



**HAL**  
open science

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

Sokhna Dieng, El-Hadj Ba, Badara Cissé, Kankoé L Sallah, Paul Milligan,  
Jean Gaudart

► **To cite this version:**

Sokhna Dieng, El-Hadj Ba, Badara Cissé, Kankoé L Sallah, Paul Milligan, et al.. Analyse de données fonctionnelles pour la classification de profils épidémiques : application au paludisme. *Epidemiology and Public Health = Revue d'Epidémiologie et de Santé Publique*, 2017, 65 (2), pp.S48. 10.1016/j.respe.2017.03.006 . inserm-01873096

**HAL Id: inserm-01873096**

**<https://inserm.hal.science/inserm-01873096v1>**

Submitted on 12 Sep 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

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

Sokhna Dieng<sup>\*1,2</sup>, El-hadj Ba<sup>3</sup>, Badara Cissé<sup>3</sup>, Kankoe L Sallah<sup>1</sup>, Paul Milligan<sup>4</sup>, Jean Gaudart<sup>1</sup>

<sup>1</sup>Laboratoire Sciences Economiques & Sociales de la Santé & Traitement de l'Information Médicale, UMR 912 SESSTIM, Aix Marseille Université, IRD, INSERM, Marseille, France.

<sup>2</sup>Ecole des Hautes Etudes en Santé Publique, Rennes, France

<sup>3</sup>IRD, Dakar, Sénégal

<sup>4</sup>London School of Hygiene & Tropical Medicine, Londres, Royaume-Uni

\*Auteur correspondant.

Adresse e-mail : [sokhna.dieng@etu.univ-amu.fr](mailto:sokhna.dieng@etu.univ-amu.fr) (sokhna Dieng), [el-hadj.ba@ird.fr](mailto:el-hadj.ba@ird.fr) (El-hadi Ba), [Badara.cisse@orange.sn](mailto:Badara.cisse@orange.sn) (Badara Cissé), [kankoe.sallah@univ-amu.fr](mailto:kankoe.sallah@univ-amu.fr) (Kankoe L Sallah), [paul.milligan@lshtm.ac.uk](mailto:paul.milligan@lshtm.ac.uk) (Paul Milligan), [jean.gaudart@univ-amu.fr](mailto:jean.gaudart@univ-amu.fr) (Jean Gaudart)

*Introduction* En épidémiologie, il est fréquent d'étudier des variables fonctionnelles pour lesquelles les observations sont réalisées de façon discrètes (courbe de croissance, ECG etc.). C'est le cas des séries temporelles d'incidences. L'étude des dynamiques épidémiques doit prendre en compte cette nature fonctionnelle, en estimant et analysant les fonctions à l'origine des observations plutôt qu'en analysant les données individuelles elle-même. L'objectif de notre travail était de rechercher des profils épidémiques en tenant compte de la nature fonctionnelle des séries temporelles d'incidences pour optimiser les actions de lutte ciblées.

*Méthodes* L'analyse a porté sur le nombre de cas de paludisme hebdomadaire et la population annuelle de 620 villages du Sénégal de janvier 2008 à décembre 2012. Les 620 fonctions continues d'incidences ont été estimées en utilisant une base de Fourier, tenant compte de leur périodicité annuelle. Les profils épidémiques ont été obtenus par classification hiérarchique ascendante sur ces fonctions utilisant la distance euclidienne adaptée aux données fonctionnelles (Formule 1) et la méthode de Ward. Les rapports d'incidences associés à chaque profil ont été estimés à l'aide d'un modèle additif généralisé.

*Résultats* La classification des courbes d'incidences a permis d'obtenir 3 profils épidémiques de paludisme (Fig. 1) : un profil d'incidences basses (taux d'incidence : 4 cas /100 000 habitants.semaine) avec des durées épidémiques restreinte (durée moyenne : 22 semaines) et pas de cas entre les épidémies annuelles, un profil d'incidences hautes (186 cas /100 000 habitants.semaine) avec des durées épidémiques importantes (33 semaines) et toujours des cas persistants, et entre les deux, un profil intermédiaire (25 cas /100 000 habitants.semaine) avec une durée épidémique moyenne de 24 semaines. Malgré un aspect spatialement très dispersé, les villages du profil « bas » étaient majoritairement au Nord de la zone d'étude, ceux du profil « haut » majoritairement au sud et ceux du profil « intermédiaire » présents aussi bien au nord qu'au sud (Fig. 2). Par rapport au profil intermédiaire, le rapport standardisé d'incidence du profil bas était de 0.17 (IC95%[0.15;0.19],  $p < 0.0001$ ), et celui du profil haut était de 7.34, (IC95%[6.21;8.67]  $p < 0.0001$ ).

*Conclusion* L'analyse réalisée a permis de prendre en compte le caractère fonctionnel des courbes épidémiques indiquant ainsi 3 dynamiques temporelles différentes du paludisme.

**Mots clés** Functional data analysis ; Données fonctionnelles ; Classification ; Profils épidémiques ; Paludisme

**Formule**

Formule 1 :  $\sqrt{\frac{1}{\sqrt{T}} \int (f(t) - g(t))^2 dt}$

**Figures**

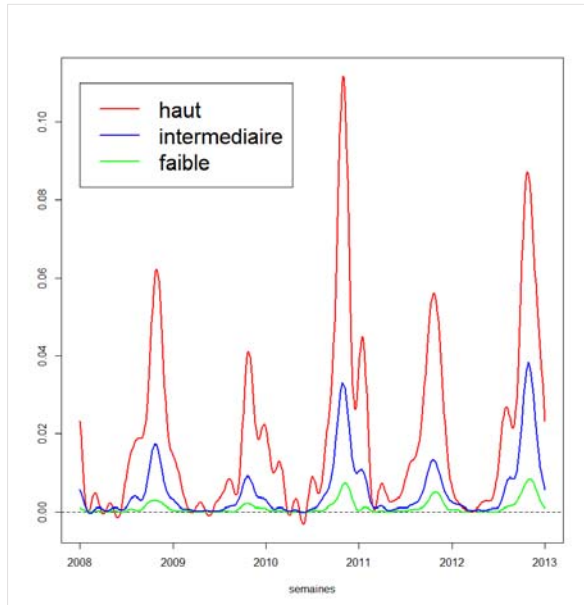


Fig.1 : Profils épidémiques : évolution temporelle des incidences

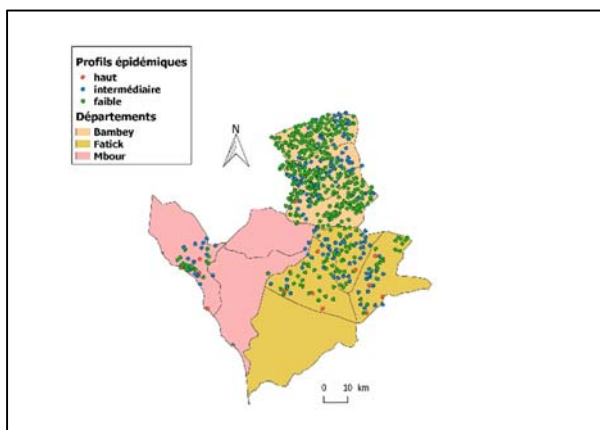


Fig.2 : La répartition spatiale des profils épidémiques

*Déclaration d'intérêt*

Les auteurs déclarent qu'ils n'ont pas perçu de conflit d'intérêt.