Spatial statistical modeling and applications

Sophie Dabo-Niang

Professor of Applied Mathematics, University of Lille, Lab Painleve and INRIA-MODAL, Chair of EMS-CDC
sophie.dabo@univ-lille.fr

AWMA Virtual meeting, January 7th, 2021
Outline

- Introduction to Spatial Data Science
- Spatial prediction: Kriging
- Space-time/Functional Prediction (Kriging)
- Open questions
Introduction to Spatial Data Science
Spatial Statistics from South Africa

'50  Danie Krige (mining engineer from South Africa) : 1919-2013

'60  Mathematical Foundation at MinesParisTech by G. Matheron
     (1930-2000)

- Firstly mining and petroleum applications
- Nowadays mostly data are spatial or spatio-temporal
Spatial Data
Covid-19 Cumulative cases World Map
Spatial data

are collected from different spatial locations (area, country, county, town, ...) on the earth, as in a variety of fields, including soil science, geology, oceanography, econometrics, epidemiology, environmental science, ...

- temperature or precipitation in given locations and impact in neighborhoods
- housing prices in one area and impacts on nearby houses
- disease spreading from one location to another
- migration of population from one region to another
- distribution of species in ecology
- soils properties, ...

**Spatial statistics** includes (statistical) techniques which study phenomenons observed on spatial sets: observations are **spatially dependent**.
Let $S$ a spatial set, $S \subset \mathbb{R}^N$, $N \geq 2$.

In particular, take $N = 2$ and set a location $s \in \mathbb{R}^2$ as 
(latitude, longitude)

Spatial data can be modeled as realization of random fields : $Z = \{Z_s, s \in S\}$, a collection of random variables indexed in $S$. 
Three types:

- **Lattices data**: spatial networks (neighbors locations are connected), generalization of time-series processes
- **Point spatial processes**: Spatial location are random, generalization of time-point processes
- **Geostatistical data**: data located at locations in $\mathbb{R}^N$, $N > 1$

The distinction between these types is not always simple, see e.g., Cressie (1995)

Data are more and more collected in space an time, see e.g, Cressie et Wikle (2011)
**Goal**: determine fish stocks in order to avoid fishing some species or in some areas, impact of climate change on stocks

- fish (different species) abundance with environmental parameters
- observations at some locations
- prediction in other locations

Map of the locations at which data are known
Massive Environmental and Oceanology Scan Fishery data in west Africa
Goal: determine contaminated soils in order to avoid farming in some areas.

- region of 14.5 km²
- concentration on Cadmium, Cuivre, Plomb,...
- 259 sites used as observations
- 100 sites to predict

Data collected by Ecole Polytechnique Fédérale de Lausanne (Suisse).

See, e.g. Goovaerts (1997) (geostatistic)
Malaria-infected erythrocytes
Disease Mapping in time and space (lattice or point process)
Spatial prediction : Kriging

- Spatial variability
- Interpolation : prediction
Kriging and Fields of applications

- Environmental Science
- Hydrogeology
- Natural Resources
- Remote Sensing
- Mining

...and more!
Main Problem: Kriging

- **A model and hypotheses**: let $Z = \{ Z_s \in \mathbb{R}, s \in \mathbb{R}^N \}$ be a real *gaussian* spatial process that we suppose *second order or intrinsic stationary*. $Z = (Z_s, s \in \mathbb{R}^N)$ is observed on $n$ sites $s_1, \ldots, s_n$.

- **Prediction of** $Z_{s_0}$ (or $g(Z_{s_0})$) **using** $Z_{s_1}, Z_{s_2}, \ldots, Z_{s_n}$

- **With the help of a covariance or Variogram model**: $\gamma(h)$ (or covariance function $C(h)$) supposed known:
  $$\gamma(h) = \frac{1}{2} \text{Var}(Z_{s+h} - Z_s)$$
Example of variogram

**Exponential:**

\[ \gamma(h) = c_0 \cdot \left(1 - \exp\left(-\frac{\|h\|}{a}\right)\right), \quad \text{for } h \neq 0, \quad (1) \]

The empirical estimator of the variogram is given by:

\[ 2\hat{\gamma}_n(h) = \frac{1}{\#N(h)} \sum_{s_i, s_j \in N(h)} (Z_{s_i} - Z_{s_j})^2, \quad h \in \mathbb{R}^N, \quad (2) \]

where \( N(h) = \{(s_i, s_j) : s_i - s_j = h; i, j = 1, \ldots, n\} \) is the set of pairs of observations for the distance \( h \), \( \#N(h) \) is the number of distinct pairs in \( N(h) \), \( h \) is the distance between two locations.
**Figure 1** – Experimental and theoretical variograms. The nugget is the variance between very close observations. The range indicates the distance after which observations are independent. At this distance, there is maximal variance which corresponds to the sill.
Prediction of $Z_{s_0}$

**Prediction (Kriging)** $Z_{s_0}$ is to find the best linear predictor (BLUP) $p(Z, s_0)$ of $Z_{s_0}$ with observations $Z = (Z_{s_1}, \ldots, Z_{s_n})$. 
Let \( \{ Z_s, s \in \mathbb{R}^N \} \) be a spatial process.

**The optimal predictor at a location** \( s_0 \) **minimizes**

\[
E(L(Z_{s_0}, p(Z, s_0))).
\]

Let the loss function be \( L(Z_{s_0}, p(Z, s_0)) \).

If

\[
L(Z_{s_0}, p(Z, s_0)) = (Z_{s_0} - p(Z, s_0))^2,
\]

then the optimal predictor is

\[
p^{opt} = E(Z_{s_0} | Z).
\]
Ordinary Kriging

Hypotheses: $Z_s = \mu + \epsilon_s$, $\{\epsilon_s\}$ is centered, intrinsic, $\mu$ unknown.

The predictor is $\hat{Z}_{s_0} = \sum_{i=1}^{n} \lambda_i Z_{s_i}$, with

$$E(\hat{Z}_{s_0}) = E(Z_{s_0}) \text{ and } E(\hat{Z}_{s_0} - Z_{s_0})^2 \text{ minimum}$$

The $(\lambda_i)_{i=1,n}$ are solutions of the system ($m$ is a Lagrange multiplier)

$$\begin{pmatrix}
0 & \gamma(s_1 - s_2) & \ldots & \gamma(s_1 - s_n) & 1 \\
\gamma(s_1 - s_2) & 0 & \ldots & \gamma(s_2 - s_n) & 1 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
\gamma(s_1 - s_n) & \gamma(s_2 - s_n) & \ldots & 0 & 1 \\
1 & 1 & \ldots & 1 & 0
\end{pmatrix}
\begin{pmatrix}
\lambda_1 \\
\lambda_2 \\
\vdots \\
\lambda_n \\
m
\end{pmatrix}
= 
\begin{pmatrix}
\gamma(s_0 - s_1) \\
\gamma(s_0 - s_2) \\
\vdots \\
\gamma(s_0 - s_n) \\
1
\end{pmatrix}$$
**Ordinary Kriging**

**Kriging Variance**

\[
\sigma_{OK}^2(s_0) = E((\hat{Z}_{s_0} - Z_{s_0})^2) = m + \sum_{i=1}^{n} \lambda_i \gamma(s_i - s_0)
\]

**Remarks**

- We don’t necessarily have \(0 \leq \lambda_i \leq 1\)
- If \(s_0 \in \{s_1, \ldots, s_n\}\) then \(\lambda_i = 1, \lambda_j = 0, j \neq i\), and \(\sigma_{K}^2(s_i) = 0\)
- The kriging weights depend on the layout of the data sites, the location of the prediction site, the number of data and the variogram function.
Extensive number of ressources: published papers, scientific events, journals, societies, ...

More than 100 free R, Python, Matlab,..., softwares

Website of Luc Anselin:
https://spatial.uchicago.edu/software

Matlab Ressources (LeSage):
https://www.spatial-econometrics.com/
Space-time/Functional Prediction (Kriging) : Application
Functional data at different locations: spatial global radiation

- 106 monitoring stations of US in 2015
- Ozone recorded hourly from July 19 at 12 am to July 20 at 11 pm
The most important data in Hydrology is the daily streamflow curve.

Streamflow data during a year or a season at a given location is named hydrograph: temporal evolution of the flow.
A random variable $X$ is called functional variable if it takes values in an functional space (usually of infinite dimensional space).

$$X = \{X_t : t \in T\}$$

$$X_t : \Omega \to \mathbb{R}$$

- If $T \subseteq \mathbb{R}$ then $X$ is a curve
- If $T \subseteq \mathbb{R}^2$ then $X$ is an image

However, raw data (i.e. what we observe) is always discrete

**Application**: biometrics, chemometrics, economics, medicine, electricity consumption, climate, environment, hydrology,...
Functional data literature

**Change-point**  Berkes et al. (2009), Hörmann and Kokoszka (2010); Aston and Kirch (2011); ...

**Classification**  Dabo et al. (2007); Delaigle et Hall (2012); ...

**Outlier**  Hyndman and Shang (2010); Sawant et al. (2012); ...

**PCA**  Ramsay and Silverman (2005); Biau and Mas (2012); Shang (2014); ...

**Regression**  Masry (2005); Dabo and Rhomari (2009), Delsol (2009); Ferraty et al. (2012); Kokoszka and Reimherr (2013); Aue et al. (2014); Dabo-Niang et al. (2017), ...


**Review**  See also the "Review of Functional data Analysis" of Wang et al. (2015), Martinez-Hernandez and Genton (2020)
Modeling spatial functional (massive) data

- Modelization of functional data basically focuses on independent data.
- In many applied domains, data are spatially correlated functions: economic, environmental, hydrology, ...
  
  **Example**: curves of daily concentration of ozone at two near stations
- Some works are developed to deal with spatially correlated functional data
  
  **Functional geostatistical data**:
  - Kriging methods: Giraldo et al. (2010), Bohorquez et al. (2016), ...
  - Nonparametric regression: Dabo et al. (2011, 2018, 2020), Ternynck (2014), ...
- **Lattice functional data**: less developed
  - Ruiz-Medina (2012) : prediction of SAR hilbertian processes
  - Pineda-Rios and Giraldo (2016) : FLMs with SAR disturbance process
We consider daily temperature data recorded at 772 stations from the meteorological monitoring network of Colombia.

We have 13,149 data at each station corresponding to daily records of maximum temperature (°C) obtained from January 1, 1980 to December 31, 2015.

Prediction of the whole temperature curve at a given station

The spatio-temporal dataset could be analyzed by using, space-time geostatistics (space-time kriging, see Cressie and Wikle, 2011).

However, given the high volume of data, FDA approach may be an alternative.
Modeling spatial massive data
Basic notations for functional spatial data

- $X = (X_s(\cdot), \ s \in \mathbb{R}^N)$, a measurable spatial process $N \geq 1$, defined on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$
- $X_s$ is valued in a space $(\mathcal{E}, \ d)$ of eventually infinite dimension
- $d(\cdot, \cdot)$ is some measure of proximity, e.g. a metric or a semi-metric
- $\mathcal{E}$ is a space of functions, typically the space of squared integrable functions defined on some finite interval $T = [0, T]$.
- $X$ is observed at a set of locations $\mathcal{I} \subseteq \mathbb{R}^N$ of cardinal $n$, $\mathcal{I} = \{s_1, \ldots, s_n\}, \ s_i \in \mathbb{R}^N, \ i = 1 \ldots n = 772$ and a set of time points $\mathcal{J} = \{t_1, \ldots, t_d\}, \ d = 13.149$
- $S$ the set of the $n \times d$ discrete observations, $S = \{x_{s_i}(t_j), s_i \in \mathcal{I}, t_j \in \mathcal{J}\}$.
- Prediction of a whole curve $X_{s_0} = \{X_{s_0}(t), \ t \in T\}$
Smoothing the data

- First FDA step: the discrete data \( \{x_{s_i}(t_j), s_i \in I, t_j \in J \} \) are converted into curves \( \{X_{s_i}(t), s_i \in I, t \in T \} \) by using smoothing methods (e.g. Splines).
- \( \{X_{s_i}(t), s_i \in I, t \in T \} \) are valued in \( \mathcal{E} = L^2[0, T] \) of infinite dimension
- Expand each \( X_{s_i}(.) \) in terms of basis functions (here Fourier basis).
- The number of basis functions chosen by using cross-validation.
**Second order stationary**

**Definition**

\( X \) is second order stationary, if

- for a fixed \( t \), \( E(X_s(t)) \), does not depend on \( s \) and the mean function \( E(X_s(.)) \) is measurable on \( \mathcal{T} \)

- for all \( t, t' \) and \( s, s' \), \( \text{Var}(X_s(t) - X_{s'}(t')) \), depends only on \( s - s' \)

\[ 2\gamma_{t,t'}(h) = \text{Var}(X_{s+h}(t) - X_s(t')) \] is the variogram function, we denote by \( \gamma_t(h) = \gamma_{t,t}(h) \)
Definition

The trace-variogram $\gamma(h) = \int_{\mathcal{T}} \gamma_t(h) \, dt$. By Fubini, we have

$$2\gamma(h) = E \int_{\mathcal{T}} (X_{s_i}(t) - X_{s_j}(t))^2 \, dt, \quad h = s_i - s_j, \ s_i, s_j \in \mathcal{I}$$
Let us suppose an isotropic variogram: \( \text{Var}(X_s(t) - X_{s'}(t')) \)
depends only on \( h = \|s - s'\| \)

**Definition**

The trace-variogram estimate is

\[
\hat{\gamma}_n(h) = \frac{1}{2 \# N(h)} \sum_{s_i, s_j \in N(h)} \int_T (X_{s_i}(t) - X_{s_j}(t))^2 dt, \quad h \in \mathbb{R}^N,
\]

where

\[
N(h) = \{(s_i, s_j) : h - \Delta \leq \|s_i - s_j\| \leq h + \Delta; \quad i, j = 1, \ldots, n\}.
\]
Ordinary functional Kriging

The predictor is \( \hat{X}_{s_0} = \sum_{i=1}^{n} \lambda_i X_{s_i} \), with

\[
E(\hat{X}_{s_0}) = E(X_{s_0}) \quad \text{and} \quad E \int_{T} (\hat{X}_{s_0}(t) - X_{s_0}(t))^2 dt \quad \text{minimum}
\]

The \( (\lambda_i)_{i=1,n} \) are solutions of the system (\( m \) is a Lagrange multiplier)

\[
\begin{pmatrix}
0 & \gamma(\|s_1 - s_2\|) & \ldots & \gamma(\|s_1 - s_n\|) & 1 \\
\gamma(\|s_1 - s_2\|) & 0 & \ldots & \gamma(\|s_2 - s_n\|) & 1 \\
\ldots & \ldots & \ldots & \ldots & \ldots \\
\gamma(\|s_1 - s_n\|) & \gamma(\|s_2 - s_n\|) & \ldots & 0 & 1 \\
1 & 1 & \ldots & 1 & 0 \\
\end{pmatrix}
\begin{pmatrix}
\lambda_1 \\
\lambda_2 \\
\ldots \\
\lambda_n \\
m \\
\end{pmatrix} =
\begin{pmatrix}
\gamma(\|s_0 - s_1\|) \\
\gamma(\|s_0 - s_2\|) \\
\ldots \\
\gamma(\|s_0 - s_n\|) \\
1 \\
\end{pmatrix}
Ordinary Kriging

Kriging Variance

\[ \sigma_{OK}^2(s_0) = E((\hat{X}_{s_0} - X_{s_0})^2) = m + \sum_{i=1}^{n} \lambda_i \gamma(||s_i - s_0||) \]
Prediction of a smoothed curve on an unvisited station

Kriging of a smoothed curve on an unvisited station (black curve).
A number of Spatial Data Science topics to investigate for real problems applications

- Continuously indexed spatial processes
- Non-parametric space-time modeling
- Spatio-functional modeling
- Spatial Survival nonparametric analysis
- Spatial nonparametric analysis of extreme data
- Recent review on complex spatial FDA: Martinez-Hernandez and Genton (2020); Mateu and Giraldo (2020)
- Large number of Applications to: Environmental sciences, Agriculture, Ecology, Hydrology, Energy, Biology, Images processing, Disease Mapping and Modeling, Epidemiology,...
- ...

...
Thank you for your attention