A spatio-temporal stick-breaking process

Clara Grazian

School of Mathematics and Statistics, University of Sydney

clara.grazian@sydney.edu.au

Bayes on the Beach

Gold Coast

06 February 2024

Introduction

Spatial data are becoming more and more popular!

With the term “spatial data”, we mean any type of data that directly or indirectly references a specific geographical area or location.

Spatial data is data that describes a surface, usually through variables on a two-dimensional plane.

Spatial Modelling

Consider a random field $\{Y(s) : s \in \mathscr{D}\}$
for $\mathscr{D} \subseteq \mathbb{R}^2$,
and let $s = (s_1, \ldots, s_n)$ be a set of distinct locations in $\mathscr{D}$, such that $s_i = (s_{i,1}, s_{i,2})$ for $i = 1, \ldots, n$.

$y(s)$ is characterized by

$$y(s) = \theta(s) + x(s)^T \beta + \varepsilon(s)$$

- $\theta(s)$ is a spatial effect
- $x(s)$ is a column-vector of p covariates
- β is a column-vector of p regression coefficients
- $\varepsilon(s) \sim \mathscr{N}(0, \sigma_\varepsilon^2)$

In the following, we assume covariates are not available, then

$$y(s) = \theta(s) + \varepsilon(s)$$

Kriging

A common approach to problems of modelling surfaces is kriging [Stein, 2012], which interpolates and smooths spatial data.

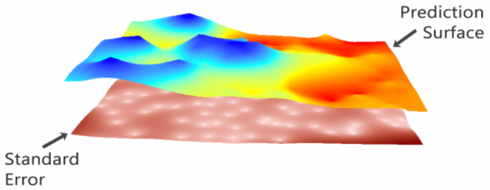


Figure: Courtesy of `gisgeography.com`

Assumptions:

- Gaussianity
- stationarity
- data cannot have trends

Kriging

In kriging, prediction $y(s^*)$ at a spatial location s^* is obtained as a weighted average, where observations at closer locations are given higher weights than observations at distant locations.

$$\theta(s) \sim N(\mu, \Sigma)$$

where $\Sigma_{ij} = \mathbb{C}\text{ov}(\theta(s_i), \theta(s_j))$ [Cressie (2015)]

Spatio-temporal processes

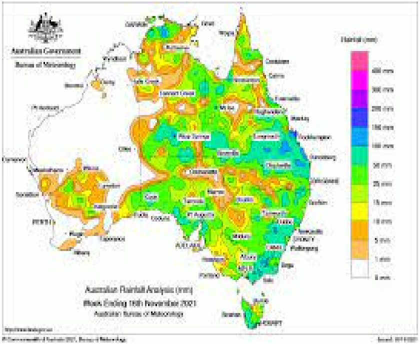
The concept of spatial process can be extended to the spatio-temporal case

$$y(s,t) = \{y(s,t), (s,t) \in \mathscr{D} \times \mathscr{T} \subset \mathbb{R}^d \times \mathbb{R}\}$$

The kriging model can be extended by defining a spatio-temporal covariance function given as $\mathbb{C}\text{ov}(\theta(s,t), \theta(s',t'))$ [Gelfand et al. (2010)],[Cressie and Wikle (2015)].
Often space and time are considered separable [Finkenstadt et al. (2006)]:

$$\mathbb{C}\text{ov}(\theta(s,t), \theta(s',t')) = \sigma^2 C_1(d(s,s')) C_2(|t-t'|)$$

Examples of non-separable covariances: [Cressie and Huang (1999)],[Gneiting (2002)]



A first relaxation of the Gaussianity assumption may be represented by finite mixture models; see [Fernandéz and Green (2002), Neelon et al. (2014) and Hossain et al. (2008)]

However, finite mixture models assume it is possible to consistently select the number of components, which may be difficult in presence of covariates.

Let's go nonparametric!

A random distribution F is distributed according to a Dirichlet process (DP) if its marginal distributions are Dirichlet distributed [Ferguson, 1973].

Taking F_0 to be a distribution on Θ, and $\alpha > 0$, consider $A_1, \dots, A_r \in \Theta$; then $F \sim DP(\alpha, F_0)$, if

$$(F(A_1), \dots, F(A_r)) \sim Dir(\alpha F_0(A_1), \dots, \alpha F_0(A_r)).$$

Sethuraman's representation

[Sethuraman, 1994] shows that $F \sim DP(\alpha, F_0)$ if

$$V_k \sim Beta(1,\alpha) \qquad k = 1, \ldots$$
$$\theta_k \sim F_0$$
$$\pi_1 = V_1 \qquad \text{and} \qquad \pi_k = V_k \prod_{j=1}^{k-1} (1 - V_j) \quad \text{for } k = 2, \ldots$$
$$F = \sum_{k=1}^{\infty} \pi_k \delta_{\theta_k},$$

where

- $\delta(\theta)$ is a Dirac mass
- V_k and θ_k are independent for $k = 1, 2, \ldots$

This can be generalised to $V_k \sim Beta(a_k, b_k)$.

Spatio-Temporal SB

Suppose $y(s,t)$ can be represented as an infinite mixture model:

$$g(y(s,t)) = \sum_{k=1}^{\infty} \pi_k(s,t) g(y(s,t)|\theta_k)$$

where the mixing probability $\pi_k(s,t)$ is the probability that the observation at location s and time t belongs to component k.

The mixing weights are built similarly to the stick-breaking process:

$$F_{s,t} = \sum_{k=1}^{\infty} \pi_k(s,t)\delta_{\theta_k} \qquad s \in \mathscr{D}, t \in \mathscr{T} \qquad \text{where}$$

$$\pi_1(s,t) = V_1(s,t) \qquad \text{and} \qquad \pi_k(s,t) = V_k(s,t)\prod_{j=1}^{k-1}(1 - V_j(s,t))$$

$$V_k(s,t) = w_k(s, \psi_k, t, \zeta_k)V_k$$

$$V_k \sim Beta(a,b) \qquad \perp \qquad \theta_k \sim F_0 \qquad \text{for } k = 1,2,\ldots$$

Conditional covariance

The covariance between $y(s,t)$ and $y(s',t')$ conditional on the mixing weights but marginalised with respect to the atoms is

$$\mathbb{C}\text{ov}(y(s,t),y(s',t')|V,\psi,\zeta) = \sigma_\theta^2 \sum_{k=1}^{\infty} \left[w_k(s,\psi_k,t,\zeta_k) w_k(s',\psi_k,t',\zeta_k) V_k^2 \right.$$
$$\cdot \prod_{j<k} [1 - (w_j(s,\psi_j,t,\zeta_j) + w_j(s',\psi_j,t',\zeta_j)) V_j +$$
$$\left. + w_j(s,\psi_j,t,\zeta_j) w_j(s',\psi_j,t',\zeta_j) V_j^2] \right]$$

The covariance function

- depends on the choice of the kernels
- does not depend on the value of $y(s,t)$ and $y(s',t')$
- can be stationary or not

Marginal covariance

If we use $V_k \sim Be(a, b)$ for all k, then $\mathbb{E}[V_k]$ and $\mathbb{E}[V_k^2]$ have a fixed form, depending on a and b, and

$$\mathbb{C}\text{ov}(y(s,t), y(s',t')) = \sigma_\theta^2 \frac{h(s,s',t,t')}{2\left(1 + \frac{b}{a+1} - h(s,s',t,t')\right)}$$

where $h(s,s',t,t')$ depends on the expectation of

- $w_k(s, \psi_k, t, \zeta_k)$
- $w_k(s', \psi_k, t', \zeta_k)$
- $w_k(s, \psi_k, t, \zeta_k) w_k(s', \psi_k, t', \zeta_k)$

Separable kernels

It is possible to consider separable kernels:

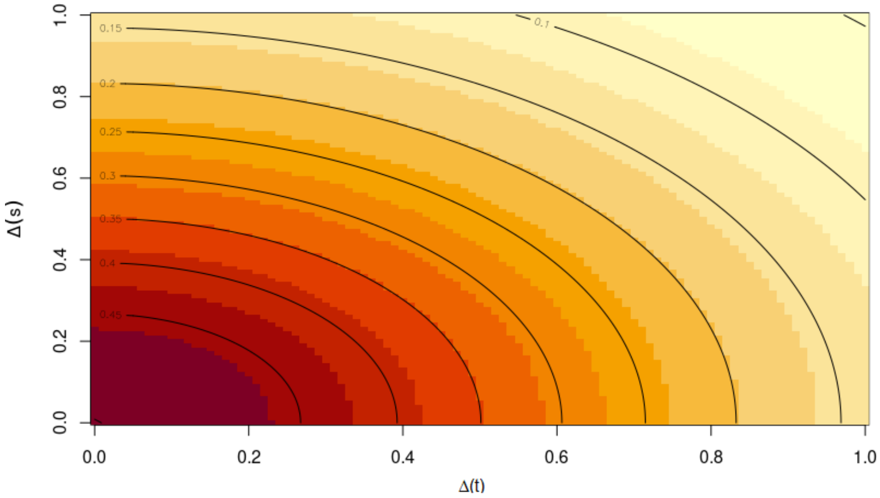
$$w_k(s,t,\psi,\zeta) = w_k(s,\psi)\cdot w_k(t,\zeta)$$

where ψ is a set of knots for the spatial domain and ζ is a set of knots for the temporal domain.
Then

$$h(s,s',t,t') = \frac{\int_\Psi w(s,\psi)w(s',\psi)p(\psi)d\psi}{\int_\Psi \int w(s,\psi)p(\psi)d\psi}\cdot\frac{\int_Z w(t,\zeta)w(t',\zeta)p(\zeta)d\zeta}{\int_Z w(t,\zeta)p(\zeta)d\zeta}.$$

While the kernel is separable, the covariance structure of the process is not separable.

Separable kernels



Non-separable kernels

Consider a non-separable kernel [Gneiting (2002)]

$$w(s,\psi,t,\zeta) = \frac{1}{\gamma|t-\zeta|+1}\exp\left(-\frac{(s_1-\psi_1)^2+(s_2-\psi_2)^2}{(\gamma|t-\zeta|+1)^{\lambda/2}}\right),$$

- γ is a nonnegative scaling parameter
- $\lambda \in [0,1]$ is a parameter controlling the space-time interaction
- if $\lambda = 0$, the kernel is separable; as $\lambda \to 1$, the space-time interaction increases and the spatial dependence at positive lags decreases more slowly

Non-separable kernels

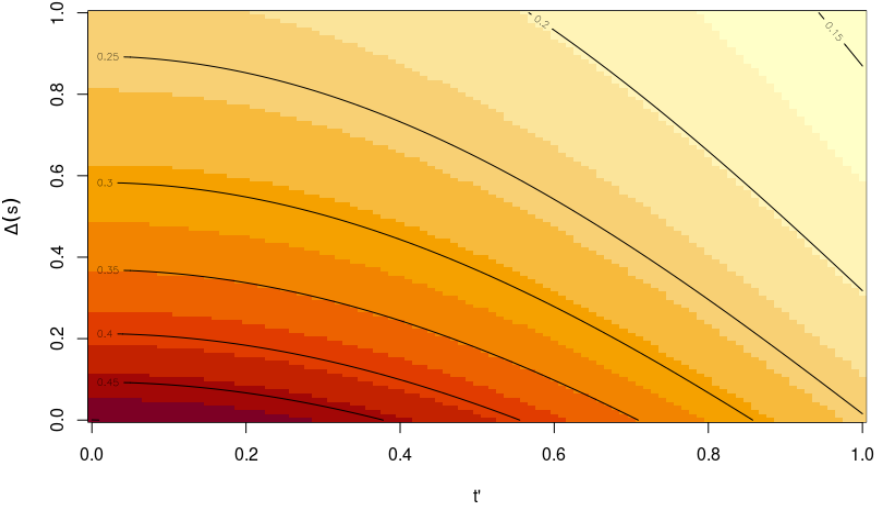
In this case

$$h(s,s',t,t') = \frac{\int_\Psi \int_\Xi w_k(s,\psi_k,t,\zeta_k) w_k(s',\psi_k,t',\zeta_k) p(\psi,\zeta) d\psi_k d\zeta_k}{\int_\Psi \int_\Xi w_k(s,\psi_k,t,\zeta_k) p(\psi,\zeta) d\psi_k d\zeta_k},$$

when considering γ and λ fixed.
In the specific case of the previous kernel, even when $\lambda = 0$

$$\begin{aligned} h(s,s',t,t') = \exp\left(-((s_1-s_1')^2 + (s_2-s_2')^2)\right) \\ \cdot \left(-\frac{\log(t') - \log(t'-1) - \log(t) - \log(t-1)}{\gamma^2(t-t')}\right), \end{aligned}$$

Non-separable kernels

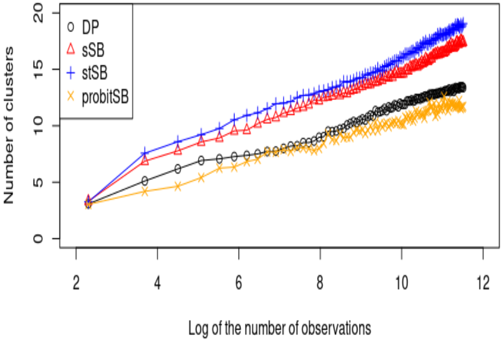


Testing for separability

The non-separable kernels proposed before allows for an automatic testing on the λ parameter, assuming a **spike-and-slab prior** [Mitchell and Beauchamp (1988), Andersen et al. (2017)]

$$\lambda \sim p_\lambda f(\lambda) + (1 - p_\lambda)\delta_0,$$

Clustering properties



Allowing atoms to depends on space and time

Atoms may also be allowed to vary by introducing a modification of the stick-breaking representation of each element

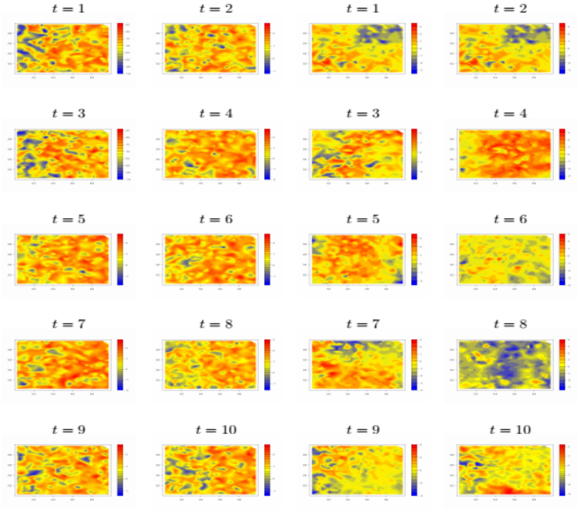
$$F_{s,t} = \sum_{k=1}^{\infty} \pi_k(s,t)\delta_{\theta_k(s,t)}$$

where $\theta_k(s,t)$ are independent stochastic processes with index set $\mathscr{S} \times \mathscr{T}$ and $F_{s,t,0}$ marginal distributions.
The atoms $\theta_k(s,t)$ can be generated from a Gaussian process with covariance

$$C(d(s,s'), d(t,t_0)) = C_s(d(s,s')) \times C_t(d(t,t_0)),$$

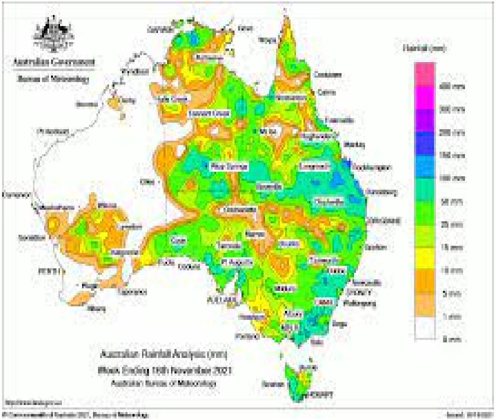
where C_s is a spatial covariance function and C_t is a temporal covariance function.

Example of spatio-temporal SB process



Modelling Rainfalls

Rainfalls in Australia are known to be highly variable [Nicholls et al. (1997)]

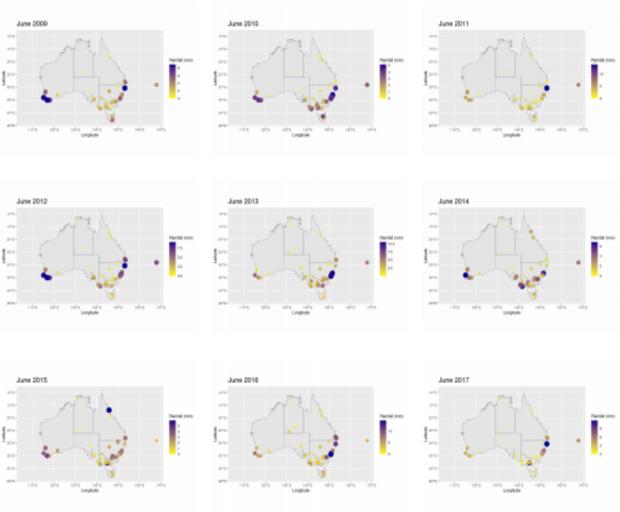


The dataset

The dataset contains

- 10 years of daily weather observations, from 31 October 2007 to 24 June 2017
- 49 locations across all Australia
- observations were drawn from weather stations
- the amount of rainfall recorded for the day in mm
- publicly available on the website of the Bureau of Meteorology of the Australian Government
- 140,793 data points
- **training dataset:** data from 2007 to 2016 (132,420 observations)
- **testing dataset:** data for 2017 (8,373 observations)

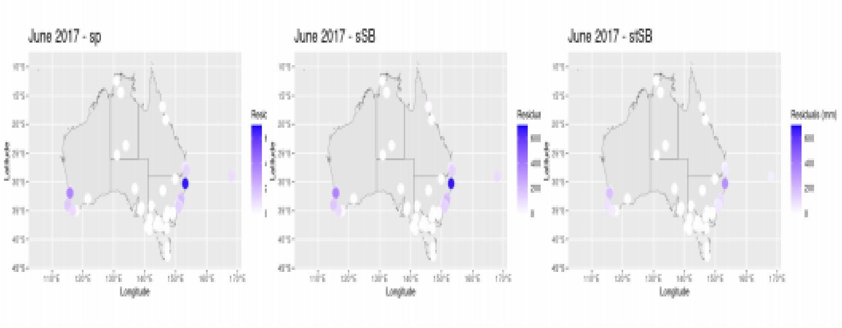
Modelling Rainfalls



Comparison between models

	ESPE	**DIC**	**WAIC**
sp	109.74	39,960,564	3,109,830
sSB	115.47	495,345	524,083
stSB	71.05	233,070	257,581

Prediction



June 2017 - sp
June 2017 - sSB
June 2017 - stSB
Longitude
Residuals (mm)

Conclusions

Modelling spatio-temporal phenomena is an important and challenging problem.

- the spatio-temporal SB avoids oversmoothing
- it can model the interaction between space and time and incorporate both spatial and temporal dependence in the definition of the stick-breaking weights.
- this allows to model nonstationary and non-normal models over time and space.
- this model avoids the need for space-time varying covariates as in [Hossain et al. (2013)] to model the weight components
- extensions to consider spatio-temporal atoms are straightforward

Where is seasonality?

Thank you!

Grazian, C. (2023+) "Spatio-temporal stick-breaking" (*work in progress*)

Grazian, C. (2023+) "A review on Bayesian model-based clustering" (*work in progress*)

Grazian, C. `https://github.com/bayesgra/spatiotempoSB`