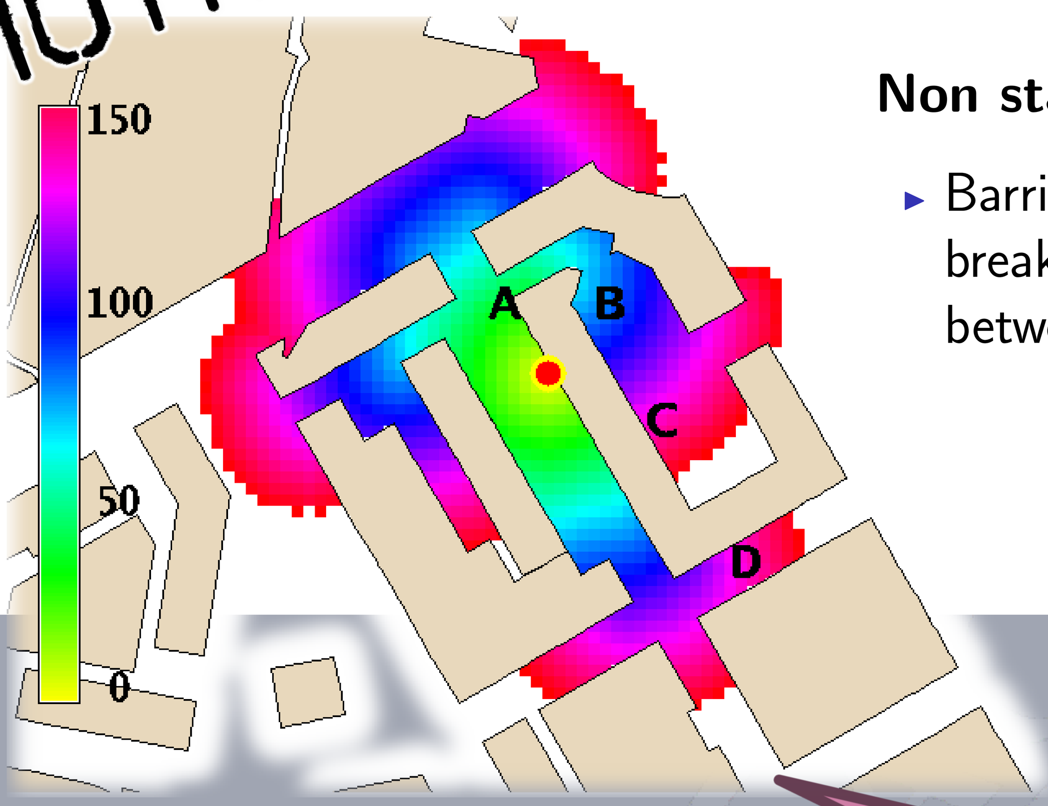
Geostatistical modelling with non-Euclidean distances

Facundo Muñoz Antonio López-Quílez

Universitat de València, Spain

MOTIVATION

Irregular locations



Non stationarity

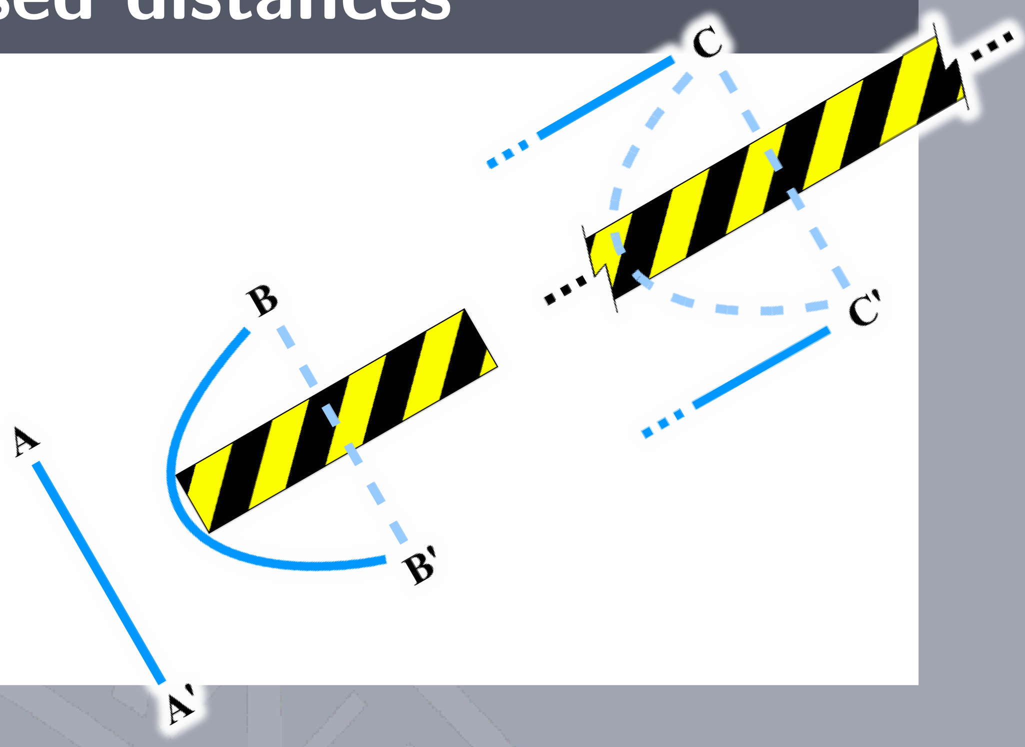
- Barriers or other irregularities break the functional relationship between correlation and distance

PROPOSAL

Cost-based distances

Stationarity w.r.t. cost-based distance

- Build cost surface $\mathfrak{c}$ from geographical characteristics
- Compute minimum-cost paths
- Set covariance model as a function of cost-based distance



PROBLEM

Positive-definiteness

Choose the covariance model C such that

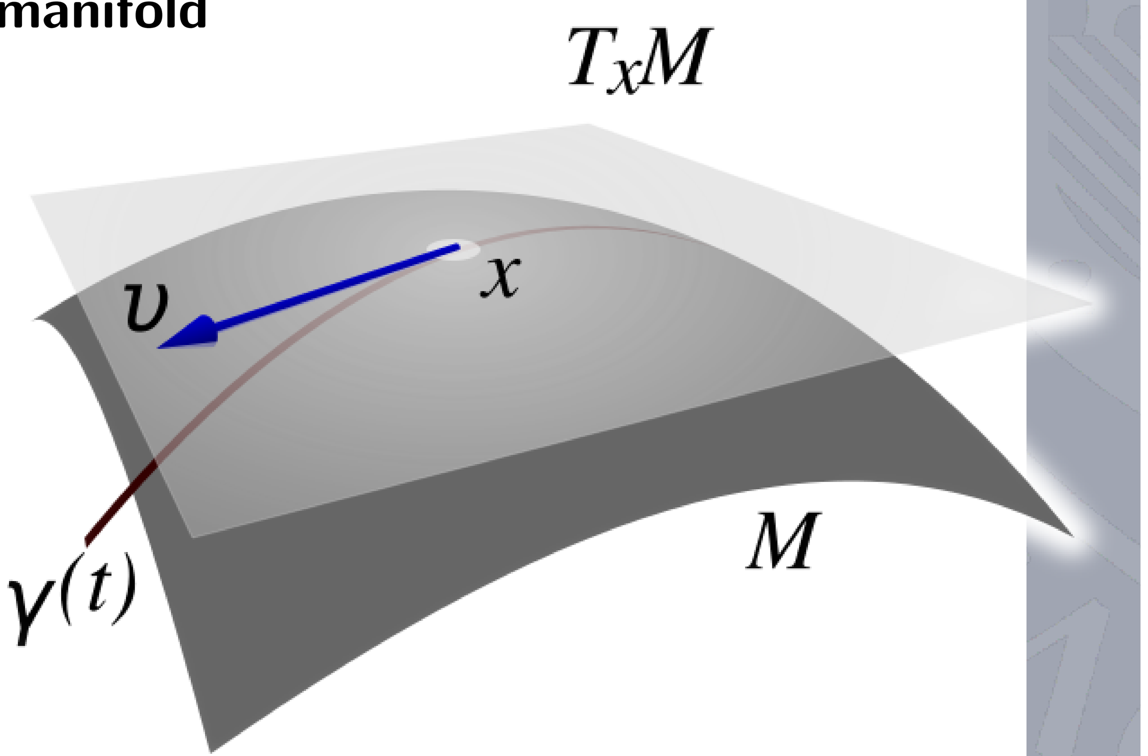
$$\left.\begin{array}{r}\forall \text{ locations } \boldsymbol{s}_1, \dots, \boldsymbol{s}_n \in \text{Region},\\ \forall \text{ scalars } \alpha_1, \dots, \alpha_n \in \mathbb{C},\end{array}\right\} \quad \sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \bar{\alpha}_j C(d_{\text{cb}}(\boldsymbol{s}_i, \boldsymbol{s}_j)) \geq 0,$$

where d_{cb} is the cost-based distance between its arguments.

Open lines of work

Riemannian Manifolds

- **Consider the region M as a Riemannian manifold**



- **Define the Riemannian metric as**

$$g_{\boldsymbol{s}}(u, v) = \mathfrak{c}(\boldsymbol{s})^2 \langle u, v \rangle \quad \forall u, v \in T_{\boldsymbol{x}}M$$

- **Metric induced**

$$\tau_g(\boldsymbol{s}, \boldsymbol{t}) = \inf \{\text{lengths of the curves connecting } \boldsymbol{s} \text{ and } \boldsymbol{t}\}$$

- **Characterise the family of positive-definite functions over M**
  In an analogous way to Bochner's and Shoenberg's theorems, this involves developing Fourier and spectral analysis in this (much) more general context, in order to compute transforms of positive measures and to integrate them out over the surfaces of constant radius.

Pseudo-Euclidean spaces

- **Definition**
  A pseudo-Euclidean space is a vector space of dimension d, say $\mathbb{R}^d$, with a non-degenerate symmetric bilinear form

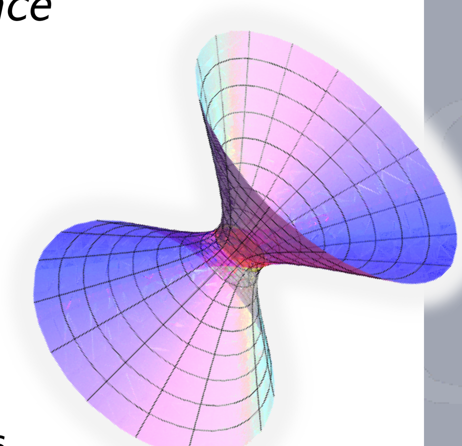
$$(\cdot,\cdot) : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$$
$$(\boldsymbol{x}, \boldsymbol{y}) = (x_1y_1 + \cdots + x_ky_k) - (x_{k+1}y_{k+1} + \cdots + x_dy_d),$$

  where k is called the *index*, while the pair $(k, d-k)$ is called the *signature* of the space. The space is denoted $E_{(k,d-k)}$.

- **Results**
  The original locations, together with their cost-based distances can be **exactly** *represented in a pseudo-Euclidean space.*

  The Bochner's theorem is still valid in the pseudo-Euclidean space

- **Too big**
  - The constant-radius surface turns into a hyperboloid, causing integration to diverge.
  - The pseudo-Euclidean space is able to represent any set of *dissimilarities*. But this is unnecessary, since the cost-based distance is a (full) *metric*.
  - The family of positive definite functions (which includes the trivial constant function 1) is a subset of those in the space M,



$$1 \in \mathfrak{P}(E_{(k,d-k)}) \subset \mathfrak{P}(M).$$

Bayesian Simulation

- **Model**

$$\boldsymbol{s}_1, \dots, \boldsymbol{s}_n \rightsquigarrow \mathbf{D}_{\text{cb}} = (r_{ij}); \quad \begin{cases} r_{ii} = 0 \\ r_{ij} \geq 0 \\ r_{ij} = r_{ji} \end{cases}$$
$$y_1, \dots, y_n \rightsquigarrow \boldsymbol{y} \sim \mathcal{N}(\boldsymbol{\mu}, \tau^2 \mathbf{I})$$
$$\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\omega}$$
$$\boldsymbol{\omega} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{P})$$
$$\mathbf{P} = f(\mathbf{D}_{\text{cb}})$$
$$f \sim \cdots$$

  where f is a random function satisfying $f(0) = 1$, $|f(r)| \leq 1$ and most importantly, the (correlation) matrix resulting from the element-wise transformation of the (cost-based) distance matrix must be positive definite.

- **Simulate f from a given family of functions**
  Maybe a non-paramteric family, honouring the restrictions, and hopefully positive-definite, most times.

- **Accept-reject**
  Check the positive-definiteness condition

- **Open questions**
  The procedure lacks theoretical foundation. The positive-definiteness of the covariance function is not guaranteed.