & Guttorp spatial deformation $\rightarrow$ Invited me to give a talk in a session at the JSM 2000, followed by a Discussion group in Seattle
- Final Workshop of the HSSS (Luminy 2000), I ended up being a discussant because Peter couldn't make it $\Rightarrow$ Alan Gelfand hired me to be his Post-Doc
- Our papers on "*cov in cov*" (Environmetrics, 2011 and JRSS Series C, 2014)
- Organization of PASI 2014
- Most recent one, my move from Rio de Janeiro to Montreal

Discussion about organizing PASI (2012)

- Let's do it **during** the World Cup in 2014? (!!!!!???@#@#@#)
- Peter's flexibility (**schedule**, Brazil's **crazy bureaucracy**, **high** prices)

PASI 2014 - And we did it!!!

http://www.stat.washington.edu/peter/PASI/PASI_2014.html

80 participants (more than a half were Graduate students) from various parts of the world: Brazil, Canada, Chile, Germany, Norway, Sweden, Switzerland, USA, Venezuela...

Geostatistics

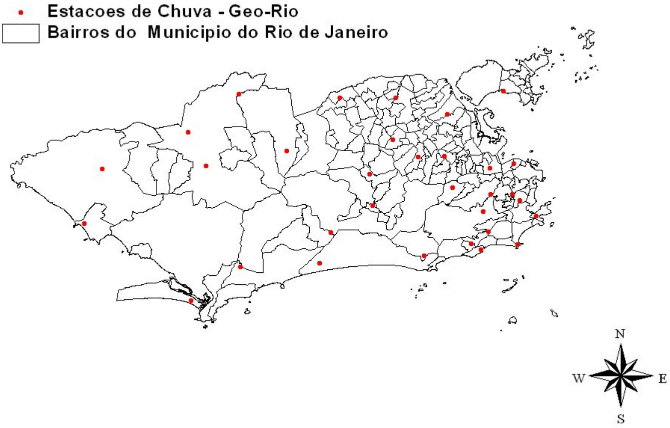

Figure: Rainfall monitoring stations in the city of Rio de Janeiro.

Stochastic Processes

Basic Model: Data (**Y**) are a (partial) realization of a random process (*stochastic process* or *random field*)

$$\{Y(\mathbf{s}) : \mathbf{s} \in D\}$$

where D is a fixed subset of R^d with positive d-dimensional volume. In other words, the spatial index **s** varies *continously* throughout the region D.

Gaussian Processes

A function $Y(\cdot)$ taking values $y(\mathbf{s})$ for $\mathbf{s} \in D$ has a Gaussian process distribution with mean function $m(\cdot)$ and covariance function $c(\cdot,\cdot)$, denoted by

$$Y(\cdot) \sim GP(m(\cdot), c(\cdot,\cdot))$$

if for any $\mathbf{s}_1, \cdots, \mathbf{s}_n \in D$, and any $n = 1, 2, \cdots$, the joint distribution of $Y(\mathbf{s}_1), \cdots, Y(\mathbf{s}_n)$ is multivariate Normal with parameters given by

$$E\{Y(\mathbf{s}_j)\} = m(\mathbf{s}_j) \text{ and } Cov(Y(\mathbf{s}_i), Y(\mathbf{s}_j)) = c(\mathbf{s}_i, \mathbf{s}_j)$$

Advantages:

- Once you specify $m(.)$ and $c(.,.)$, the distribution of $Y(\cdot)$ is fully specified
- "Easy" interpretation of associations between outcome and covariates
- Spatial interpolation follows easily from properties of the multivariate normal distribution

Intrinsic Stationarity

It is defined through first differences:

$$\begin{aligned} \mathrm{E}(Y(\mathbf{s}+\mathbf{h}) - Y(\mathbf{s})) &= 0, \\ \mathrm{Var}(Y(\mathbf{s}+\mathbf{h}) - Y(\mathbf{s})) &= 2\gamma(\mathbf{h}) \end{aligned}$$

- The quantity $2\gamma(\mathbf{h})$ is known as the *variogram.*
- $\gamma(.)$ *is known as the semi-variogram.*
- *In geostatistics,* $2\gamma(.)$ *is treated as a parameter of the random process* $\{Y(\boldsymbol{s}) : \boldsymbol{s} \in D\}$

Second Order Stationarity

A random function $Y(.)$ satisfying:

$$\begin{aligned} \mathrm{E}(Y(\mathbf{s})) &= \mu \quad \forall \mathbf{s} \in D \\ \mathrm{Cov}(Y(\mathbf{s}), Y(\mathbf{s}')) &= c(\mathbf{s}-\mathbf{s}') \quad \forall \mathbf{s}, \mathbf{s}' \in D \end{aligned}$$

is defined to be second-order stationary. Furthermore, if $c(\mathbf{s}-\mathbf{s}')$ is a function only of $\| \mathbf{s}-\mathbf{s}' \|$ (it is not a function of the locations), then $C(.)$ is said to be isotropic.

A standard approach

In the analysis of most spatio-temporal processes in environmental studies, observations present skewed distributions, with a heavy right or left tail. Commonly it is assumed

- a transformation of the response variable to follow a GP (the same one at all sampling locations)
- the process to be stationary and isotropic $\rightarrow$ distribution is unchanged when the origin of the index set is translated, and the process is invariant under rotation about the origin

Is this a reasonable approach?

Effect of data transformation

- Wallin and Bolin (2015) point out that a transformation $Y(\mathbf{s}) = g(Z(\mathbf{s}))$ of the spatial process may induce dependence between the mean and covariance structures of $Z(\mathbf{s})$

- E.g.: $\sqrt{Z(\mathbf{s})}$, such that $Y(\mathbf{s}) = \mathbf{X}(\mathbf{s})'\beta + S(\mathbf{s}) + \varepsilon(\mathbf{s})$, then Wallin and Bolin (2015) show that

$$\begin{aligned} E(Z(\mathbf{s})) &= C_S(0) + 2\left(\mathbf{X}(\mathbf{s})'\beta\right)^2 \\ C(Z(\mathbf{s}), Z(\mathbf{s}')) &= 2C_S(\mathbf{s}-\mathbf{s}')^2 + 4(\mathbf{X}(\mathbf{s})\beta)(\mathbf{X}(\mathbf{s}')\beta)C_S(\mathbf{s}-\mathbf{s}') \end{aligned}$$

- Covariance structure depends on the mean structure $\Rightarrow$ nonstationary $\Rightarrow$ gets more complicated when we assume more complex models (nonlinear mean, nonstationary covariance...)

Maybe we should model the process in its original scale

Transformed Gaussian Random Fields
De Oliveira et. al (1997)

Let $\{Z(s), s \in D\}$, $D \subset \mathbb{R}^2$ be the random field of interest. They propose to model

$$\{Y(s) = g_\lambda(Z(s)),\ s \in D\}$$

where each $g_\lambda(\cdot)$ is a nonlinear monotone transformation, $g'_\lambda(\cdot)$ exists and is continuous in $\Lambda \times \mathbb{R}$ $\rightarrow$ a possible family is the Box-Cox family of power transformations

- Pros
  - inference procedure is performed under a single framework $\rightarrow$ λ is jointly estimated with other model parameters
  - spatial interpolation is performed integrating out the posterior distribution of the parameters
- Cons
  - Fall back into the problem mentioned by Wallin and Bolin (2015)
  - How to describe the estimated spatial correlation? And what about the association between covariates and the outcome?

Gaussian-log-Gaussian model - Palacios and Steel (2006)

$$Z(s_i) = \mathbf{w}(s_i)'\beta + \sigma \frac{\varepsilon(s_i)}{\sqrt{\lambda(s_i)}} + \tau\delta(s_i),$$

- $\varepsilon = (\varepsilon(s_1), \cdots, \varepsilon(s_n))' \sim N(0, C_\theta)$
- $\log(\lambda) = (\log(\lambda_1), \cdots, \log(\lambda_n))' \sim N\left(\frac{\nu}{2}\mathbf{1}, \nu C_\theta\right)$
  - $E(\lambda(s)) = 1$
  - $Var(\lambda(s)) = \exp(\nu) - 1$
  - Large values of ν induce thicker tails, and as $\nu \to 0$ Normal tails are retrieved
- $\delta(s_i) \sim N(0,1)$ i.i.d's

Properties:

- $Cov(d) = \sigma^2 C_\theta(d) \exp\left(\nu\left\{1 + \frac{1}{4}[C_\theta(d) - 1]\right\}\right)$
- small values of $\lambda(s)$ are regions of the space where the observations tend to be relatively far away from the estimated mean surface $\to$ spatial heteroskedasticity

Fonseca and Steel (Environmetrics and Biometrika 2011)

- Fonseca and Steel (Environmetrics, 2011) propose non-separable, stationary covariance functions through mixing over separable models
- Fonseca and Steel (Biometrika 2011) extend Palacios and Steel (2006) and Fonseca and Steel (2011) to construct non-Gaussian with non-separable covariance structures $\rightarrow$ allows to identify regions in space and time with large variances
- Zareifard and Kaledi (2013) combine the Unified Skew-Normal representation of Arellano-Valle and Azzalini (2006) with the model by Palacios and Steel to propose the unified skew Gaussian-log Gaussian (SUGLG) to account for skewness and heavy-tails
- Bueno, Fonseca, and Schmidt (under revision, 2017) allows ν to depend on covariates

Multivariate skew-normal distribution

Definition (Azzalini & Dalla Valle, 1996)

A random vector **H** follows a multivariate skew normal distribution if each element of **H** can be written as

$$H_l = \delta_l |\eta_0| + (1-\delta_l^2)^{1/2} e_l, \quad l = 1, \cdots, L$$

where $\eta_0 \sim N(0,1)$, $\boldsymbol{e} = (e_1, \cdots, e_L)$ is a random vector following a multivariate normal distribution with correlation matrix **M**, and with standard normal marginals, the elements of **e** are independent of η_0

Some references that use this distribution: Kim & Mallick(2004) and Frühwirth-Schnatter & Pyne(2010).

Multivariate skew-normal distribution

- Some care must be taken when using the multivariate skew-normal of Azzalini & Dalla Valle (1996) for spatial observations
  - Zhang & El-Shaarawi (2010) mention that, when the asymmetry is high, the spatial correlation between 2 locations gets close to 1 regardless of the distance between these two locations
  - When there is a single realization of the spatial process, Genton & Zhang (2012) mention that the parameters are not well identified, even if the number of locations is high. They also show that the model proposed by Zhang & El-Shaarawi (2010) avoids this problem

Zhang & El-Shaarawi (2010)

Let $Z(\boldsymbol{s})$ be a process defined in a continuous region, $\boldsymbol{s} \in \mathbb{G} \subset \mathbb{R}^p$. Assume

$$Z(\boldsymbol{s}) = m(\boldsymbol{s}) + \sigma|\eta(\boldsymbol{s})| + e(\boldsymbol{s})$$

where

- $\sigma \in \mathbb{R}$
- $\eta(\boldsymbol{s})$ is independent of $e(\boldsymbol{s}+\boldsymbol{r})$, $\forall\{\boldsymbol{s}, \boldsymbol{s}+\boldsymbol{r} \in \mathbb{G}\}$,
- $\{\eta(\boldsymbol{s}) \sim PG(0, 1, \rho_\eta(d)), \boldsymbol{s} \in \mathbb{D}\}$,
- $\{e(\boldsymbol{s}) \sim PG(0, 1, \tau I(d=0) + V\rho(d)), \boldsymbol{s} \in \mathbb{D}\}$

- It is also challenging to estimate models' parameters if there is only a single partial-realization of the process

Schmidt, Gonçalves and Velozo (2017)

Let $\{Z_t(\boldsymbol{s}); \boldsymbol{s} \in \mathbb{G}; t = 1, 2, \ldots\}$ be a stochastic process in discrete time t and location $\boldsymbol{s} \in \mathbb{G}$, with $\mathbb{G} \subset \mathbb{R}^p$, $p = 1$, 2, or 3. Let

$$Z_t(\boldsymbol{s}) = m_t(\boldsymbol{s}) + \sigma(\boldsymbol{s})|\eta_t(\boldsymbol{s})| + \sqrt{V_t}\omega_t(\boldsymbol{s}) + \sqrt{\tau}\varepsilon_t(\boldsymbol{s})$$

$$\begin{aligned} m_t(\boldsymbol{s}) &= \boldsymbol{X}_t(\boldsymbol{s})\boldsymbol{\theta}_t + \boldsymbol{A}_t(\boldsymbol{s})\boldsymbol{\gamma}, \\ \boldsymbol{\theta}_t &= \boldsymbol{G}_t\boldsymbol{\theta}_{t-1} + \boldsymbol{e}_t, \quad \boldsymbol{e}_t \sim N_K(0, \boldsymbol{W}) \end{aligned}$$

Schmidt, Gonçalves and Velozo (2017)

Let $\{Z_t(\mathbf{s}); \mathbf{s} \in \mathbb{G}; t = 1, 2, \ldots\}$ be a stochastic process in discrete time t and location $\mathbf{s} \in \mathbb{G}$, with $\mathbb{G} \subset \mathbb{R}^p$, $p = 1$, 2, or 3. Let

$$Z_t(\mathbf{s}) = m_t(\mathbf{s}) + \sigma(\mathbf{s})|\eta_t(\mathbf{s})| + \sqrt{V_t}\omega_t(\mathbf{s}) + \sqrt{\tau}\varepsilon_t(\mathbf{s})$$

- $\sigma(\mathbf{s}) \in \mathbb{R}$
- $\sigma(\mathbf{s}) = 0 \Rightarrow$ GP model
- *A priori*, we assume

$$\sigma(\mathbf{s}) \mid \mu_\sigma, V_\sigma \sim^{iid} N(\mu_\sigma, V_\sigma)$$

- $\mu_\sigma \sim N(m_{00}, C_{00})$, and $V_\sigma \sim \exp(C_v)$
- Then, marginally, the prior for $\sigma(\mathbf{s})$ has

$$E(\sigma(\mathbf{s})) = m_{00} \quad \text{and} \quad V(\sigma(\mathbf{s})) = 1/C_v + C_{00}$$

$$Cov(\sigma(\mathbf{s}), \sigma(\mathbf{s}')) = C_{00} \quad \text{and} \quad \rho_\sigma = Corr(\sigma(\mathbf{s}), \sigma(\mathbf{s}')) = \frac{C_{00}}{1/C_v + C_{00}}$$

- Fix C_v according to its 1-to-1 relationship with ρ_σ, as $C_v = \frac{\rho_\sigma}{C_{00}(1-\rho_\sigma)}$

Schmidt, Gonçalves and Velozo (2017)

Let $\{Z_t(\boldsymbol{s}); \boldsymbol{s} \in \mathbb{G}; t = 1, 2, \ldots\}$ be a stochastic process in discrete time t and location $\boldsymbol{s} \in \mathbb{G}$, with $\mathbb{G} \subset \mathbb{R}^p$, $p = 1$, 2, or 3. Let

$$Z_t(\boldsymbol{s}) = m_t(\boldsymbol{s}) + \sigma(\boldsymbol{s})|\eta_t(\boldsymbol{s})| + \sqrt{V_t}\omega_t(\boldsymbol{s}) + \sqrt{\tau}\varepsilon_t(\boldsymbol{s})$$

- $V, \tau > 0$ are scale parameters
- $\boldsymbol{\eta}_t$ and $\boldsymbol{\omega}_t$ are independent, zero mean GPs, with variance equals 1 and spatial correlation functions ρ_η and ρ, respectively
- $\varepsilon_t(\boldsymbol{s}) \sim N(0, 1)$ i.i.d.

Schmidt, Gonçalves and Velozo (2017)

Let $\{Z_t(\boldsymbol{s}); \boldsymbol{s} \in \mathbb{G}; t = 1, 2, \ldots\}$ be a stochastic process in discrete time t and location $\boldsymbol{s} \in \mathbb{G}$, with $\mathbb{G} \subset \mathbb{R}^p$, $p = 1$, 2, or 3. Let

$$Z_t(\boldsymbol{s}) = m_t(\boldsymbol{s}) + \sigma(\boldsymbol{s})|\eta_t(\boldsymbol{s})| + \sqrt{V_t}\omega_t(\boldsymbol{s}) + \sqrt{\tau}\varepsilon_t(\boldsymbol{s})$$

$$\log V_t = \log V_{t-1} + e_t^v, \quad e_t^v \sim N(0, V_v)$$

Resultant Covariance Structure

Let $W_t(\boldsymbol{s}) = \sigma(\boldsymbol{s})|\eta_t(\boldsymbol{s})| + \sqrt{V_t}\omega_t(\boldsymbol{s}) + \sqrt{\tau}\varepsilon_t(\boldsymbol{s})$, $\forall \boldsymbol{s} \in \mathbb{G}$, then

If $\sigma(\boldsymbol{s})$ and V_t

$$\text{cov}(W_t(\boldsymbol{s}), W_{t'}(\boldsymbol{s}')) = \begin{cases} \frac{2}{\pi}\sigma(\boldsymbol{s})\sigma(\boldsymbol{s}')\left(\sqrt{1-\rho_\eta^2(d)} + \rho_\eta(d)\arcsin(\rho_\eta(d)) - 1\right) + V_t\rho(d) + \tau I(d), & t = t' \\ 0, & t \neq t' \end{cases}$$

$d = ||\boldsymbol{s} - \boldsymbol{s}'||$, and $I(d)$ is an indicator function, such that $I(d) = 1$, if $d = 0$.

- Pros:
  - Parameter for asymmetry $\alpha_t(\boldsymbol{s}) = \frac{\sigma(\boldsymbol{s})}{V_t + \tau} \to$ process at different locations might have different distributions
  - GP is a particular case
  - Nonstationary covariance structure
- Cons:
  - There are two spatial processes $\to$ two covariance matrices
  - Estimated spatial correlation seems smoother than the empirical one
  - How much skewness can it handle?

Xu and Genton (2017)

- Spatial process based on Tukey's g-and-h transformation, that assumes

$$\tau_{g,h}(z) = g^{-1}\{\exp(gz) - 1\}\exp(hz^2/2)$$

- Let $Z(\cdot)$ be a standard Gaussian random field with some correlation function $\rho_Z(||h||)$, and

$$T(\mathbf{s}) = \tau_{g,h}(Z(\mathbf{s}))$$

they define a general random field $Y(\mathbf{s})$ as

$$Y(\mathbf{s}) = \beta_0 + \mathbf{X}(\mathbf{s})\beta + \omega T(\mathbf{s})$$

$\omega > 0$ is a scale parameter, $\beta \in \mathbb{R}^p$

Xu and Genton (2017)

Properties:

- $E(T(\mathbf{s}))$ is a function of g and h, and $C_T(\mathbf{s},\mathbf{s}')$ is a function of $\rho_Z(\mathbf{s},\mathbf{s}')$, g, h and $E(T(\mathbf{s}))$
- If $g = h = 0$ $Y(\mathbf{s})$ reduces to a Gaussian random field
- For $h = 0$ and $g > 0$ $Y(\mathbf{s})$ is a shifted log-Gaussian random field
- For $g = 0$ and $h > 0$ $Y(\mathbf{s})$ becomes a Pareto-like marginal random field
- It can model left-skewed data with $g < 0$
- If $h < 1/2$ and $Z(\mathbf{s})$ is a second-order stationary, then $T(\mathbf{s})$ is also second-order stationary; (b) $T(\mathbf{s})$ is mean-square continuous iff $Z(\mathbf{s})$ is; (c) $T(\mathbf{s})$ is m-times mean-square differentiable if $Z(\mathbf{s})$ is
- Inference is via MLE, and if measurement error is present it only allows inference for the model

$$Y(\mathbf{s}) = \beta_0 + \mathbf{X}(\mathbf{s})\beta + \omega\,\tau_{g,h}(V(\mathbf{s}) + \varepsilon(\mathbf{s}))$$

Bolin (2014) and Wallin and Bolin (2015)

- Based on the fact that a Matérn field $X(s)$ can be viewed as a solution to the SPDE

$$(\kappa^2 - \Delta)^{\alpha/2} X = \dot{M}$$

  The Gaussian Matérn fields are recovered by choosing $\dot{M}$ as a Gaussian white noise scaled by a variance parameter ϕ
- Bolin (2014) extends this for the case where $\dot{M}$ is non-Gaussian, focused on SPDE's driven by generalized asymmetric Laplace noise
- Proposed model

$$\begin{aligned} \mathbf{y} &= \mathbf{B}\beta + \mathbf{A}\mathbf{w} + \varepsilon \\ \mathbf{w} &= \mathbf{K}_{\alpha}^{-1}\left(\tau \mathbf{B}_{\gamma}\gamma + \mathbf{I}_{v}\mathbf{B}_{\mu}\mu + \sigma\sqrt{\mathbf{V}}\mathbf{Z}\right) \end{aligned}$$

- Xu and Genton (2017) mention that "Although this provides an interesting alternative, their approach is mathematically involved and its statistical properties are much less understood than the transGaussian random field."

Discussion

- Exploratory Data Analysis
- Interpretation of fixed effects under different distributions
- Are we able to estimate parameters from these complex models from a single partial realization of the field?
- Approaches based on Copulas
- Random effects in Generalized Spatial Models: should we explore non-Gaussian random effects?
- Big data

Acknowledgements

Thanks to

- Peter Craigmile and co-organizers for organizing this workshop and for the invitation
- CNPq, CAPES, and FAPERJ, Brazil, and NSERC, Canada for the financial support

Acknowledgements

Thanks to

- Peter Craigmile and co-organizers for organizing this workshop and for the invitation
- CNPq, CAPES, and FAPERJ, Brazil, and NSERC, Canada for the financial support
- (Last, but not least) Peter Guttorp for the great collaborations and advises!

Acknowledgements

Thanks to

- Peter Craigmile and co-organizers for organizing this workshop and for the invitation
- CNPq, CAPES, and FAPERJ, Brazil, and NSERC, Canada for the financial support
- (Last, but not least) Peter Guttorp for the great collaborations and advises!

Muito obrigada!