

Spatial-functional statistics with applications to environment and health in Africa

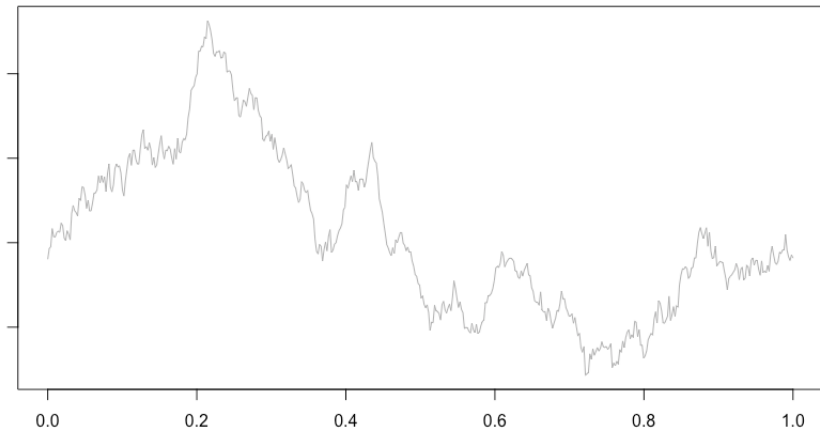
Sophie Dabo-Niang

Professor of Applied Mathematics, University of Lille, Lab Painleve and
INRIA-MODAL, Chair of EMS-CDC

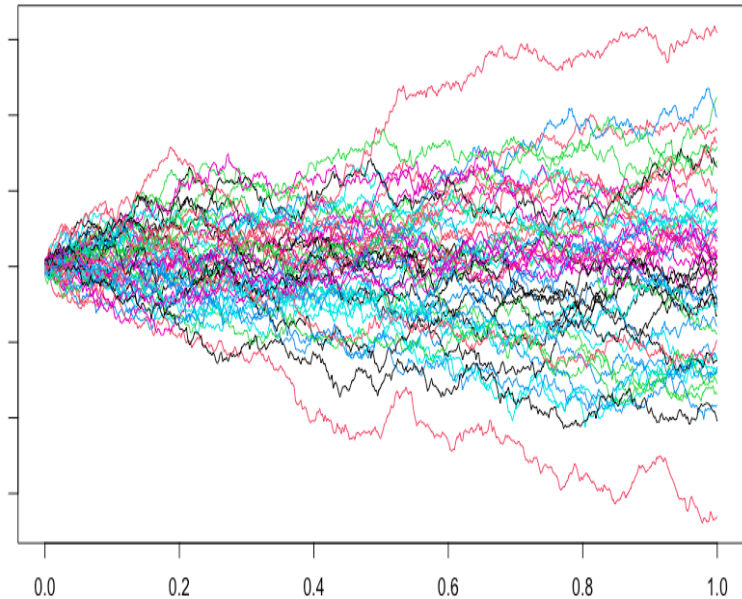
sophie.dabo@univ-lille.fr

LIRIMA seminar, February 3th, 2021

FDA (Functional Data Analysis) object



Several objects = FDA sample



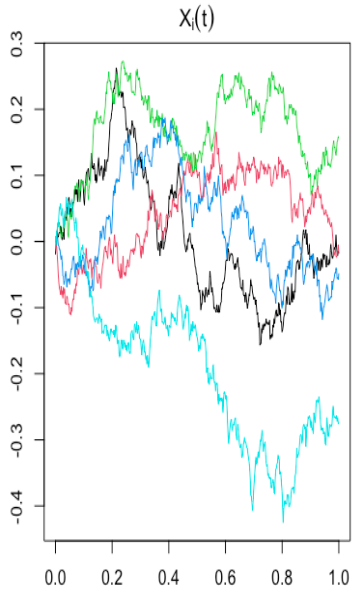
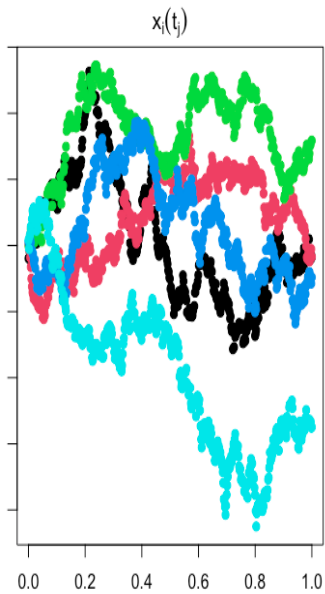
Observed sample (raw data) is always discrete

The raw sample is of the form :

$$x_i(t_j), \quad t_j \in \mathcal{T}, \quad i = 1, \dots, n \quad j = 1, \dots, p$$

- ① Transform $x_i(t_j)$ as n observations $\{X_i(t), t \in \mathcal{T}\}$, of a stochastic process $\{X(t), t \in \mathcal{T}\}$:
- ② $\{X_i(t), t \in \mathcal{T}\}$, valued in a functional space ($L^2(\mathcal{T})$; space of square integrable functions in \mathcal{T} , Hilbert, Banach space, ...)
- ③ by a functional Basis expansion (Splines, Fourier Basis,...)

Raw to FDA data



- Spatio-Functional Data
- Parametric modeling of spatial data and Application to Malaria automatic diagnostic
- Nonparametric spatial-functional estimation and Application to fisheries in West Africa
- Conclusion

Spatio-Functional Data

Functional Random Variable

A random variable X is called **functional** if it takes values in an **functional space** (of eventually infinite dimensional space) :

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

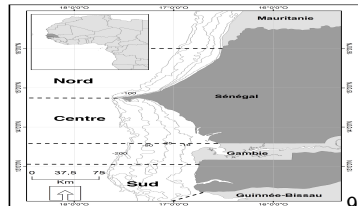
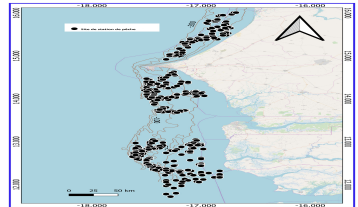
- If $\mathcal{T} \subseteq \mathbb{R}$ then X is a curve
- If $\mathcal{T} \subseteq \mathbb{R}^2$ then X is an image
- If $\mathcal{T} \subseteq \mathbb{R}^p$ ($p > 2$); X is a more complex structure

Applications : Environmental science, biometrics, chemometrics, mining, natural sciences, economics, medicine, . . .

Application to fisheries in West African Coast

Scientific campaigns of **CRODT**, **IRD** (Senegal) : mapping of fish stocks in order to avoid fishing some species, impact of climate change on stocks

- **Environmental Data** : Bottom Temperature (SBT), Surface temperature (SST), Bottom Salinity (SBS), Surface Salinity (SSS).
- **Biological Data** : different species
Dentex angolensis ; Pagrus caeruleostictus ; Galeoides decadactylus.
Pagellus bellottii ; Octopus Vulgarus ; Epinephelus-aeneus
- **Spatial Information** : latitude and longitude.



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
- images analysis
- 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 phenomena observed on spatial sets : observations are **spatially dependent**.

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

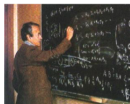


D. Krige



D. Krige wearing the back, 1938.

'60 Mathematical Foundation at MinesParisTech by G. Matheron

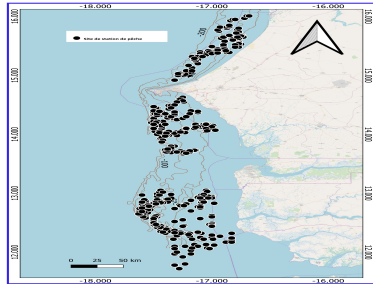


(1930-2000)

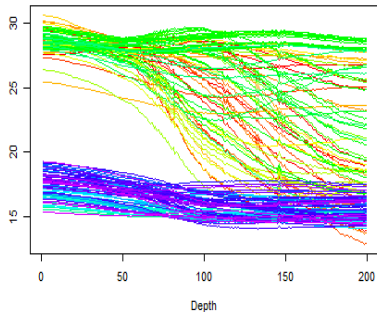
- Firstly mining and petroleum applications
- Nowadays mostly data are spatial or spatio-temporal
- Famous spatial prediction method (**Kriging**) by D. Krige, **intensively** used in **Statistical and Machine Learning Theories and Applications**.

Spatial data

- Let S a spatial set, $S \subset \mathbb{R}^N$, $N \geq 2$.
- In particular, $N = 2$ and a location $s \in \mathbb{R}^2$ is (latitude, longitude)
- Spatial data can be modeled as realizations of random fields :
 $Z = \{Z_s, s \in S\}$, collection of random variables indexed in S .



S



$Z = \{Z_s, s \in S\}_{12}$

Types of spatial data

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 literature on spatial is abundant : Guyon (1995), Anselin and Florax (1995), Cressie (1991) or Ripley (1981), Cressie and Wikle (2013),...

In about **20 years** : most of usual statistical methods, technics, have been extended to FDA

PCA, Change-point, Classification, Outlier detection, Regression,...

See the Monographs : Bosq (2000), Ramsay and Silverman (2002, 2005), Ferraty and Vieu (2007), Dabo-Niang and Ferraty (2008), Horváth and Kokoszka (2012), Kokoszka and Reimherr (2017), Mateu and Giraldo (2021),...

Review See also the reviews on FDA of Wang et al. (2016), Martinez-Hernandez and Genton (2020)

More and More Spatial, Spatio-Functional software, scientific events, resources

- Extensive number of [resources](#) : published papers, [scientific events](#), journals, societies, ...
- More than 100 free R, Python, Matlab,..., [spatial](#), and [FDA](#) resources
- [Workshop](#) on "Recent FDA Advances", AIMS, Senegal, April 2022, with [most of researchers working on FDA](#)

Parametric modeling of spatial data
and Application to Malaria automatic
diagnostic

Spatial parametric prediction : Kriging

- 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, \dots, s_n .
- Prediction of Z_{s_0} (or $g(Z_{s_0})$) using $Z_{s_1}, Z_{s_2}, \dots, Z_{s_n}$
- With the help of a covariance (variogram) model to account the spatial dependency

Intrinsic stationary

Second order stationary is not always satisfied.

One way to establish the second order stationarity hypothesis is to consider the h - increments process from Z : intrinsic process

Definition

- Z is an intrinsic process, if for every h :
 - Z is of constant mean
 - the increment process $I^{(h)} = \left\{ I_s^{(h)} = Z_{s+h} - Z_s, \right\}$ is a second order stationary process in s :

$$E(I_s^{(h)}) = 0 \quad \text{Var}(Z_{s+h} - Z_s) = 2\gamma(h),$$

where 2γ is called **variogram** function.

Variogram

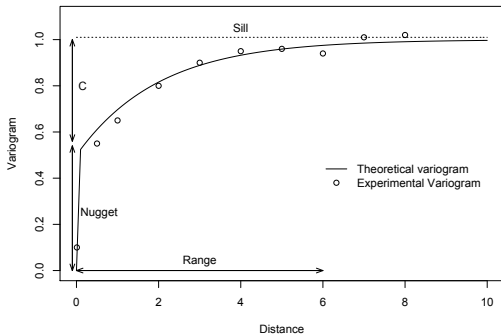
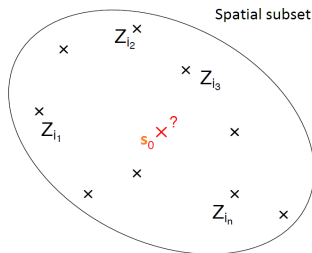


Figure 1 – Experimental and theoretical variograms. The nugget is the variance between very close observations. The range indicates the distance after which observations are independent.

Prediction of Z_{s_0}



Prediction (Kriging) Z_{s_0} is to find the best linear predictor (BLUP) of Z_{s_0} based on observations $\mathcal{Z} = (Z_{s_1}, \dots, Z_{s_n})$ and a variogram model $\gamma(\cdot)$:

$$\hat{Z}_{s_0} = \sum_{i=1}^n \lambda_i Z_{s_i}$$

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

Isotropic admissible variograms

Examples of isotropic variograms associated with a covariance $C(h) = \sigma^2 - \gamma(h)$, where σ^2 is the associated stationary variance.

- nugget $\gamma(h) = C$
- exponential $\gamma(h) = C(1 - \exp(-\|h\|/a))$ $C > 0$ $a > 0$
- spherical $\gamma(h) = \begin{cases} C\left(\frac{3}{2} \frac{\|h\|}{a} + \frac{1}{2} \frac{\|h\|^3}{a^3}\right) & \text{if } 0 \leq \|h\| \leq a \\ C & \text{if } \|h\| \geq a \end{cases}$
- gaussian $\gamma(h) = C(1 - \exp(-\|h\|^2/a))$ $C > 0$ $a > 0$
- power $\gamma(h) = C\|h\|^\alpha$, $\alpha < 2$

Detection and segmentation of erythrocytes in multispectral label-free blood smear images for automatic cell counting : (Doumoun, Dabo-Niang and Zoueu ; 2020)

- **Kriging+ Sampling** : Reduction of 66% of acquisition time

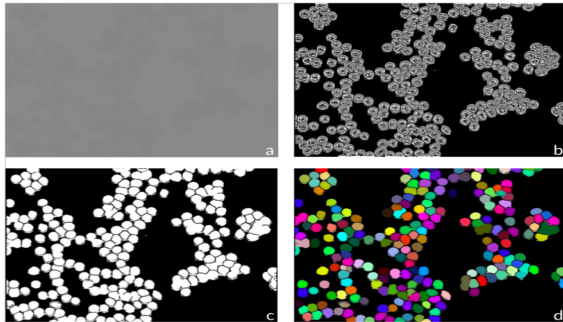
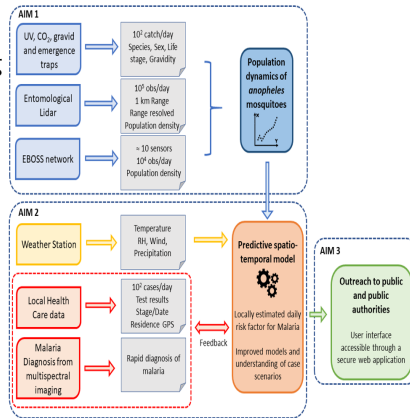


Figure 2 – Segmentation process illustrated with a 435 nm image ; (a) sample image, (b) adaptive thresholding, (c) post-processing and (d) counting

Mapping malaria risk of exposure in Ivory Coast

Spatio-temporal **malaria risk of exposure mapping** and **rapid diagnosis** of malaria (Doumoun, Dabo-Niang and Zoueu ; 2020)

- **Population dynamics of anopheles mosquitoes** : innovative entomological lidar + bistatic optical sensors + trapping methodologies to study population dynamics of anopheles mosquitoes.
- **Spatial Information** : different urban and rural areas in Ivory Coast.
- **Climat and environment** : weather data
- **Spatio-temporal big data** from physical traps, optical sensors, weather stations



Nonparametric spatial-functional
estimation and Application to fisheries
in West Africa

Preliminary step : estimation of a regression function $r(\cdot)$

$$Y = r(X) + \epsilon$$

$Y \in \mathbb{R}$, $X \in \mathbb{R}^d$, ϵ , the error term.

→ with the help of spatial observations $(Z_{s_i}, s_i \in \mathbb{Z}^N)$, $N \geq 1$,
 $Z_{s_i} = (X_{s_i}, Y_{s_i})$, same law as (X, Y) .

Context

Let the spatial process $(Z_{s_i} = (X_{s_i}, Y_{s_i}) \in \mathbb{R}^d \times \mathbb{R}, s_i \in \mathbb{Z}^N)$, observed on some discrete set

$$\mathcal{I}_n = \{s_i \in \mathbb{N}^N, 1 \leq s_k \leq n_k, k = 1, \dots, N\}$$

- the Z_{s_i} have unknown density, $f_{X,Y}$ w.r.t to Lebesgue measure on \mathbb{R}^{d+1} ;
- X_{s_i} of density f ;
- the regression function is

$$r(x) = \begin{cases} \varphi(x)/f(x) & \text{if } f(x) \neq 0; \\ \mathbb{E}(Y) & \text{otherwise,} \end{cases}$$

where $\varphi(x) = \int_{\mathbb{R}} y f_{X,Y}(x, y) dy$ with $y \in \mathbb{R}$ and $x \in \mathbb{R}^d$.

Classical Kernel estimation of $f(\cdot)$ and $r(\cdot)$

With observations $\{(X_{s_i}, Y_{s_i}), s_i \in \mathcal{I}_n\}$, we have for $x \in \mathbb{R}^d$:

$$r_n(x) = \begin{cases} \varphi_n(x)/f_n(x) & \text{if } f_n(x) \neq 0; \\ \frac{1}{\hat{n}} \sum_{s_i \in \mathcal{I}_n} Y_{s_i} & \text{otherwise,} \end{cases}$$

with

$$\varphi_n(x) = \frac{1}{\hat{n}h_n^d} \sum_{s_k \in \mathcal{I}_n} Y_{s_k} K\left(\frac{x - X_{s_k}}{h_n}\right)$$

and

$$f_n(x) = \frac{1}{\hat{n}h_n^d} \sum_{s_k \in \mathcal{I}_n} K\left(\frac{x - X_{s_k}}{h_n}\right)$$

- K a kernel on \mathbb{R}^d , with compact support, valued in \mathbb{R}
- $\lim_{n \rightarrow \infty} h_n = 0$
- $\hat{n} := n_1 \times \dots \times n_N$ (sample size)

Drawback of this kernel regression estimator

- Do not take into account the spatial information in the data
- Strict stationarity assumption
- Estimate the regression function in the same as if the processes were independent

More flexible estimates

Based on product of 2 kernels :

one on observations

the other on the locations (sites)

Used in Hall, Muller and Wu (2006), Wang and Wang (2009), Wang, Wang and Huang (2012), Dabo-Niang, Hamdad, Ternynck and Yao (2014), Ternynck (2014), , Dabo-Niang et al. (2016), ...

Kernel Spatial prediction and local stationarity

- Let $\{Z_{s_i} = (X_{s_i}, Y_{s_i}) \in \mathbb{R}^d \times \mathbb{R}, s_i \in \mathbb{N}^N\}$ be a spatial process observable in

$$\mathcal{I}_n = \{s_i \in \mathbb{N}^N : 1 \leq s_r \leq n_r, r = 1, \dots, N\}$$

- Regression model :

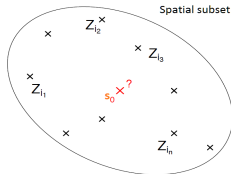
$$Y_{s_i} = r(X_{s_i}) + \varepsilon_{s_i}, \quad s_i \in \mathbb{N}^N \quad \text{with} \quad r(\cdot) = E(Y_{s_i} | X_{s_i} = \cdot)$$

- $\{\varepsilon_{s_i}\}$ is centred, α -mixing process

\Leftrightarrow Goal : prediction of $\{Y_{s_i}, s_i \in \mathbb{N}^N\}$ at an unobserved site $s_0 \in \mathcal{I}_n$, using :

- X_{s_0} and
- $(X_{s_i}, Y_{s_i})_{s_i \in \mathcal{O}_n}$, with $s_0 \notin \mathcal{O}_n \subset \mathcal{I}_n$

under the **assumption** : $(X_{s_i}, Y_{s_i})_{s_i \in \mathcal{O}_n}$ are locally identically distributed around s_0 (Klemelä, 2008) instead of the classical stationarity condition.



Let $\tilde{s}_j = \frac{s_j}{\mathbf{n}}$ (\rightarrow normalized).

For fixed $X_{s_j} = x \in \mathbb{R}^d$, observed at s_j (denoted x_{s_j}) :

$$f_{\mathbf{n}}(x_{s_j}) = \frac{1}{\widehat{\mathbf{n}} b_{\mathbf{n}}^d \rho_{\mathbf{n}}^N} \sum_{s_i \in \mathcal{I}_{\mathbf{n}}} K_1 \left(\frac{x_{s_j} - X_{s_i}}{b_{\mathbf{n}}} \right) K_2 \left(\frac{\|\tilde{s}_j - \tilde{s}_i\|}{\rho_{\mathbf{n}}} \right)$$

where

- K_1 et K_2 : kernels on \mathbb{R}^d and \mathbb{R} , valued in \mathbb{R}
- $\lim_{\mathbf{n} \rightarrow \infty} b_{\mathbf{n}} = 0$ and $\lim_{\mathbf{n} \rightarrow \infty} \rho_{\mathbf{n}} = 0$
- $\mathbf{n} = (n_1, \dots, n_N) \rightarrow \infty$ means that $\min_i n_i \rightarrow \infty$, $i = 1, \dots, N$

Estimator of $r(\cdot)$

$$r_{\mathbf{n}}(x_{s_j}) = \begin{cases} \varphi_{\mathbf{n}}(x_{s_j})/f_{\mathbf{n}}(x_{s_j}) & \text{if } f_{\mathbf{n}}(x_{s_j}) \neq 0; \\ \frac{1}{\widehat{\mathbf{n}}} \sum_{s_i \in \mathcal{I}_{\mathbf{n}}} Y_{s_i} & \text{else,} \end{cases}$$

with

$$\varphi_{\mathbf{n}}(x_{s_j}) = \frac{1}{\widehat{\mathbf{n}} b_{\mathbf{n}}^d \rho_{\mathbf{n}}^N} \sum_{s_i \in \mathcal{I}_{\mathbf{n}}} Y_{s_i} K_1 \left(\frac{x_{s_j} - X_{s_i}}{b_{\mathbf{n}}} \right) K_2 \left(\frac{\|\tilde{s}_j - \tilde{s}_i\|}{\rho_{\mathbf{n}}} \right)$$

Kernel spatial predictor

Let s_0 a location where Y_{s_0} is not available, the kernel prediction is

$$\hat{Y}_{s_0}^{NW} = \frac{\sum_{s_i \in \mathcal{O}_n} Y_{s_i} K_1 \left(\frac{X_{s_0} - X_{s_i}}{h_n} \right) K_2 \left(\rho_n^{-1} \left\| \frac{s_0 - s_i}{\mathbf{n}} \right\| \right)}{\sum_{s_i \in \mathcal{O}_n} K_1 \left(\frac{X_{s_0} - X_{s_i}}{h_n} \right) K_2 \left(\rho_n^{-1} \left\| \frac{s_0 - s_i}{\mathbf{n}} \right\| \right)},$$

- ▶ K_1 and K_2 are two kernels and $\frac{s_i}{\mathbf{n}} = \left(\frac{s_{i,1}}{n_1}, \dots, \frac{s_{i,N}}{n_N} \right)$.
- ▶ h_n, ρ_n are non random bandwidths.

↪ The fact that h_n is deterministic, does not allow to capture some heterogeneity.

k -NN kernel spatial predictor

Explore k -NN kernel method (Collomb, 1980) to deal with an eventual heterogeneity :

$$\hat{Y}_{\mathbf{s}_0} = \frac{\sum_{s_i \in \mathcal{O}_{\mathbf{n}}} Y_{s_i} K_1 \left(\frac{X_{\mathbf{s}_0} - X_{s_i}}{H_{\mathbf{n}, X_{\mathbf{s}_0}}} \right) K_2 \left(h_{\mathbf{n}, \mathbf{s}_0}^{-1} \left\| \frac{\mathbf{s}_0 - s_i}{\mathbf{n}} \right\| \right)}{\sum_{s_i \in \mathcal{O}_{\mathbf{n}}} K_1 \left(\frac{X_{\mathbf{s}_0} - X_{s_i}}{H_{\mathbf{n}, X_{\mathbf{s}_0}}} \right) K_2 \left(h_{\mathbf{n}, \mathbf{s}_0}^{-1} \left\| \frac{\mathbf{s}_0 - s_i}{\mathbf{n}} \right\| \right)},$$

- ▶ $H_{\mathbf{n}, X_{\mathbf{s}_0}}$: random bandwidth on the covariate :

$$H_{\mathbf{n}, X_{\mathbf{s}_0}} = \min \left\{ h \in \mathbb{R}_+^* \mid \sum_{s_i \in \mathcal{O}_{\mathbf{n}}} \mathbb{I}(\|X_{s_i} - X_{\mathbf{s}_0}\| < h) = k(\mathbf{n}) \right\}$$

- ▶ $h_{\mathbf{n}, \mathbf{s}_0}$: bandwidth on locations :

$$h_{\mathbf{n}, \mathbf{s}_0} = \min \left\{ h \in \mathbb{R}_+^* \mid \sum_{s_i \in \mathcal{O}_{\mathbf{n}}} \mathbb{I} \left(\left\| \frac{s_i - \mathbf{s}_0}{\mathbf{n}} \right\| < h \right) = k_{\mathbf{n}}^1 \right\}$$

- ▶ The smoothing parameters $k_{\mathbf{n}}^1$ and $k(\mathbf{n})$ are positive integers.

Application to fisheries data of West African Coast

Data from coastal demersal surveys off Senegalese coasts performed by the scientific team of the CRODT in cold and hot seasons, in North, Center and South areas, from 2001 to 2015.

↔ At 495 stations :

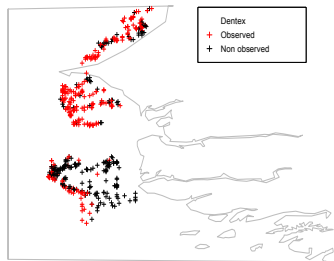
- **Environmental data** : sea bottom temperature (SBT), sea surface temperature (SST), sea bottom salinity (SBS) and sea surface salinity (SSS).
- **Spatial information** : latitude and longitude.
- **Biological data** : three species

Dentex : *Dentex angolensis*

Pagrus : *Pagrus caeruleostictus*

Thiekem : *Galeoides decadactylus*

↔ Goal : **predict** the presence of species



Classification

The following methods are compared :

Method	Covariates	Spatial param
Basic k-NN SVM	SBT, SST, SBS, SSS, latitude and longitude	No
Logit model	SBT, SBS, latitude and longitude	No
Kernel k-NN Kernel	SBT, SST, SBS, SSS	Yes

For the three species :

- Dataset is split into : training and testing samples, with respective sizes equal to 80% and 20% of the whole sample size.
- Tuning parameters of each method are selected through cross validation method or AIC (Logit model).
- Correct classification rate (CCR) is computed for each method.

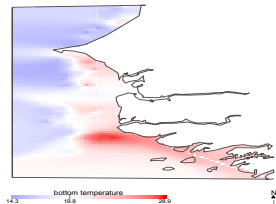
Results

Specie	Method	CCR for :		
		All	Observed	Non observed
Dentex	Basic k-NN	0.776	0.811	0.733
	SVM	0.816	0.887	0.733
	Logit model	0.714	0.774	0.644
	Kernel	0.745	0.736	0.756
	Kernel k-NN	0.847	0.868	0.822
Pagrus	Basic k-NN	0.735	0.516	0.836
	SVM	0.786	0.452	0.940
	Logit model	0.704	0.258	0.910
	Kernel	0.827	0.581	0.940
	Kernel k-NN	0.796	0.581	0.896
Thiekem	Basic k-NN	0.857	0.400	0.974
	SVM	0.908	0.750	0.950
	Logit model	0.878	0.750	0.910
	Kernel	0.898	0.700	0.949
	Kernel k-NN	0.959	0.850	0.987

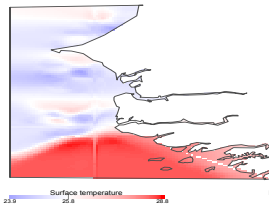
Application to prediction

Prediction of SBT, SST, SBS and SSS in several positions :

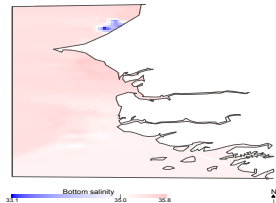
Sea bottom temperature



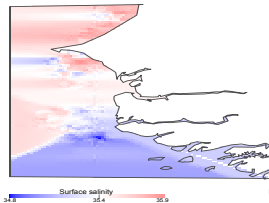
Sea surface temperature



Sea bottom salinity



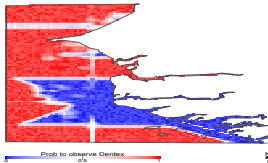
Sea surface salinity



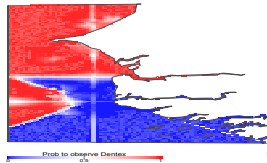
Application on fisheries data of West African Coast

Use the previous predictions of SBT, SST, SBS and SSS to predict the probability to observe each specie :

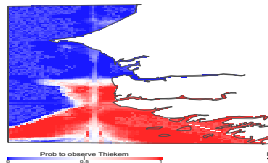
Dentex



Pagrus



Thiekem



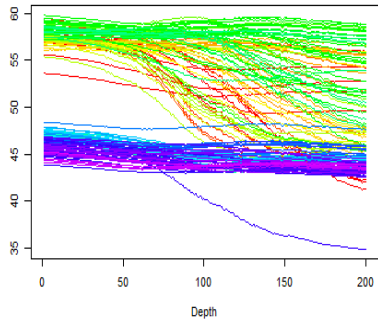
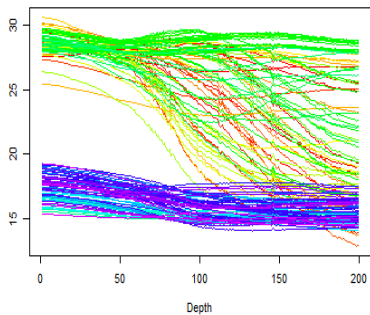
Spatio-functional kernel regression and prediction

- $X = (X_s(\cdot), s \in \mathbb{R}^N)$, a measurable spatial process $N \geq 1$, defined on some probability space $(\Omega, \mathcal{A}, \mathbf{P})$
- 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 $\mathcal{T} = [0, T]$.
- X is observed at a set of locations $\mathcal{I} \subseteq \mathbb{R}^N$ of cardinal n , $\mathcal{I} = \{s_1, \dots, s_n\}$, $s_i \in \mathbb{R}^N$, $i = 1 \dots n$ and a set of depth points $\mathcal{J} = \{t_1, \dots, t_p\}$
- In practice, we have a set of $n \times d$ discrete observations, $S = \{x_{s_i}(t_j), s_i \in \mathcal{I}, t_j \in \mathcal{J}\}$.

Smoothing the data

- **First FDA step** : the discrete data $\{x_{s_i}(t_j), s_i \in \mathcal{I}, t_j \in \mathcal{J}\}$ are converted into curves $\{X_{s_i}(t), s_i \in \mathcal{I}, t \in \mathcal{T}\}$ by using smoothing methods (e.g. Splines).
- $\{X_{s_i}(t), s_i \in \mathcal{I}, t \in \mathcal{T}\}$ are valued in \mathcal{E} of infinite dimension
- **Reduce the infinite dimension** : expand each $X_{s_i}(\cdot)$ in terms of basis functions (here Fourier basis).
- The number of basis functions chosen by using cross-validation.

Temperature and Pressure curves



Kernel-based procedure for estimating the nonlinear functional regression model :

$$Y_{s_i} = r(X_{s_i}) + \epsilon_{s_i}, \quad i = 1, \dots, n$$

- Y_{s_i} : scalar response $\in \mathbb{R}$
- X_{s_i} : **functional predictor** (e.g. curves) $\in (\mathcal{E}, d(\cdot, \cdot))$ where $d(\cdot, \cdot)$ is a semi-metric
 α -mixing process (general case of weakly dependence)
- $\{\epsilon_s\}$: **stationary** residual process
independent of the process $\{X_s\}$

Kernel regression for functional data

Let $X_{s_i} \in (\mathcal{E}, d)$ (dimension ∞) and $Y_{s_i} \in \mathbb{R}$, the kernel estimator :

$$\hat{r}(x) = \frac{\sum_{i=1}^n Y_{s_i} K\left(\frac{d(x, X_{s_i})}{h}\right)}{\sum_{j=1}^n K\left(\frac{d(x, X_{s_j})}{h}\right)}, \quad x \in (\mathcal{E}, d)$$

using observations $(X_{s_i}, Y_{s_i})_{i=1, \dots, n}$, where $d(\cdot, \cdot)$ is a semi-metric.

→ Investigated first by Ferraty and Vieu (2000) and then

for independent data by e.g. Ferraty and Vieu (2002);

Dabo-Niang and Rhomari (2003); ...

for dependent data by e.g. Masry (2005); Ternynck (2014); ,

Dabo-Niang et al (2016), Ndiaye et al. (2020), ...

Account the spatial feature

Use the spatial locations of the sample to improve estimation of $r(\cdot)$ to a prediction goal :

$$\hat{Y}_{\mathbf{s}_0} = \frac{\sum_{s_i \in \mathcal{O}_{\mathbf{n}}} Y_{s_i} K_1 \left(\frac{d(x, X_{s_i})}{b_{\mathbf{n}}} \right) K_2 \left(\rho_{\mathbf{n}}^{-1} \left\| \frac{\mathbf{s}_0 - s_i}{\mathbf{n}} \right\| \right)}{\sum_{s_i \in \mathcal{O}_{\mathbf{n}}} K_1 \left(\frac{d(x, X_{s_i})}{b_{\mathbf{n}}} \right) K_2 \left(\rho_{\mathbf{n}}^{-1} \left\| \frac{\mathbf{s}_0 - s_i}{\mathbf{n}} \right\| \right)}, \quad (1)$$

Table 1 – *ER* and CCR with functional covariate salinity

Functional Methods	ER	CCR \rightarrow Y=1,0	TCC \rightarrow Y=1	TCC \rightarrow Y=0
MFC	0,34	0,44	0,83	0,37
FGLM	0,44	0,59	0,98	0,14
FGSAM	0,36	0,64	0,90	0,34
KNNF	0,43	0,53	1,00	0,00
DD^G	0,27	0,57	0,75	0,37
MFSD	0,09	0,75	0,78	0,70

Another alternative : reduce the dimension

- Expansion of $X(\cdot)$ using a set of basis functions (for instance PCA basis) $\phi_k(\cdot)$, $k = 1, \dots, K_x$.
- Observed curve X_{s_i} from $X(\cdot)$ is expressed

$$X_{s_i}(t) = \sum_{k=1}^K c_{ik} \phi_k(t)$$

- Do kernel regression estimate with the coefficients c_{ik} as covariates
- Functional kernel problem is then reduced to regress a finite number of coefficients c_{ik}

Other ongoing projects in Africa

- **Analysis of marine pollution marine in West Africa using acoustic halieutic data** : Spatio-functional Modeling.
Y. Kandé, P. Brehmer, M. Ndiaye. [IRD](#), [CROD](#), [Senegal](#)
- **Sea Surface temperature Analysis** : Spatio-functional Model Based-Clustering (Dabo-Niang et al. (2021)).
F. Adjibode, M. Ndiaye,...,[Benin](#)
- **Economic growth and Environment in MENA countries** : by Spatial Econometrics modeling.
A. El Ghini, S. Yacour. [Morroco](#)
- **Diffusion of Covid-19** : by Influential-imitator diffusion models.
T. Tchouya, S. Nasini [Benin](#)

Other ongoing projects in Africa : Cancer cell characterisation

Rapid Cancer detection : by FDA modeling and Bio-MEMS

C. Agonkou, Benin

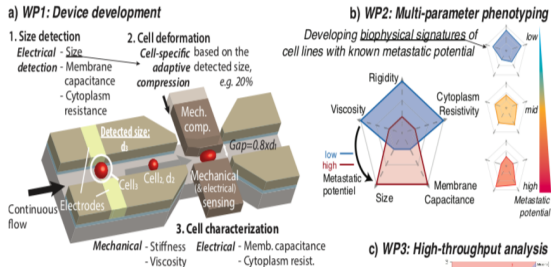
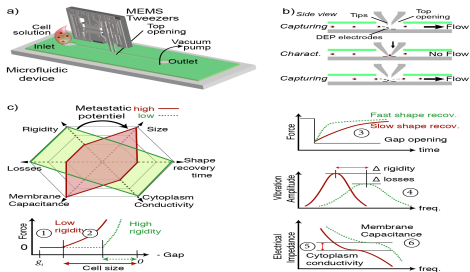


Fig. 1: Overview of HiPCIC.

- WP1: MEMS device for high-throughput single-cell characterization using autonomous adaptive sensing.
- WP2: Multi-parameter phenotyping to classify biophysical cell properties according to known metastatic potential of cancer cell lines
- WP3: Predicting the metastatic potential (low, mid & high) of cancer cell lines & patient samples.

Other ongoing projects in Africa : Cancer cell characterisation



Hot topics to investigate

- Non-parametric space-time modeling
- Multivariate Spatio-functional data modeling
- Spatial Survival nonparametric analysis
- Spatial nonparametric analysis of extreme data
- See the recent reviews on Spatial FDA :
 [Martinez-Hernandez and Genton \(2020\)](#)
 [Mateu and Giraldo \(2021\)](#)
- ...

Thank you for your attention