Modelling epidemiosurveillance with a marked sequential point process

François d'Alayer de Costemore d'Arc

Coll. Edith Gabriel & Samuel Soubeyrand

Biostatistique BIO/Π & Processus Spatiaux INRAE

Réunion annuelle ModStatSAP 2023

Plan

What is epidemiological surveillance ?

Surveillance is the monitoring of diseases and pathogens.
We suppose that the disease under study **doesn't spread**.

What are its objectives ?

- **Evaluation objective**
  Making disease status assessment easy.
- **Control Objective**
  Fighting against the disease.
- **Containment Objective**
  Optimally allocating strategies of control.

Method of surveillance

Getting a sample in the population and testing it.
The result of the test is binary : 1 if infected and 0 if not infected.

What is epidemiological surveillance ?

Surveillance is the monitoring of diseases and pathogens.
We suppose that the disease under study **doesn't spread**.

What are its objectives ?

- **Evaluation objective**
  Making disease status assessment easy.
- **Control Objective**
  Fighting against the disease.
- **Containment Objective**
  Optimally allocating strategies of control.

Method of surveillance

Getting a sample in the population and testing it.
The result of the test is binary : 1 if infected and 0 if not infected.

What is epidemiological surveillance ?

Surveillance is the monitoring of diseases and pathogens.
We suppose that the disease under study **doesn't spread**.

What are its objectives ?

- **Evaluation objective**
  Making disease status assessment easy.
- **Control Objective**
  Fighting against the disease.
- **Containment Objective**
  Optimally allocating strategies of control.

Method of surveillance

Getting a sample in the population and testing it.
The result of the test is binary : 1 if infected and 0 if not infected.

What is epidemiological surveillance ?

Surveillance is the monitoring of diseases and pathogens.
We suppose that the disease under study **doesn't spread**.

What are its objectives ?

- **Evaluation objective**
  Making disease status assessment easy.
- **Control Objective**
  Fighting against the disease.
- **Containment Objective**
  Optimally allocating strategies of control.

Method of surveillance

Getting a sample in the population and testing it.
The result of the test is binary : 1 if infected and 0 if not infected.

The necessity of sampling

Why do we need to sample the population ?

Because having perfect knowledge of the disease requires **testing at all times everywhere**.

But each test is costly and time consuming.

So we need an **efficient way of sampling** in a population $W \subset \mathbb{R}^2$.

Saliency map

- **Saliency map**

$$\begin{array}{rcl} \alpha : W & \mapsto & [0,1] \\ x & \to & \alpha(x) \text{ probability that } x \text{ is infected.} \end{array}$$

- We know factors of risk $\alpha_0, \ldots, \alpha_r$ (population density,...). We can then model α :

$$\alpha(x) = \text{sig}(\sum_{i=0}^{r} \lambda_i \alpha_i(x))$$

with λ_i real coefficients and sig : $\mathbb{R} \mapsto [0,1]$ the sigmoid function.

Surveillance designs

- A surveillance design $Y = (X_i, M_i)_i$ is a marked point process in W, such that $X_i \in W$ is a spatial location and $M_i \sim \text{Ber}(\alpha(X_i))$ is the status of point X_i.

- **Our goal :** find **efficient surveillance designs** with respect to the objectives of epidemiological surveillance (**disease evaluation** and **control**).

- When defining a surveillance design $Y = (X_i, M_i)_i$, we only specify the positions of the samples $X = (X_i)_i \in W$.
  We do not control the distribution of $M = (M_i)_i \in \{0, 1\}$.

Control action

Given a saliency map α (probability of being infected) and a surveillance design $Y = (X_i, M_i)_i$ (location of individuals being tested), we define control actions.

Control actions model the action we do after sampling positive individuals (confinement for example). We denote by S the set of surveillance designs.

Control action

A control action A is an application $A : [0,1]^w \times S \mapsto [0,1]^w$ such that $A(\alpha, Y)$ is a new saliency map.

Exemple

$$A(\alpha, Y)(u) = \alpha(u)\tau(u, Y)$$

with $0 \leq \tau(u, Y) \leq 1$ that is small near positive points of Y and moves towards 1 elsewhere.

Evaluation of the disease

In the **Evaluation Objective** and for now on, the true saliency map α is unknown.

A surveillance design Y gives us an estimate α_Y of α. We aim to reduce the uncertainty of this estimate.

$$\min_{Y \in S} \text{IBV}(\alpha_Y) = \min_{Y \in S} \int_W \alpha_Y(x)[1 - \alpha_Y(x)]dx$$

where α_Y is the estimate of α from the surveillance design Y and IBV stands for Integrated Bernoulli Variance.

Goal : Uncertainty reduction

Optimization of the Control

In the **Control Objective**, we look for surveillance designs that minimize the integral of the saliency map after applying a control.

We want to maximize the reduction of the risk globally.

$$\min_{Y \in S} \int_W A(\alpha, Y)(x) dx$$

Goal : Reduction of the risk

Containment Objective

In some areas, it is **not efficient** to apply a control.
It is thus important to **estimate** these areas.

In some cases, such areas are where α is too important :

$$L(c) = \{x \in W | \alpha(x) > c\}, \text{ for } 0 < c < 1$$

We then search surveillance designs $Y \in S$ such that the estimate $L_Y(c)$ of $L(c)$ is the most accurate.

Goal : Estimation of random areas

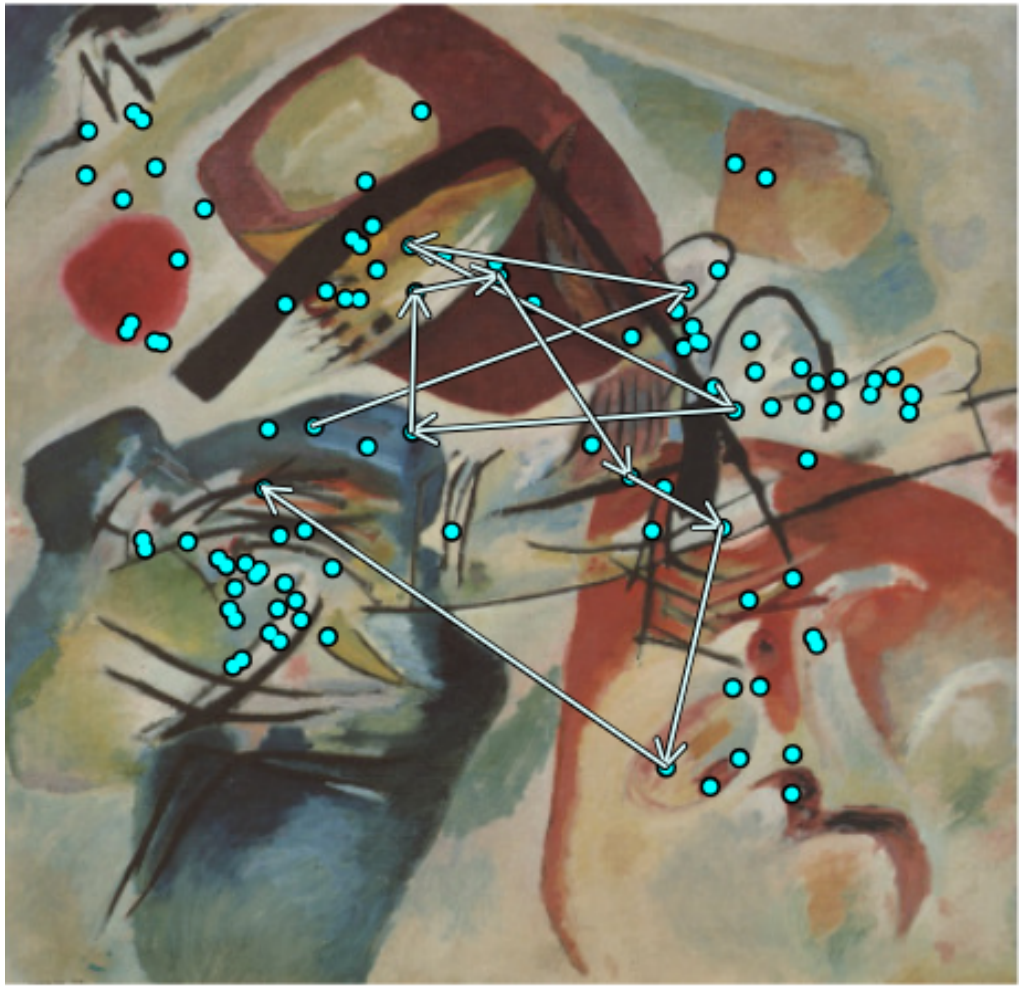
Sequential approach

Sequential point process are generally defined through their conditional densities :

$$f(\overrightarrow{x_n}) = f_1(x_1) \prod_{k=1}^{n-1} f_{k+1}(x_{k+1}|\overrightarrow{x_k}), \tag{1}$$

with $\overrightarrow{x_n} = \{x_1, \ldots, x_n\}$ an ordered sequence of points in W. Each point x_i depends on all the precedent points.

Eye Movement Model (Penttinen, 2016)



Mit dem schwarzen Bogen(Avec l'arc noir), Wassily Kandinsky (1912)

The observer inspects the whole scene at the beginning and gradually focuses on a few details.

Characteristics of surveillance designs

In modeling surveillance designs, 3 characteristics are important :

- Spatial heterogeneity : Saliency map.
- Short term dependency : Kernel.
- Self interaction : Reweighting function.

Kernel

A kernel K such that $\int_W K(x, y)dy = 1$ creates a markovian dependency between points of the sequential design.

This is wanted since the position of the next sample generally depends on the previous point.

Example

The Gaussian Kernel

$$K(x_k, x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp^{-\frac{1}{2\sigma^2}||x_k - x||^2}$$

Self interaction

Self interaction : implicit dependency between a new point y_k and its past
$\overrightarrow{y_{k-1}} = (y_1, \ldots, y_{k-1}) = ((x_1, m_1), \ldots, (x_{k-1}, m_{k-1}))$.

This is done by putting **more or less weight** on areas **according to the proportion of disease** in these areas.

Example

for $E \subset W, 0 < \theta < 1$,

$$\pi(y, E) = \begin{cases} \theta \text{ if } y \in E \\ 1 - \theta \text{ sinon} \end{cases}$$

Sequential point process for epidemiological surveillance

Density of our Sequential Surveillance Process for an **ordered** sequence $\overrightarrow{y_n} = ((x_1, m_1), \ldots, (x_n, m_n))$:

$$f(\overrightarrow{y_n}) = f_1(y_1) \prod_{k=1}^{n-1} f_{k+1}(y_{k+1}|\overrightarrow{y_k}),$$

with

- $f_{k+1}(u|\overrightarrow{y_k}) \propto \alpha(u) K(y_k, u) \pi(u, p(u, \overrightarrow{y_k}))$,
- $p(u, \overrightarrow{y_k})$ the proportion of positive points of $\overrightarrow{y_k}$ at distance $\mathbf{r} > 0$ of u.
- $\pi(u, p(u, \overrightarrow{y_k})) \propto p(u, \overrightarrow{y_k})^{\mathbf{a}} (1 - p(u, \overrightarrow{y_k}))^{\mathbf{b}}$, $\mathbf{a}, \mathbf{b} \geq 0$.

Sequential point process for epidemiological surveillance

Density of our Sequential Surveillance Process for an **ordered** sequence $\overrightarrow{y_n} = ((x_1, m_1), \ldots, (x_n, m_n))$:

$$f(\overrightarrow{y_n}) = f_1(y_1) \prod_{k=1}^{n-1} f_{k+1}(y_{k+1}|\overrightarrow{y_k}),$$

with

- $f_{k+1}(u|\overrightarrow{y_k}) \propto \alpha(u) K(y_k, u) \pi(u, p(u, \overrightarrow{y_k}))$,
- $p(u, \overrightarrow{y_k})$ the proportion of positive points of $\overrightarrow{y_k}$ at distance $\mathbf{r} > 0$ of u.
- $\pi(u, p(u, \overrightarrow{y_k})) \propto p(u, \overrightarrow{y_k})^{\mathbf{a}} (1 - p(u, \overrightarrow{y_k}))^{\mathbf{b}}$, $\mathbf{a}, \mathbf{b} \geq 0$.

Sequential point process for epidemiological surveillance

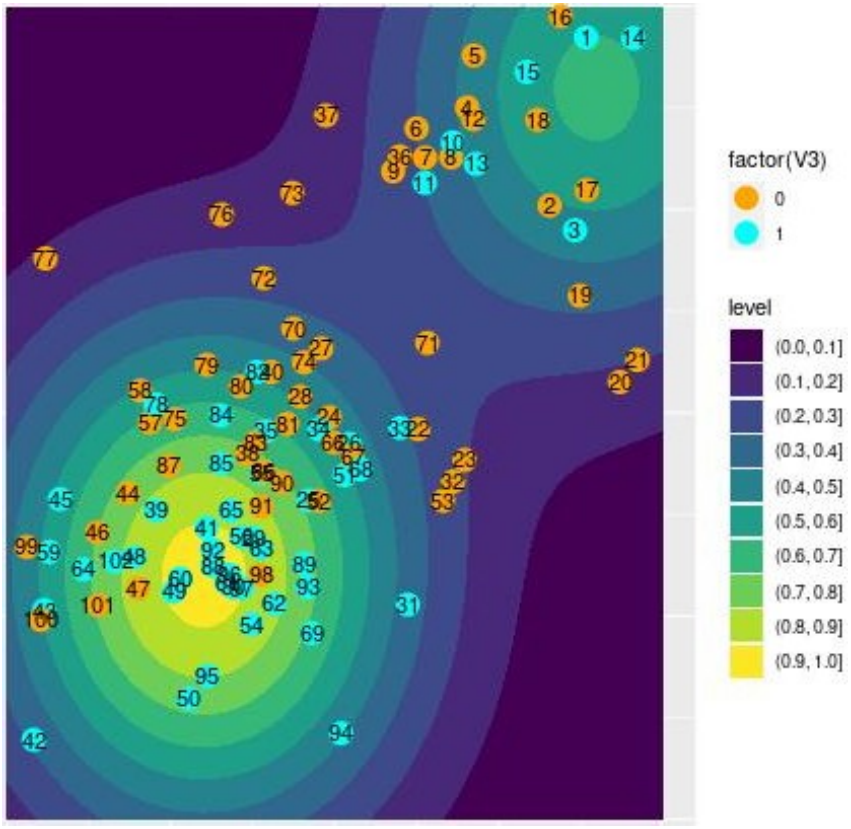
Density of our Sequential Surveillance Process for an **ordered** sequence $\overrightarrow{y_n} = ((x_1, m_1), \ldots, (x_n, m_n))$:

$$f(\overrightarrow{y_n}) = f_1(y_1) \prod_{k=1}^{n-1} f_{k+1}(y_{k+1}|\overrightarrow{y_k}),$$

with

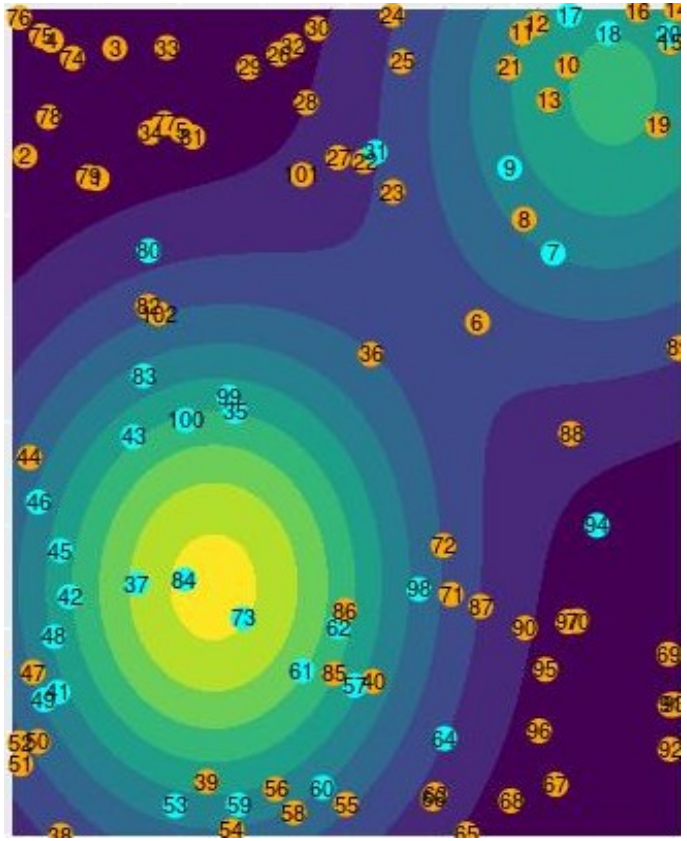
- $f_{k+1}(u|\overrightarrow{y_k}) \propto \alpha(u)K(y_k, u)\pi(u, p(u, \overrightarrow{y_k}))$,
- $p(u, \overrightarrow{y_k})$ the proportion of positive points of $\overrightarrow{y_k}$ at distance $\mathbf{r} > 0$ of u.
- $\pi(u, p(u, \overrightarrow{y_k})) \propto p(u, \overrightarrow{y_k})^{\mathbf{a}}(1 - p(u, \overrightarrow{y_k}))^{\mathbf{b}}$, $\mathbf{a}, \mathbf{b} \geq 0$.

Simulation of the model



(a,b)=(0,0)

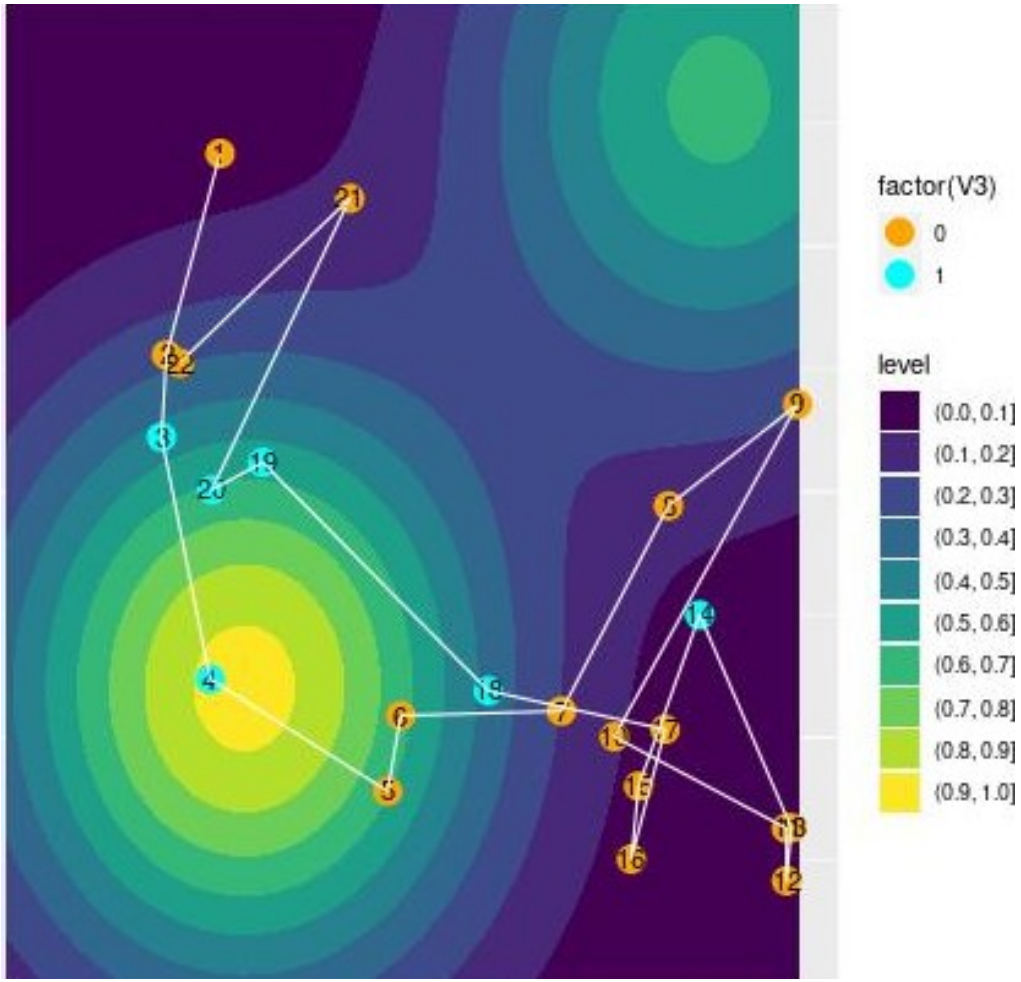
All proportions are equally likely. $\pi = 1$



(a,b)=(1,200)

Small proportions are more likely.

Illustration of the sequential component



New summary statistics

We need new summary statistics to characterize possible behaviours of our model.

The usual ones don't accommodate the sequential component.

We use them to compare three sets of parameters of our model :

- **Set 1** : (a,b)=(1,200). Color : BLACK
- **Set 2** : (a,b)=(0,100). Color : BLUE
- **Set 3** : (a,b)=(0,0). Color : RED

Sets 1 and 2 favor areas of low positivity of the disease, whereas set 3 favors no area in particular.

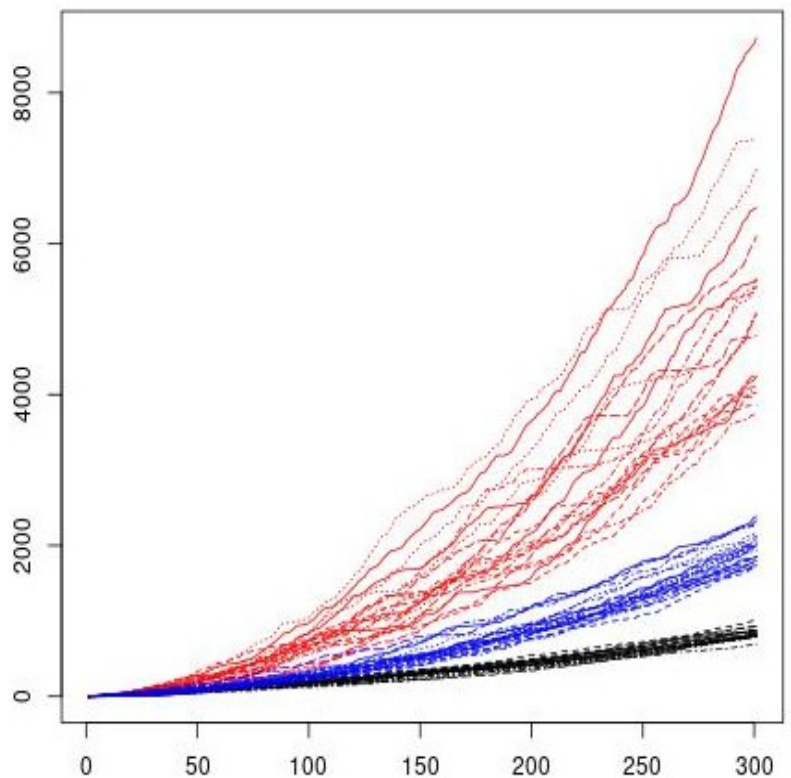
Cumulative Lagged Positive Proportion(**CLPP**)

- $\mathrm{CLPP}(\vec{y_k}) = \sum_{i=2}^{k} \sum_{\substack{j \le i-1 \\ m_j = 1}} \mathbb{1}_{B(x_i,r)}(x_j)$
- $f_{k+1}(u|\vec{y_k}) \propto \alpha(u) K(y_k, u) p(u, \vec{y_k})^a (1 - p(u, \vec{y_k}))^b$

Set 1 : (a,b)=(1,200)

Set 2 : (a,b)=(0,100)

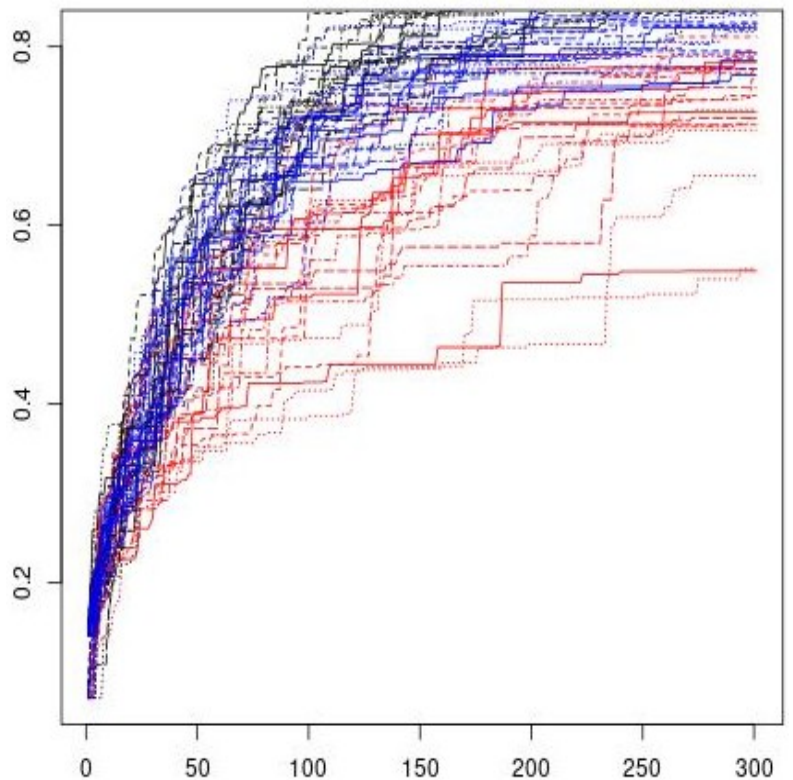
Set 3 : (a,b)=(0,0)



Ball Union Coverage of Positive Points (**BUC**)

- $\mathrm{BUC}(\overrightarrow{y_k}) = \frac{|\bigcup_{i|m_i=1}^{k} B(x_i,r)\cap W|}{|W|}$
- $f_{k+1}(u|\overrightarrow{y_k}) \propto \alpha(u) K(y_k, u) p(u, \overrightarrow{y_k})^a (1 - p(u, \overrightarrow{y_k}))^b$

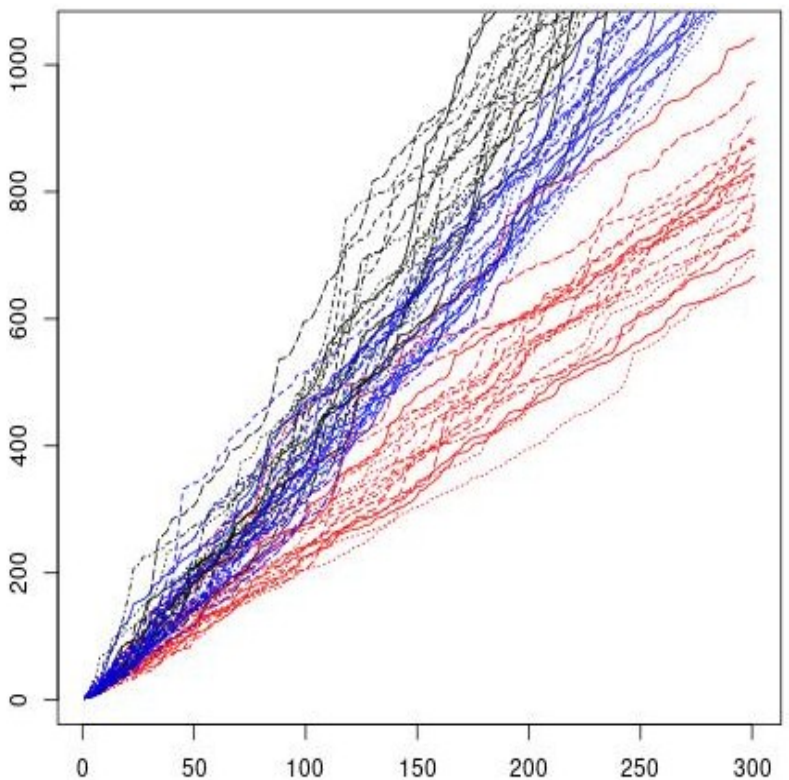
Set 1 : (a,b)=(1,200)
Set 2 : (a,b)=(0,100)
Set 3 : (a,b)=(0,0)



Cumulative Scanpath Length (**SLC**)

- $\mathrm{SLC}(\overrightarrow{y_k}) = \sum_{i=1}^{k} \sum_{j=0}^{M_i} ||x_{i+1-j} - x_{k-j}||$
  where $M_i = \inf\{n \in \mathbb{N}^* | m_{i+1-n} = 1\} - 1$
- $f_{k+1}(u|\overrightarrow{y_k}) \propto \alpha(u) K(y_k, u) p(u, \overrightarrow{y_k})^a (1 - p(u, \overrightarrow{y_k}))^b$

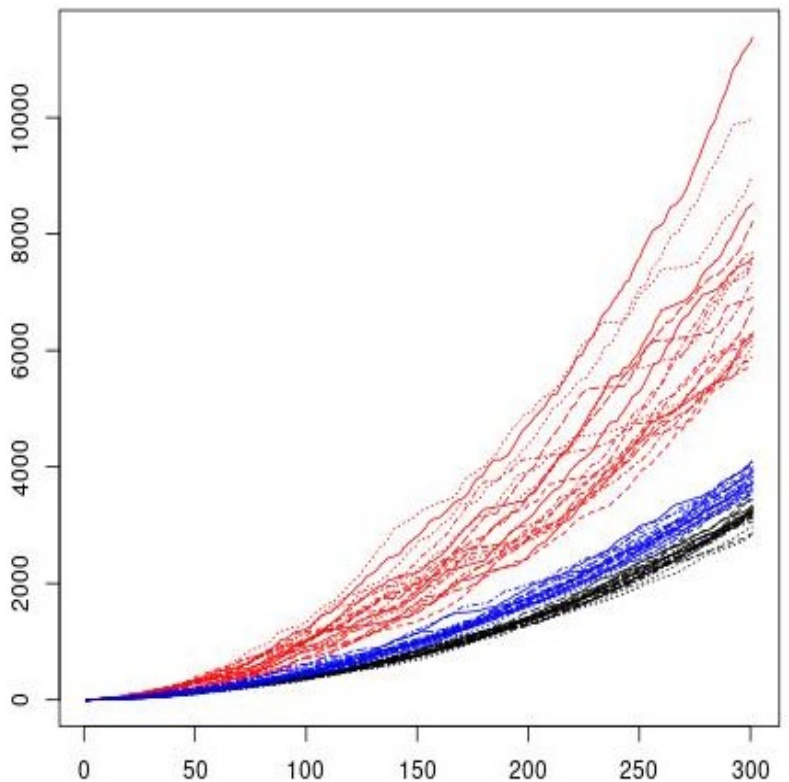
Set 1 : (a,b)=(1,200)
Set 2 : (a,b)=(0,100)
Set 3 : (a,b)=(0,0)



Cumulative Recurrence (**CR**)

- $\mathrm{CR}(\overrightarrow{y_k}) = \sum_{i=2}^{k} \sum_{j=1}^{i-1} \mathbb{1}_{x_j \in B(x_i, r)}$
- $f_{k+1}(u|\overrightarrow{y_k}) \propto \alpha(u) K(y_k, u) p(u, \overrightarrow{y_k})^a (1 - p(u, \overrightarrow{y_k}))^b$

Set 1 : (a,b)=(1,200)
Set 2 : (a,b)=(0,100)
Set 3 : (a,b)=(0,0)



Conclusion

- There exists **no** single surveillance **design optimum** for the 3 objectives we fixed at once.
- The class of our model being smaller than the class of all surveillance designs, we could **find different sets of parameters such that our model has good performance** in the 3 problems of epidemiosurveillance.
- The **parametric** nature of the model allows the **induction of characteristics for a good surveillance design**. Having found the optimal set of parameters, we can describe optimal designs with the help of new summary statistics.

Perspectives

- Computation of the optimal surveillance designs.
  Are there significant differences between these designs ?

- Case study on the surveillance of *Xylella fastidiosa*.
  By studying the data of surveillance samples in Corsica and PACA region, we'll determine if the surveillance done in practice is optimal.

Questions

Contact for references, precisions :
francois.dalayer@inrae.fr
Questions ?