



Sciences Economiques et Sociales de la Santé
& Traitement de l'Information Médicale

sesstim.univ-amu.fr

Sokhna DIENG

Doctorante au sein de l'équipe **QuantIM**

**Analyse de données fonctionnelles pour la classification de profils épidémiologiques :
Application au paludisme**

décembre 2018



Cliquez ici pour voir l'intégralité des ressources associées à ce document



Réseau doctoral
en santé publique



Analyse de données fonctionnelles pour la classification de profils épidémiques : application au paludisme

Dieng S^{1,2}, Sallah K¹, Michel P³, Ba E⁴, Cissé B⁴, Milligan P⁵, Gaudart J⁶

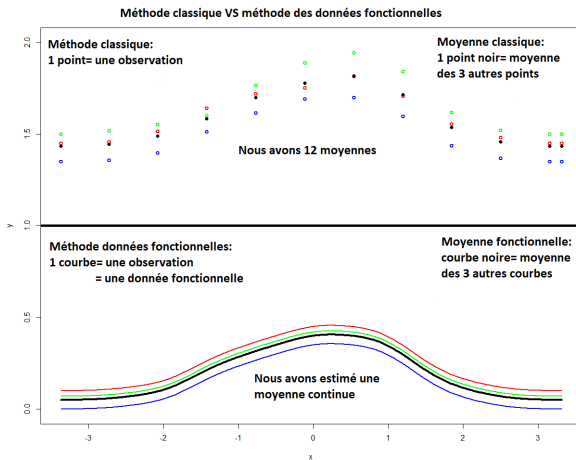
1.Aix Marseille Univ, IRD, INSERM, SESSTIM, Marseille, France 2.Ecole des Hautes Etudes en Santé Publique, Rennes, France 3.Aix-Marseille University, CNRS, EHESP, Centrale Marseille, AMSE 4.IRD, Dakar, Sénégal 5.London School of Hygiene Tropical Medicine, Londres, Royaume-Uni 6.Aix Marseille Univ, APHM, INSERM, IRD, SESSTIM, Hop Timone, BioSTIC, Biostatistic & ICT, Marseille, France.

Seminaire SESSTIM 14 décembre 2018

Contexte et Objectif

● Contexte

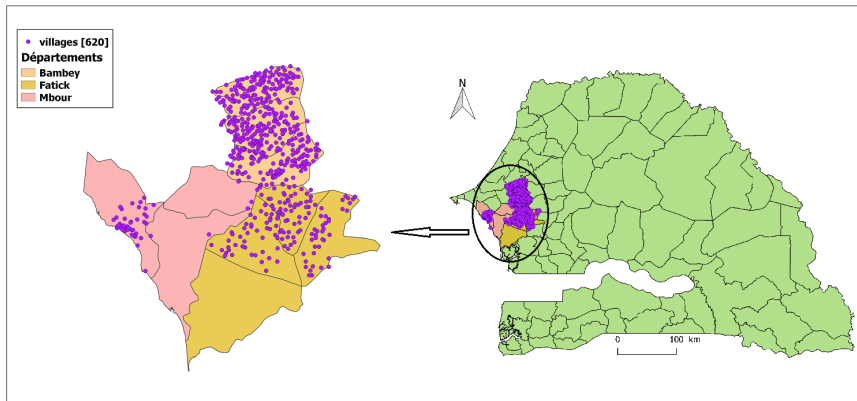
- En 2016, 90% des 216 million cas de paludisme et 91% des 445000 décès \Rightarrow en Afrique subsaharienne. [OMS, 2017]
- Entre 2009-2016, incidence paludisme de 14 à 23.6 p.1000 p.a au Sénégal [PNLP, 2016]
- Hétérogénéité du risque sur zones géographiques
- Longues séries temporelles par unité géographique :
 - **Données de grandes dimensions**
 - **Nature fonctionnelle des données**
- Analyse de données fonctionnelles (FDA)
Passer de $[x(t_1), x(t_2), \dots, x(t_p)]$ à $\tilde{x}(t)$



● Objectif

Rechercher des profils épidémiques de paludisme pour optimiser les actions de luttes ciblées

Site d'étude



● \approx 600 000 habitants et 54 postes de santé

Données

- **Période:** janvier 2008- décembre 2012

Données	Échelle spatiale	Échelle temporelle	Collecte
Paludisme	village	semaine	poste de santé
Population	village	année	recensement

- **Centralisation:** système de suivi démographique et sanitaire de Bambey-Fatick-Mbour

Méthodes statistiques et logiciels

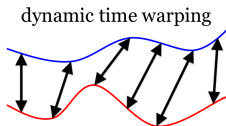
- **Données fonctionnelles avec bases BSpline cubique**

y_{ij} : taux d'incidence du village i à la semaine $j \Rightarrow y_{ij} = x_i(t_{ij}) + \epsilon_{ij}$

avec $\mathbf{x}_i(\mathbf{t}) = \sum_{k=1}^K \mathbf{c}_{ik} \phi_k(\mathbf{t})$: fonction "d'évolution de l'incidence" du village i où K est le nombre de base, c_{ik} les coefficients et ϕ_k BSpline cubique

ϵ_{ij} l'erreur d'approximation suivant une loi normale de moyenne 0 et variance constante

- **Classification hiérarchique ascendante (WARD) avec la distance DTW (Dynamic Time Warping)**



Source image: MathWorks.com

- **Modèle additif généralisé**

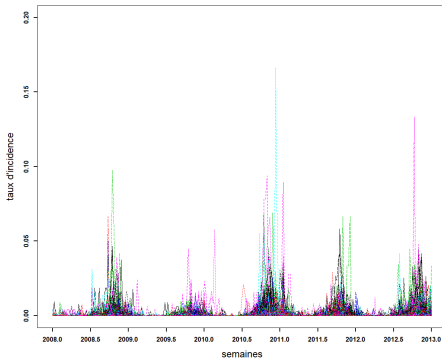
- Surdispersion \Rightarrow loi quasipoisson
- Variabilité géographique \Rightarrow fonction spline bivariée sur longitudes et latitudes

- **Logiciels**

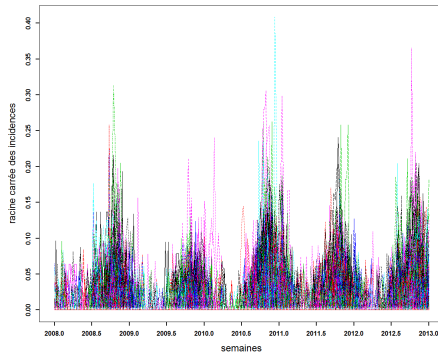
- R: analyses statistiques (packages: fda, fda.usc, FactoMineR, mgcv)
- QGIS: représentation profils palustres

Les incidences fonctionnelles

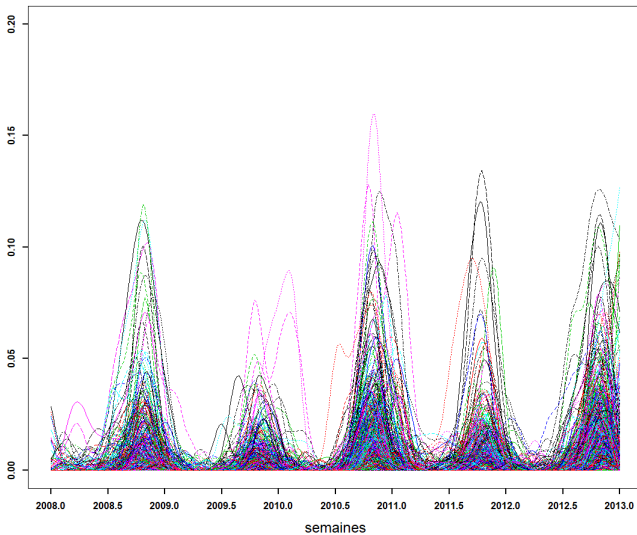
Taux d'incidence observés par village



Transformation racine carrée

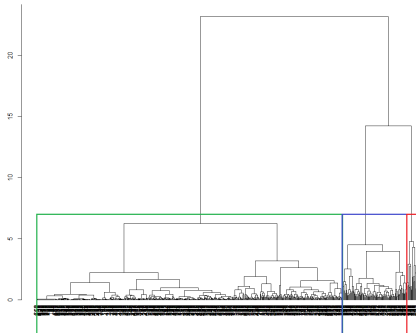


Taux d'incidence fonctionnels par village

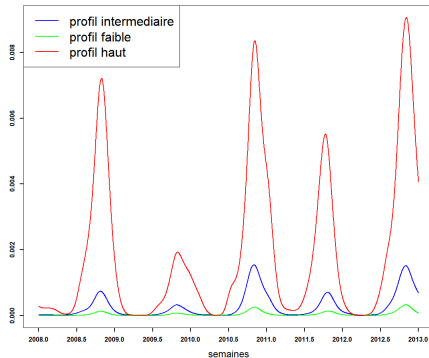


Les profils et leurs évolutions temporelles

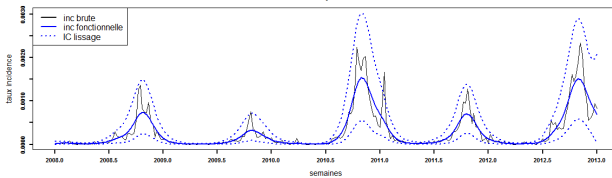
Dendrogramme de la classification hiérarchique ascendante avec DTW



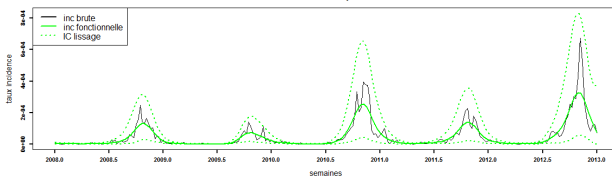
Taux d'incidence fonctionnels par profil



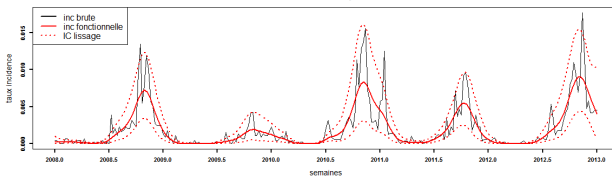
Incidence et IC du profil intermédiaire



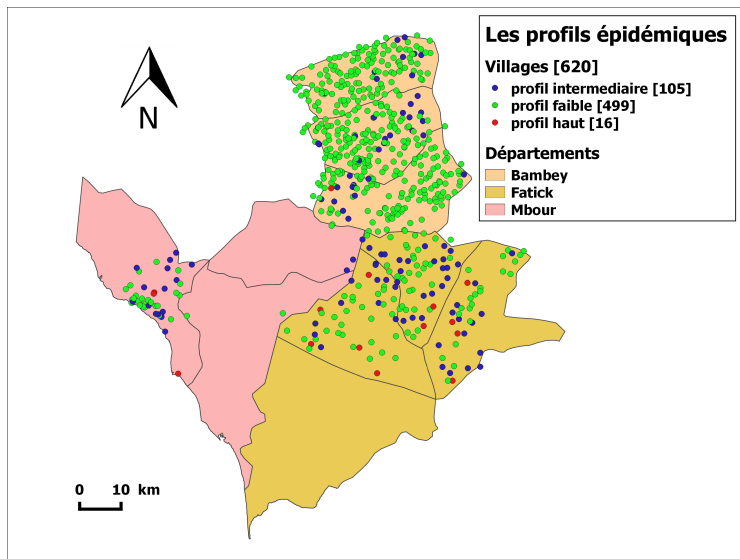
Incidence et IC du profil faible



Incidence et IC du profil haut



La répartition spatiale des profils épidémiques

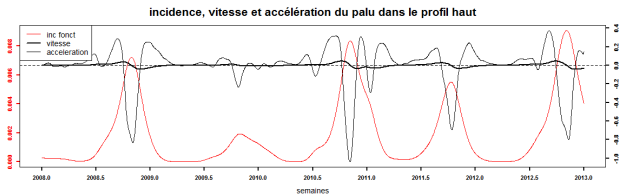
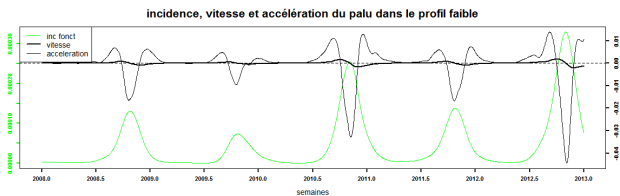
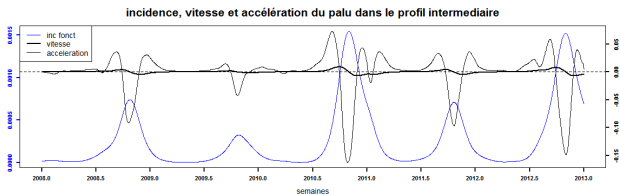


Les caractéristiques des profils épidémiques

Profils (nombre de villages)	Durée épidémies annuelles moyennes en semaines (Ecart-Type)	Taux d'incidence (100.000 pers.sem) (min; med; max)	Rapport d'incidence standardisé (I.C 95%)
Faible (499)	26.6 (2.3)	5 (0; 6; 127)	0.18 (0.16,0.20)
Intermédiaire (105)	26.8 (3.7)	30 (10; 31; 104)	Réf
Haut (16)	32.4 (4.2)	212 (122; 206; 350)	7.5 (6.25,8.92)

Discussion et Conclusion

- Trois profils épidémiques avec dynamiques temporelles différentes du paludisme
- Contexte ressources limitées \Rightarrow cibler des actions de luttés en fonction des caractéristiques des profils(régression logistique sur les profils avec facteurs entomologiques, sociodemographiques et environnementaux)
- Autres utilisations de la méthode des données fonctionnelles: étude des dérivées pour vitesse et accélération épidémie, régressions fonctionnelles...



Remerciements

- Réseau Doctoral en Santé Publique animé par École des Hautes Étude en Santé Publique
- UMR SESSTIM Marseille
- London School of hygiene and tropical medicine
- IRD DAKAR

Références

- OMS, "Rapport sur le paludisme dans le monde 2017", 2017
- Programme National de Lutte contre le Paludisme au Sénégal," Rapport d'activité", 2016
- Programme National de Lutte contre le Paludisme au Sénégal," Plan stratégique National de Lutte Contre le Paludisme (2016-2020)",2015
- J.O.Ramsay, Giles Hooker, Spencer Graves : "Functionnal Data Analysis with R and Matlab", Springer Science & Business Media, 2009
- Julien Jacques, Cristian Preda : "Functional data clustering : a survey",Springer Verlag, 2014
- André Bouchier : "L'analyse des données multivariées à l'aide du logiciel R: La classification ascendante hiérarchique (C.A.H)", Mars 2010
- Simon WOOD. Generalized additive models: an introduction with R. CRC press, 2006.

découpage temporel du taux d'incidence global du site

