

## Cartographie du Risque Appliquée aux Fièvres d'Origines Inconnues à Djibouti

Mohamed ABDI KHAIRE<sup>1</sup>, Hawa ADEN FARAH<sup>2</sup> et Anne-Françoise YAO<sup>2</sup>

**Résumé :** Le rapport présente une étude épidémiologique faite sur les fièvres d'origines inconnues à Djibouti. Ces dernières sont définies par une température corporelle supérieure à 38 degré qui ne s'associe à aucune maladie transitoire ou autolimitée. Le nombre de personnes atteintes par la fièvre d'origine inconnue n'a cessé d'augmenter durant les 10 dernières années.

Nous nous intéressons à l'estimation du risque et à son évolution dans la capitale à travers les risques (taux de mortalité standardisée). Ainsi nous les projetérons sur une carte de Djibouti-ville avant de les utiliser pour l'exploration des données via des analyses statistiques.

**Abstract:** The report presents an epidemiological study on fever of unknown origin in Djibouti. The latter is defined by a body temperature greater than 38 degrees which does not associate with any transient or spontaneously resolving disease. The number of people suffering from fever of unknown origin has been steadily increasing for 10 years.

We are interested in the estimation of risk and its evolution in the capital through risks (standardized mortality rate). Thus we will project them on a map of Djibouti and we will then use them for the exploration of the data via a statistical analysis

**Mots-clés :** Etude épidémiologique, Fièvre d'origine inconnue, Cartographie du risque, Analyse de données, Classification, Régression linéaire, Taux de Mortalité Standardisé (SMR), Apprentissage non supervisé, Régression PLS, Analyse de variance (ANOVA)

**Keywords:** Epidemiological Study, Fever of Unknown Origin, Risk Mapping, Data Analysis, Classification, Linear Regression, Standardized Mortality Rate (SMR), Unsupervised Learning, PLS Regression, Analysis of Variance (ANOVA)

---

<sup>1</sup> Laboratoire de Mathématiques, Université Clermont Auvergne et Département de Statistiques, Université de Djibouti.

E-mail : [Mohamed.Abdi.Khaire@etu.uca.fr](mailto:Mohamed.Abdi.Khaire@etu.uca.fr)

<sup>2</sup> Département de Statistiques, Université de Djibouti et Laboratoire de Mathématiques, Université Clermont Auvergne.

E-mail : [anuom\\_hawa@hotmail.com](mailto:anuom_hawa@hotmail.com) and E-mail : [Anne-francoise.Yao@math.univ-bpclermont.fr](mailto:Anne-francoise.Yao@math.univ-bpclermont.fr)

## Introduction

Djibouti est un petit pays de l'Afrique de l'Est (environ  $23000\text{km}^2$ ), bordé par la mer rouge d'une part et de l'océan indien d'autre part [Chiré \(2012\)](#). Situé à la limite de l'équateur, son climat est chaud, aride et sec sur quasi les  $\frac{3}{4}$  de l'année (saison chaude). Durant cette période de l'année, les températures peuvent atteindre les 48 degrés celsius, provoquant ainsi des sécheresses et des déplacements des populations nomades vers la capitale (Sédentarisation de la population).

La population Djiboutienne est concentrée à 85% dans la capitale car c'est le siège de toutes les administrations et biens sûr de son atout majeur le port. De part sa position, le port de Djibouti est le plus important de la région et joue un rôle indispensable dans le commerce international puisque toutes les lignes maritimes reliant l'Europe à l'Asie y transitent.

Depuis les années 2000, des forces militaires sont installées à Djibouti afin d'assurer la sécurité du commerce maritimes et la lutte contre le terrorisme. De plus, Djibouti est une terre d'accueil car c'est un pays stable dans une région sous tension et au climat très difficile. Beaucoup de personnes transitent dans le pays pour des raisons commerciales, militaires, touristiques et biens sûr fuyant les guerres, famines et sécheresses. Ces individus constituent une grande mixité en termes d'épidémiologie car ils peuvent être porteurs de tout type de pathologie.

Djibouti a connu une importante vague de réfugié survenue en 1980 à cause des guerres. Cela a conduit à un pic d'urbanisation incontrôlée à Djibouti. Une stratégie anti-vectorielle a été mise en place pour lutter contre la malaria. Ce phénomène était quasi rare de 1973 à 1978 mais à partir de 1989, le pays a connu une éclosion avec plus de 3000 cas confirmés et en 1991, 7338 cas ont été déclarés. Donc, pour lutter contre cette maladie une stratégie anti-vectorielle est appliquée ainsi qu'un programme nationale de lutte contre le paludisme en 2006 ayant pour but l'interruption de la transmission de la malaria [KHAIREH et FARAH \(2012\)](#).

Les fièvres d'origines inconnues sont des pathologies assimilées aux arboviroses qui sont des affections transmises par des *arthropodes hématophages* constituant l'essentiel des vecteurs. Ces maladies vectorielles peuvent provoquer des épidémies touchant une partie importante de la population africaine, entraînant ainsi une morbidité et une mortalité élevée. [Andayi \(2009\)](#).

En épidémiologie, une méthode de quantification du risque est la méthode de la cartographie du risque. Elle permet d'estimer et de représenter le risque associé à la survenue de maladies, sur des échelles de temps et d'espace importantes [COLY \(2015\)](#), [Mouhaddach et al. \(2015\)](#), [Mason et al. \(1975\)](#), [Lawson et Zhou \(2005\)](#) et [Lowe et al. \(2011\)](#). Son rôle est d'identifier les régions présentant des risques significativement plus faibles ou plus élevés, et d'investiguer sur les raisons de ces différences.

Dans cet article, les outils comme **Excel**, **R** et **Qgis** ont été utilisés lors des différents analyses afin de réaliser une carte des cas de Djibouti puis, pour une visualisation d'une structure évidente du risque, on a calculé le taux de mortalité standardisé de la maladie afin de le projeter sur la carte de Djibouti-ville.

Par la suite, sans se limiter par une simple représentation des cartes de risque, nous allons réaliser une exploration des données à travers une analyse en composante principale qui nous permettra en premier lieu de représenter la répartition des variables (risques sociodémographiques). Ensuite, des méthodes de régressions seront effectuées dans le but de comprendre les variables expliquant le risque moyen et de constituer un modèle par la suite [Sanitaire \(2016\)](#).

## 1. Présentation des données

La période d'étude s'étend de 2007 à 2016 et les structures d'accueil sont les Centres de Santé. Le critère d'inclusion dans la base de données est de présenter une fièvre de plus de 38 degrés et de la confirmer par un test sanguin.

L'enregistrement se fait au premier niveau de la chaîne de soins puis les informations sont regroupées par les services de la Direction de l'Information Sanitaire et transmises à l'Institut National de Santé Publique de Djibouti.

### 1.1. Présentation des Structures Sanitaires

La ville de Djibouti est découpée en 3 grandes communes qui se subdivisent pour former 92 secteurs (quartiers ou zones) [Sanitaire \(2016\)](#) et [Urbanisme \(2015\)](#). Ces Zones se distinguent entre-elles selon s'elles qui sont habitées, habitables ou non ; nous les noterons comme étant nos Zones d'études ( $Z_i$ ).

Le système sanitaire à Djibouti est subdivisé en 3 niveaux :

- Niveau 1 : Centres de Santé Communautaire.
- Niveau 2 : Centres Médicaux Hospitaliers.
- Niveau 3 : Hôpitaux de références.

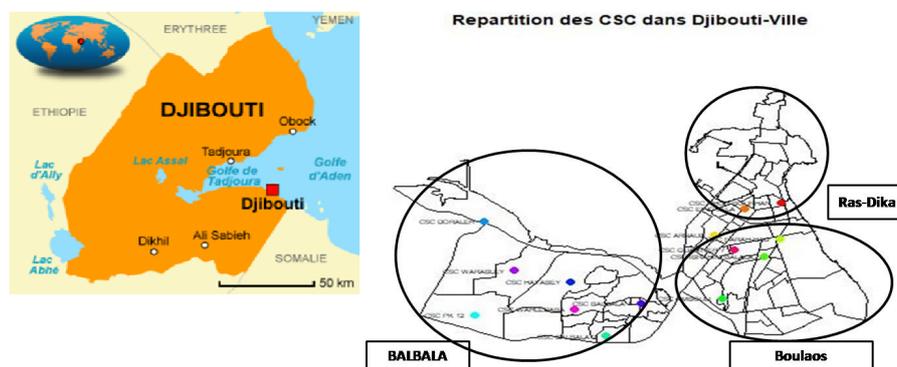


FIGURE 1. Répartition des Centres de Santé dans la capitale.

Strate ( $S_j$ )	Nombre de Zones à couvrir(quartiers)	Commune	Population d'accueil ( $P_j$ )
1- CSC HAYABLEH	6	BALBALA	80702
2 - CSC DORALEH	1	BALBALA	1434
3 - CSC PK 12	1	BALBALA	43020
4 - CSC BALBALA 2	4	BALBALA	64159
5- CSC AMBOULI	8	BOULAOS	33955
6 - CSC IBRAHIM BALBALA	5	BOULAOS	25795
7- CSC FARAH HAD	6	BOULAOS	56686
8 - CSC ARNAUD	3	BOULAOS	35577
9 - CSC EINGUELA	6	BOULAOS	37556
10 - CSC KHOR BOURHAN	9	BOULAOS	25129
11 - CSC BALBALA 1	6	BALBALA	69755
12 - CSC WARABLEH	3	BALBALA	22353
13 - CSC WAHLEDABA	6	BALBALA	38062
14 - CSC QUARTIER 7	6	BOULAOS	42508

TABLE 1. Les zones de recouvrement par les différents centres de santé communautaires

Les Zones sont référencées sous des Centres de Santé Communautaire (CSC). Ces Centres de Santé Communautaires sont au nombre de 14 et regroupent plusieurs Quartiers en termes des soins ; nous les noterons comme étant nos Strates d'études ( $S_j$ ). La présentation des ces différents centres de santé communautaires et leurs zones de couvertures sont représentées dans le **Tableau 1**.

## 1.2. Estimation du Risque

Dans cette étude les fièvres d'origines inconnues peuvent être considérées comme des phénomènes aléatoires pouvant toucher n'importe quel individu, étant donné, qu'il n'y'a apparemment aucun critère de sélection.

Une définition des entrants et une configuration particulière est proposée pour les données à savoir :

- $P_{i,j}$  la Population dans la Zone  $i$  de la Strate  $j$ .
- $O_{i,j}$  le cas Observé dans la Zone  $i$  de la Strate  $j$ .
- $P_j$  la Population Cumulée de la Strate  $j$ .
- $O_j$  les Observations Cumulées de la Strate  $j$ .
- $P_+$  la Population Totale.
- $O_+$  le Nombre total de Cas Observé.
- $r_+$  le Risque total.

Cette configuration est proposée par [Fagot-Campagna et al. \(2016\)](#), [Meyer et al. \(2017\)](#) et [Pebesma \(2012\)](#) et il s'agit de comparer le nombre de cas Observé ( $O_{i,j}$ ) à un nombre de cas espéré ( $E_{i,j}$ ).

- Si  $P_j$  et  $O_j$  connus, alors  $E_j = P_j \times r_+$  avec  $r_+ = \frac{O_+}{P_+}$ .
- S'il s'agit de cas en configuration sous Stratifiée,  $r_j = \frac{\sum_i^n O_{i,j}}{\sum_i^n P_{i,j}}$ ,  $E_j = \sum_i^j P_{i,j} \times r_j$ .

Soumis au Journal de la Société Française de Statistique

File: Soumission\_text\_long\_ABDI\_KHAIRE\_Mohamed\_Medecine\_Epidemiologie\_SFDS\_2019.tex,

compiled with jsfds, version : 2018/06/13

date: 12 avril 2019

Année	$S_1$	$S_2$	$S_3$	$S_4$	$S_5$	$S_6$	$S_7$	$S_8$	$S_9$	$S_{10}$	$S_{11}$	$S_{12}$	$S_{13}$	$S_{14}$
2007	203	7	519	63	27	17	173	63	14	22	84	—	—	—
2008	16	299	332	119	62	—	234	58	9	65	693	—	—	—
2009	12	—	60	14	36	19	8	33	1	319	28	—	—	—
2010	9	30	76	58	—	517	13	194	117	213	36	—	—	—
2011	—	86	279	42	—	164	15	284	210	74	112	—	—	—
2012	39	455	37	401	153	380	113	528	393	40	98	—	—	—
2013	7	236	50	14	93	427	21	192	427	34	121	—	—	—
2014	12	722	328	46	129	1216	331	287	1298	767	443	219	29	—
2015	12	683	85	67	150	1456	931	462	800	654	431	731	131	15
2016	87	34	102	20	840	144	1797	541	929	509	27	18	28	64

TABLE 2. Répartition des cas sur les 10 années selon les numéros des CSC.

Le tableau ci-dessus représente les états fébriles recensés sur une période de 10 années dans les différentes strates ; nous pouvons noter la présence de valeurs manquantes dans nos données. Cela s'explique par le fait que certaines strates ont été constituées récemment et que parfois il n'y a simplement aucun cas à déclarer.

Nous utiliserons par la suite une méthode de computation afin de générer ces valeurs pour compléter notre tableau. Pour finir, on peut calculer le Standardised Mortality Ratio (SMR) afin d'avoir un lissage sur les estimations ; il est obtenu comme-suit :  $SMR_j = \frac{O_j}{E_j}$ .

Nous avons regroupé sur un tableau le nombre des cas enregistré dans chaque centre de santé communautaire ( Tableau [2]).

Nous avons pu observer que la fièvre d'origine inconnue a fortement évolué durant ces 10 dernières années (de 2007 à 2016). Le taux de mortalité standardisé (SMR) a été calculé grâce à une extrapolation des données ; ainsi nous obtenons un risque pour les 92 quartiers et pour les 10 années.

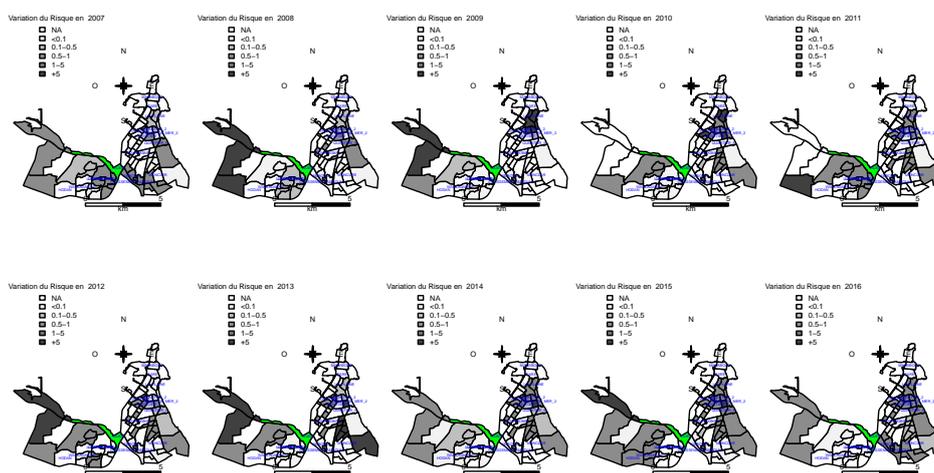


FIGURE 2. Représentation de l'évolution du risque sur les 10 ans (de gauche à droite et de haut vers le bas : 2007 à 2016).

Soumis au Journal de la Société Française de Statistique

File: Soumission\_text\_long\_ABDI\_KHAIRE\_Mohamed\_Medecine\_Epidemiologie\_SFDS\_2019.tex,

compiled with jsfds, version : 2018/06/13

date: 12 avril 2019

Les cartes sur la **Figure [2]** représentent l'évolution du risque selon les SMR dans la capitale de Djibouti durant les 10 années (2007-2016). D'une part, nous observons une structure Spatiale et de l'autre nous avons une vue Temporelle.

Le risque est très élevé pour les différents années à Doraleh et au Port de Doraleh, sauf pour les années 2010 et 2011.

En revanche, un risque s'intense est observé pour les quartiers de PK12 et de Hayableh.

Nous observons que les risques évoluent de manière diagonale du Nord Est de Djibouti au Sud Ouest pendant la période de 2007 à 2011. Puis le risque dévient très important sur toute la partie de l'Est (se situant à l'Est de de l'OUED) pour la période de 2012 à 2015.

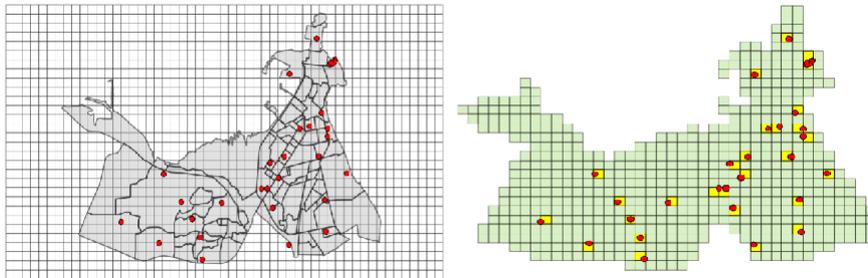


FIGURE 3. **Représentation des nids de prolifération et découpe de la carte en mailles.**

Par ailleurs, les sources d'eau, les puits et autres (éléments provoquant la propagation de la fièvre) sont appelés des nids ainsi leurs longitudes et latitudes ont été enregistrées.

Puis les distances entre les nids et les quartiers sont calculées et une codification est réalisée selon une échelle de 1 à 5.

L'enregistrement de ces codifications se fait sur une variable appelée *Niveau d'exposition* et une représentation est faite sur la **Figure [3]**.

Le fait que parfois le risque est sur dispersé dans certaines zones alors il est pertinent d'harmoniser cette représentation en considérant plusieurs facteurs d'ordre sociodémographique.

Nous avons découpé la carte de Djibouti selon une grille (maille), disposé les positions des nids de prolifération, sélectionné les zones de contact ou de proximité afin de déterminer les zones tampons.

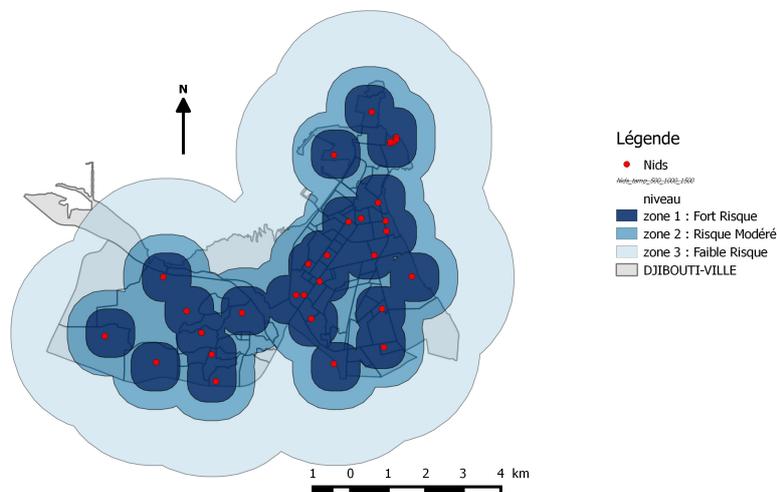
Soumis au Journal de la Société Française de Statistique

File: Soumission\_text\_long\_ABDI\_KHAIRE\_Mohamed\_Medecine\_Epidemiologie\_SFDS\_2019.tex,

compiled with jsfds, version : 2018/06/13

date: 12 avril 2019

**Représentation des zones tampons selon la proximité des nids de proliférations dans Djibouti-ville**



**FIGURE 4. Représentation des zones tampons pour les niveaux de risque.**

L'objectif de l'étude est de modéliser les fièvres à Djibouti et pour cela, il est proposé le cheminement suivant :

- Découpage de la ville en strate.
- Etude de ces states sur 10 années.
- Calcul et projection du risque sur les zones déterminées.
- Quantification des facteurs associés à l'évolution du risque.
- Analyse statistique et modélisation prédictive.

## 2. Analyses multidimensionnelles et Classification hiérarchique ascendante (CHA)

Une analyse permettant de tester le lien entre les variables caractérisant la structure des données à été proposée et le tableau suivant fait le recueil des résultats.

Variabes	V-test	P-value	Interprétation
Risque Vs exposition	87.653	$6.75 \times 10^{-12}$	Dépendants
Années Vs Strates	23656	$2.2 \times 10^{-16}$	Dépendants

TABLE 3. Test du khi-2 pour les variables d'intérêt

Nous observons un fort lien entre les variables **Risque** et **Niveau d'exposition** d'une part puis d'autre part entre **Années** et **Strates**.

Par la suite, le package missMDA a été utilisé car il est dédié aux valeurs manquantes dans l'analyse exploratoire de données multivariées et particulièrement pour l'imputation unique en ACP ou en AFC.

Il permet d'estimer le nombre de dimensions nécessaires pour imputer les données par Validation Croisée avec le critère "one leave out". L'avantage de cette méthode d'imputation est que les liens entre les variables et les similitudes entre les individus sont pris en compte.

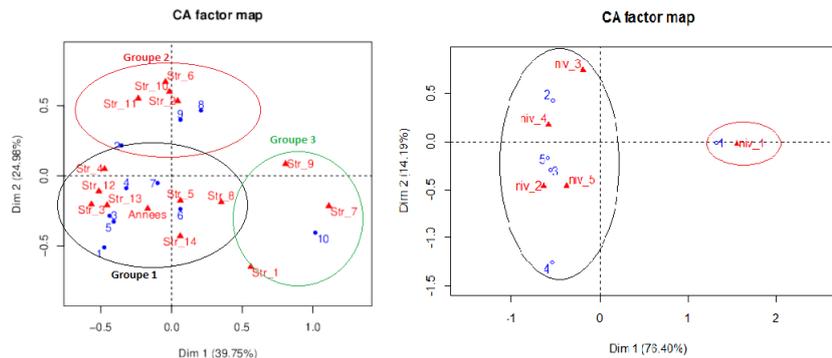


FIGURE 5. Projections du lien entre les variables d'intérêt.

Pour le graphique de droite, on visualise les groupes suivants :

- les années 2007 à 2013 regroupent les strates 3,4,5,8,12,13 et 14.
- les années 2014 et 2015 regroupent les strates 2, 6, 10 et 11.
- l'année 2016 regroupe les strates 1, 7 et 9.

Pour le graphique de gauche, on visualise les groupes suivants :

- la classe du risque élevé correspond au niveau d'exposition élevé.
- les autres classes du risque correspondent aux autres niveaux d'exposition.

Soumis au Journal de la Société Française de Statistique

File: Soumission\_text\_long\_ABDI\_KHAIRE\_Mohamed\_Medecine\_Epidemiologie\_SFDS\_2019.tex,

compiled with jsfds, version : 2018/06/13

date: 12 avril 2019

Ayant quantifier le lien entre les caractéristiques de la structure de nos données (AFC), il est à présent nécessaire d'effectuer les Analyses sur les facteurs éventuellement explicatifs du risque (ACP).

Les variables sont au nombre de 16 ; d'une part nous avons les variables socio-démographique et d'autre part les 10 variables du risques (taux de mortalités standardisés calculés précédemment) ; sans oublier la variable du niveau d'exposition. Les 92 quartiers représentent les individus. Les variables sont la plupart corrélées entre elles, la matrice de corrélation est visualisée dans la **Figure [6]**.

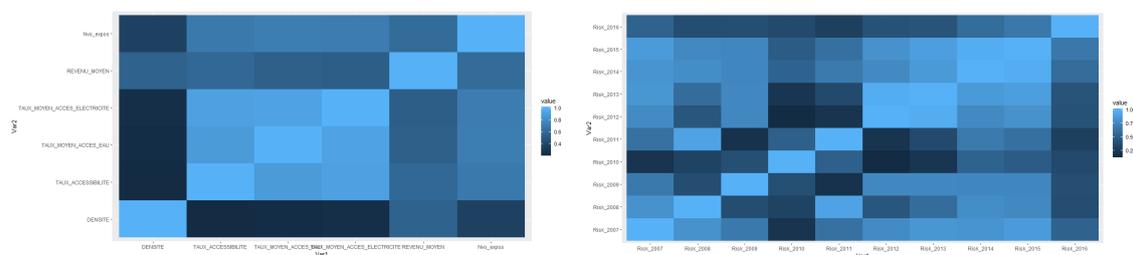


FIGURE 6. **Matrice de corrélation.**

Cette dernière montre une corrélation répartie en deux parties, d'une part les variables socio-démographique sont toutes fortement corrélées sauf la variable *Densité* qui est faiblement corrélée avec ces variables. Et d'autre part, nous avons Les variables risques qui sont la plupart fortement corrélées.

Par ailleurs, nous observons que certains risques se regroupant entre eux et forment un troisième groupe ; ce sont les risques des années 2011 et 2008. Puis nous observons que les risques de l'année 2010 et 2016 sont faiblement corrélés avec les autres.

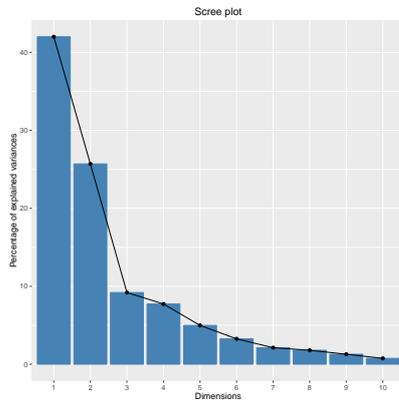
Le nombre de variables formant le troisième groupe ne sont pas très important dans ce cas, nous déduisons que le nombre des dimensions peuvent être réduit à deux.

Soumis au Journal de la Société Française de Statistique

File: Soumission\_text\_long\_ABDI\_KHAIRE\_Mohamed\_Medecine\_Epidemiologie\_SFDS\_2019.tex,

compiled with jsfds, version : 2018/06/13

date: 12 avril 2019



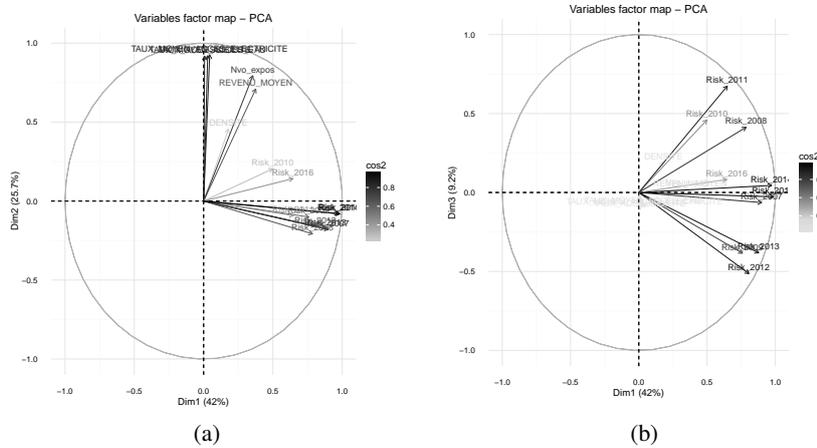
(a)

Variable	Cos <sup>2</sup>			Contrib		
	Dim.1	Dim.2	Dim.3	Dim.1	Dim.2	Dim.3
DENSITE	0.3257393	0.000821464	0.080829697	4.741556	0.02604966	5.0919484
TAUX_ACCESSIBILITE	0.2317043	0.668509668	0.011847592	3.372756	21.19928716	0.7463510
TAUX_MOYEN_ACCES_EAU	0.2925969	0.590698764	0.002629207	4.259125	18.73180499	0.1656296
TAUX_MOYEN_ACCES_ELECTRICITE	0.3261389	0.555167522	0.006753430	4.747373	17.60506436	0.4254391
REVENU_MOYEN	0.4238837	0.190283762	0.027199098	6.170175	6.03413878	1.7130566
Nvo_expos	0.3994443	0.359367752	0.002206498	5.814428	11.39600599	0.1390006
Risk_2007	0.2210094	0.001673018	0.317609934	3.217077	0.05305350	20.0081584
Risk_2008	0.4728183	0.137116572	0.222389080	6.882481	4.34813993	14.0096245
Risk_2009	0.2176576	0.035445271	0.183944632	3.168287	1.12401437	11.5877778
Risk_2010	0.4257605	0.051278117	0.346371414	6.197494	1.62609394	21.8200168
Risk_2011	0.4541736	0.066679919	0.119181423	6.611084	2.11450456	7.5079541
Risk_2012	0.5493575	0.124023507	0.009578791	7.996608	3.93294227	6.034256
Risk_2013	0.5706182	0.097012457	0.002668583	8.306085	3.07638772	0.1681101
Risk_2014	0.7279828	0.108122461	0.063677323	10.596731	3.42870000	4.0114172
Risk_2015	0.7802011	0.104652568	0.003623662	11.356836	3.31866531	0.2282763
Risk_2016	0.4507950	0.062600702	0.186897776	6.561904	1.98514744	11.7738140

(b)

FIGURE 7. (a) Courbe de coude et (b) Résultats de l'ACP.

Cependant, la Figure [7] qui visualise le graphique de coude de l'ACP montre un fort décrochage au deuxième axe, ce qui signifie que nous pouvons aussi prendre trois axes pour l'analyse. Le choix des axes par le graphique est aussi utilisé dans FRONTIER (1976) et Bertrand (2004)).



(a)

(b)

FIGURE 8. (a) Projection des variables sur le premier plan factoriel (1,2) et (b) Projection des variables sur le second plan factoriel (1,3).

Nous avons tenté de visualiser les données en faisant abstraction sur sa structure d'espace et de temps. Les cercles de corrélations pour les plans factoriels (1,2) et (1,3) sont représentés sur la Figure [8].

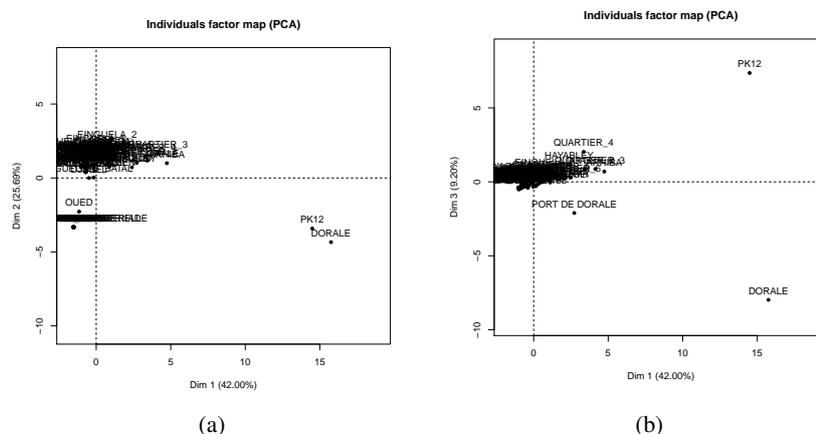


FIGURE 9. (a) Projection des individus sur le premier plan factoriel (1,2) et (b) Projection des individus sur le second plan factoriel (1,3).

Les variables socio-démographique sont toutes positivement et fortement corrélées sur le deuxième axe sauf pour la Densité et le Revenu. Le Niveau d'exposition est corrélée positivement mais moyennement.

La majorité des risques sont corrélés positivement et fortement sur le premier axe sauf les risques des années 2011, 2008, 2013, 2012 et 2009 qui présentent un cos-carrée dans les environs 75% sur le troisième axe. Ce dernier axe oppose les risques 2011, 2008 aux risques 2013, 2012, 2009 (voir la **Figure [7]**).

Nous observons une structure de répartition semblable à celle montrée par la matrice de corrélation, les variables risques sont, la plupart, isolées sur le premier axe et les autres variables socio-démographique sur le deuxième axe tandis que la variable Densité est très faiblement corrélée sur le deuxième axe.

La projection des individus est représentée sur la **Figure [9]** pour les différents axes et la majorité des individus sont regroupés ensemble et sont difficiles à lire à cause des individus Port et Doraleh que nous considérons atypique.

Puis nous relançons l'ACP et choisissons de travailler sur deux axes simplement à cause de la répartition visualisée des variables dans le cercle de corrélation car aucun individu contribue ou est corrélés avec le troisième axe.

Par la suite, nous appellerons l'axe 1 ; l'**axe des risques de prolifération de l'épidémie** ; et l'axe 2 ; l'**axe socio-démographique**.

Soumis au Journal de la Société Française de Statistique

File: Soumission\_text\_long\_ABDI\_KHAIRE\_Mohamed\_Medecine\_Epidemiologie\_SFDS\_2019.tex,

compiled with jsfds, version : 2018/06/13

date: 12 avril 2019

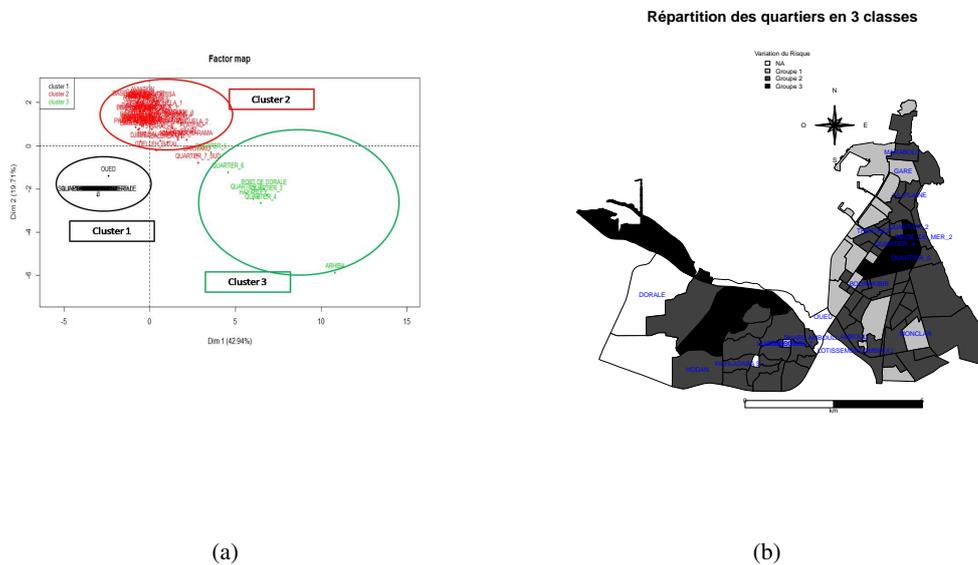


FIGURE 10. (a) Les classes d'individus projetées sur le plan factoriel (1,2) et (b) Le risque moyen de 2007 à 2016 classifié en 3 classes puis projeté sur la carte de Djibouti-Ville.

Après, nous passons à une classification hiérarchique ascendante dans le but de classer les individus et nous avons choisis trois classes.

Nous procédons ensuite à la projection de ces classes du risque sur une carte de Djibouti-Ville (voir la Figure [10]).

Les individus sont structurés en 3 groupes partageant des caractéristiques communes décrits comme-suit :

- **Le groupe 1** s'oppose aux risques contribuant à la conception de l'axe 1. Le groupe n'est touché presque par aucune épidémie et représente les quartiers les plus aisés de Djibouti-Ville.
- **Le groupe 2** représente en partie les quartiers de la commune de Boulaos et de Balbala se rejoignent dans l'exposition aux vecteurs de prolifération de la maladie, au revenu moyen, à l'accessibilité aux quartiers, le taux moyen d'accès aux ressources telles que l'électricité et Niveau d'exposition.
- **Le groupe 3** représente en partie les quartiers de la commune de Balbala et de Boulaos s'opposent aux variables corrélés avec le deuxième axe. Ce groupe représente en partie les extrêmes, c'est à dire qu'il regroupe les quartiers les plus denses, non-accessibles, ceux du centre-ville et ceux des zones industrielles.

Par exemple, le quartier Arhiba qui est le plus éloigné des individus du groupe3 : il est très peuplé, difficilement accessible ; ce groupe est quasi constant en terme d'épidémies car pratiquement toutes les épidémies le touche.

### 3. Méthodes de régressions

Dans cette section, nous allons utiliser différentes méthodes régressives. La première est la régression linéaire simple ou multiple afin de déterminer les variables qui pourraient expliquer le risque.

La seconde est une régression PLS, elle sera complémentaire à l'analyse en composante principale. Elle permet de choisir le nombre des composantes avec un apprentissage et nous permettra ainsi de déduire le nombre de variables à utiliser.

La troisième est l'analyse de la variance VASEY et THAYER (1987) et Faraway (2002) qui expliquera la variable *Risque moyen* par les autres variables déterminées.

Nous joindrons à cela une comparaison de moyenne par le test de Tukey HSD afin de voir distinguer s'il y'a similitudes dans les groupes.

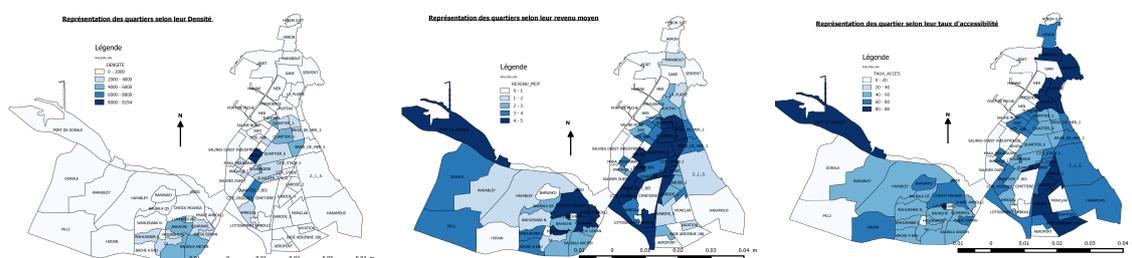


FIGURE 11. carte selon les caractéristiques sociodémographiques.

Pour ces analyses, une récodification de toutes les variables explicatives du modèle a été réalisée.

Les modalités de chacune des variables est au nombre de trois pour avoir un modèle équilibré et complet (Faible, Moyen et Fort).

Pour la récodification nous avons regardé les valeurs de chacune des variables en se référant à aux effectifs par quartiles voir la Figure[12,11].

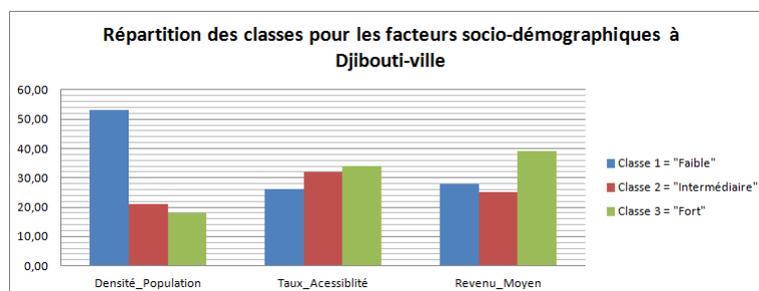


FIGURE 12. classes facteurs socios.

### 3.1. Régression linéaire

Dans cette partie, nous avons intégré toutes les variables sociodémographiques (6 variables) étudiées précédemment dans la modélisation et la méthode **StepWise** a été utilisée. Cette méthode consiste à introduire ou retirer dans le modèle (au fur et à mesure) une variable (significative) de telle sorte à tester la stabilité de ce-dernier; ainsi l'algorithme s'arrête une fois la structure du modèle déterminée.

Modèle	R <sup>2</sup>	Variables significatives
linéaire	0.28	Densité, Accessibilité, Revenu, Exposition
Logarithme	0.29	Accessibilité, Revenu, Exposition
Racine	0.39	Accessibilité, Revenu, Exposition

TABLE 4. Proposition de quelques modèles.

A travers ces résultats, nous remarquons le fait que l'on ait systématiquement les variables **Densité, Accessibilité, Revenu et Exposition**; ce qui nous oriente sur le choix de nos facteurs candidats pour la suite.

### 3.2. Régression PLS

Dans cette sous-section, nous allons approfondir le choix du nombre de variable à utiliser par la suite.

La régression par les moindres carrés partiels (PLS) est une technique qui combine les caractéristiques et la généralisation de l'analyse par composante principale (ACP) avec la régression linéaire multiple [Abdi \(2010\)](#), [WOLD \*et al.\* \(1984\)](#) et [GELADI et KOWALSKI \(1986\)](#).

La PLSR est une approches similaires pour fournir une modélisation quantitative multivariée avec des possibilités inférentielles similaires à plusieurs régression, tests t et ANOVA [Wold \*et al.\* \(2001\)](#).

Nous avons lancé la régression PLS sur le jeux de données au complet c'est à dire sur une structure de dimensions 6 par 92 d'une part (X : 6 variables sociodémographiques pour 92 quartiers) et d'une autre de dimension 1 par 92 d'autre part (Y : 1 variable à expliquer pour 92 quartiers) et ceci par le biais de la méthode **kernelpls**.

Nous obtenons ainsi les résultats suivants :

Valeur	cte	dim1	dim2	dim3	dim4	dim5
% d'inertie	—	38	80	91	92	—
CV	0.76	0.75	0.73	0.713	0.717	—

TABLE 5. Résultats de la PLS.

Nous procédons à une cross-validation par *leave-one-out* et calculons en même temps l'écart moyen (MSEP) voir la **Figure[13]**.

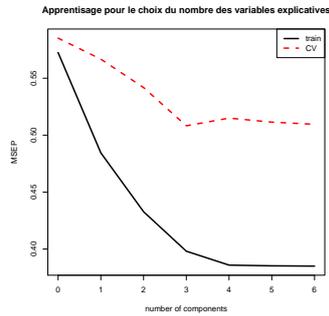


FIGURE 13. Un apprentissage par validation croisée a été réalisée sur les données.

Ces dernières ont été divisées en deux parties égales, d'une part les données pour l'apprentissage (train sur l'affichage) et d'autre part le test (appelé CV dans la fenêtre).

L'objectif de cette méthode est de trouver le nombre adéquat de composantes pour reconstruction des données estimées.

On remarque un décrochage au niveau de la troisième composante car on a une bonne représentation de l'information avec près de 91% d'inertie avec 3 composantes.

### 3.3. Analyse de variance (ANOVA) à plusieurs facteurs

Pour plus d'approfondissement, nous nous aidons de la matrice de corrélation et déduisons que les taux sont tous corrélés entre eux mais aussi avec la variable **niveau d'exposition**.

Donc, nous utiliserons un seul taux dans le modèle d'estimation, le taux d'accessibilité est très informatif par rapport aux autres.

Nous travaillerons, à présent, sur trois variables explicatives de types qualitatives. Avant de lancer l'ANOVA, une représentation des variables est effectuée sur la **Figure[14]**.

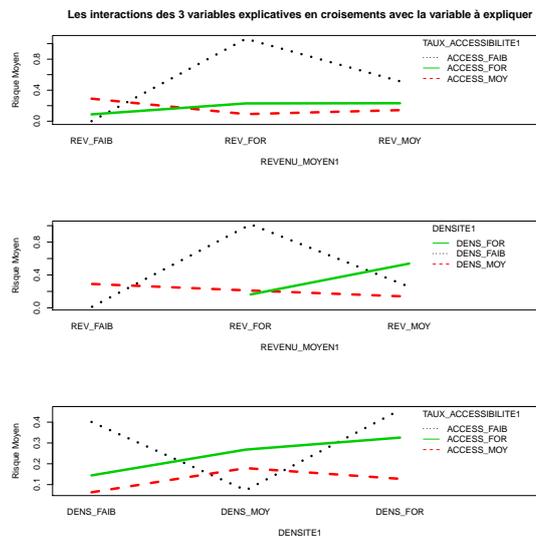


FIGURE 14. Les différentes combinaisons des interactions entre les modalités des 3 variables explicatives.

Pour chaque graphique, un croisement entre deux variables explicatives et la variable à expliquer a été réalisée.

- **1<sup>er</sup> graphe** : croisement entre les variables **Taux d'accessibilités, Revenu et Risque**. Il montre qu'un revenu fort associé à un taux d'accessibilité faible dans le quartier élève le risque moyen de prolifération de la fièvre.
- **2<sup>ème</sup> graphe** : Croisement entre les variables **Densité, Revenu et Risque**. Il montre qu'une densité faible avec un niveau de revenu fort modère le risque moyen.
- **3<sup>ème</sup> graphe** : Croisement entre les variables **Taux d'accessibilités, Densité et Risque**. Nous observons qu'une forte densité avec un taux d'accessibilité faible augmente le risque de prolifération.

Dans l'estimation du modèle, nous allons faire intervenir l'interaction de toutes les variables et regarder la significativité de chacune des variables.

Un regroupement des différents résultats de l'analyse de variance a été faite dans le **Tableau[6]**.

Soumis au Journal de la Société Française de Statistique

File: Soumission\_text\_long\_ABDI\_KHAIRE\_Mohamed\_Medecine\_Epidemiologie\_SFDS\_2019.tex,

compiled with jsfds, version : 2018/06/13

date: 12 avril 2019

Noms des variables	Mean	F-test	P-v	Signif
<b>Revenu Moyen</b>	<b>1.40854</b>	<b>4.7246</b>	<b>0.0118040</b>	<b>*</b>
<b>Taux Accessibilité</b>	<b>1.71087</b>	<b>5.7387</b>	<b>0.0048694</b>	<b>**</b>
Densité	0.74801	2.5090	0.0884364	.
<b>Revenu Moyen : Taux Accessibilité</b>	<b>1.17277</b>	<b>3.9338</b>	<b>0.00605667</b>	<b>**</b>
<b>Revenu Moyen : Densité</b>	<b>1.84605</b>	<b>6.1922</b>	<b>0.0033002</b>	<b>**</b>
<b>Taux Accessibilité : Densité</b>	<b>2.25053</b>	<b>7.5489</b>	<b>3.892e-05</b>	<b>***</b>
<b>Revenu Moyen : Taux Accessibilité : Densité</b>	<b>2.02871</b>	<b>6.8049</b>	<b>0.0004206</b>	<b>***</b>

TABLE 6. Tableau des résultats du test de l'ANOVA.

Toutes les variables introduites sont significatives à 5% ainsi les interactions entre les variables sont significatives et la qualité du modèle est acceptable avec un  $R^2$  de près de 60%. Pour vérifier cette qualité du modèle nous avons représenté l'ajustement du modèle (voir Figure[15]).

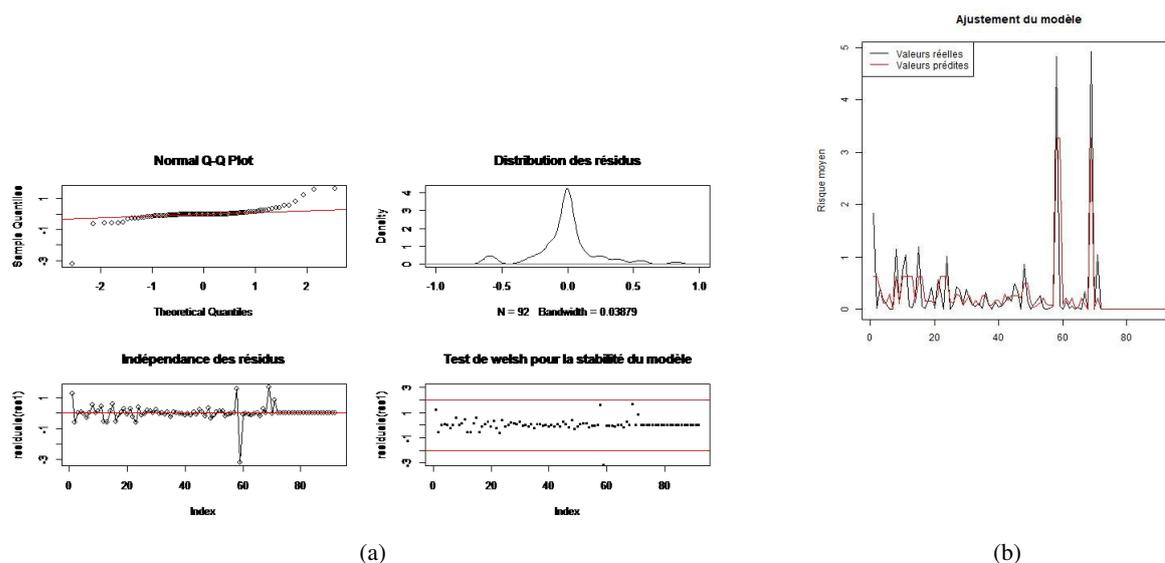


FIGURE 15. (a) Tests de la qualité du modèle et (b) L'ajustement du modèle sur les données réelles.

Par conséquent, les variables explicatives du modèle sont bien **taux d'accessibilité, revenu et densité** afin d'estimer le risque dans les différents quartiers.

Nous pouvons observer que le modèle choisit s'ajuste bien avec la variable à expliquer **Risque moyen**.

Un test de Tukey HSD est réalisé pour affiner plus l'interprétation de l'analyse de variance afin de voir si les différents groupes du modèle ont des moyennes différentes.

Les comparaisons sont regroupées dans le **Tableau[7]**.

Soumis au Journal de la Société Française de Statistique

File: Soumission\_text\_long\_ABDI\_KHAIRE\_Mohamed\_Medecine\_Epidemiologie\_SFDS\_2019.tex,

compiled with jsfds, version : 2018/06/13

date: 12 avril 2019

<b>Modalités</b>	<b>p-adj</b>
Densité Moyenne - Faible	0.8766537
Densité Forte - Faible	0.8543135
Densité Forte - Moyenne	0.6531519
Taux Accessibilité Moyen - Faible	0.4993850
Taux Accessibilité Fort - Faible	0.7075458
Taux Accessibilité Fort - Moyen	0.9442753
Revenu Moyen - Faible	0.0299880
Revenu Fort - Faible	0.159971
Revenu Fort - Moyen	0.7035875

TABLE 7. **Tableau des résultats du test de Tukey pour les comparaisons des moyennes.**

Le test réalisé est un test de moyenne pour comparer les différents groupes de chaque variable. Nous observons que les moyennes des modalités sont toutes égales et n'ont pas de différence significatives ainsi le test rejette l'hypothèse **d'égalité des moyennes**, sauf pour les modalités **moyen** et **faible** de la variable **Revenu** c'est à dire que le risque de prolifération est différent entre les quartiers selon les classe du revenu.

#### 4. Discussions et Perspectives

La fièvre d'origine inconnue a vraiment augmenté durant ces années dû à plusieurs facteurs économique ou environnementales.

Le manque d'infrastructures dans certains quartiers de Djibouti-Ville sont à l'origine de cette augmentation et c'est ce qui nous a poussé à étudier le risque de prolifération selon le quartier pour la dernière décennies (2007 à 2016).

Dans le but d'avoir une vue d'ensemble et de la structure évidente de l'évolution du risque selon les différents quartiers de Djibouti-ville, des cartes des risques sont réalisées.

La cartographie du risque de notre étude s'appuie sur le calcul des taux des mortalités standardisés qui sont des risques relatifs, et sont proportionnels aux nombres des cas atteints par la fièvre d'origine inconnue.

Une analyse en composante principale a été réalisée pour classifier les différents quartiers selon leurs risque de prolifération de la maladie et aussi pour comprendre la distribution des facteurs socio-démographique à notre disposition pour expliquer la forte présence ou non de la maladie dans certains quartiers.

Il en sort que les quartiers présentant des fortes risques de proliférations sont ceux avec une forte densité et ces sont les quartiers très proche du centre ville et qui manque d'infrastructure. Ces quartiers abritent parfois un nombre importants de personnes venant des pays voisins (Somalie, érythrée et éthiopie).

La croissance urbaine est exceptionnelle à Djibouti [Chiré \(2012\)](#) car c'est un territoire qui attire une très nombreuses populations pastorales en provenance des pays frontaliers et porteuses des très nombreux maladies.

Des méthode des régressions ont été effectuées pour expliquer le risque de prolifération moyennisé par les variables socio-démographiques et comprendre aussi les différentes interactions possibles entre les variables socio-démographique.

La régression PLS nous a permit de déduire que simplement trois variables (revenu moyen, taux d'accessibilité et la densité) expliquent à elles seules le risque de prolifération dans les différents quartiers avec 60%.

La [Figure\[14\]](#) montre qu'en moyenne la population présentant un revenu élevé avec un d'accessibilité du domicile faible ajouté à une densité élevée dans le quartier présente un risque important. Nous avons supposé par la suite que les fièvres suivaient un mouvement saisonnier par rapport au climat particulier de Djibouti. Par ce fait, nous avons considéré les données d'une structure de soins quelconque et nous avons réalisé une projection conjointe des températures avec l'évolution du nombre de cas sur une période de 3 années (2012-2014) voir la [Figure\[16\]](#).

Il en ressort que durant les mois correspondants aux la saisons fraîches (pluies et humidités), on note des variations considérables (pics et creux) dans le graphique représentatif du nombre de cas de fièvres.

Soumis au Journal de la Société Française de Statistique

File: Soumission\_text\_long\_ABDI\_KHAIRE\_Mohamed\_Medecine\_Epidemiologie\_SFDS\_2019.tex,

compiled with jsfds, version : 2018/06/13

date: 12 avril 2019

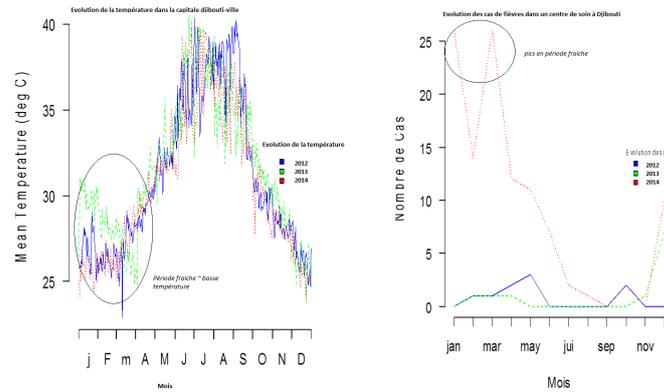


FIGURE 16. **Evolution des températures et du nombre de cas entre 2012 et 2014.**

Au vu de ces observations préliminaires, une étude mêlant le climat et l'évolution des fièvres pourrait aboutir à des résultats intéressants et complémentaires dans la détermination d'un éventuel lien.

## Conclusion

Les quartiers avec des risques de prolifération importants sont expliqués par un revenu moyen très élevé avec un taux d'accessibilité très faible et une densité très élevée.

La cartographie du risque est un bon moyen de suivi de l'évolution du risque de la fièvre d'origine inconnu dans une ville comme Djibouti-ville. Elle est considérée comme un outil permettant de classer, de comparer et hiérarchiser des risques mais également d'analyser, d'évaluer et d'identifier les risques qui peuvent impacter la santé de la population.

Nous nous sommes simplement intéressés aux taux de mortalités standardisés pour réaliser la cartographie du risque. Ces analyses seront un plus pour le domaine sanitaire ainsi que les domaines sécuritaires, financiers et économiques.

Ces modèles permettront de faciliter la suivie de l'évolution des fièvres d'origines inconnues dans les différents quartiers de Djibouti ville et d'aider les responsables des santés à mettre en place des mesures des préventions mais aussi d'élever la sécurité ou la surveillance en termes sanitaires.

## Remerciements

Nous remercions les personnels du ministère de la santé qui nous ont permis de récolter le maximum des données possible ainsi que les personnels du ministère du fond d'habitation et d'aménagement.

Nous remercions particulièrement les responsables de l'Université de Djibouti et ceux de l'Université de Clermont Auvergne d'avoir construit ce partenariat fructueux qui a permis la réalisation de ce travail.

## Références

- ABDI, H. (2010). Partial least squares regression and projection on latent structure regression (pls regression).
- ANDAYI, F. (2009). Répartition des cas de fièvres avec les séroprévalences à djibouti.
- BERTRAND, R. (2004). *Analyse en composantes principales*.
- CHIRÉ, A. S. (2012). *Le nomade et la ville à Djibouti*.
- COLY, S. (2015). Méthodes spatio-temporelles de cartographie du risque pour les maladies contagieuses.
- FAGOT-CAMPAGNA, A., ROMON, I., FOSSE, S. et ROUDIER., C. (2016). Prévalence et incidence du diabète, et mortalité liée au diabète en france.
- FARAWAY, J. J. (2002). *Practical Regression and Anova using R*.
- FRONTIER, S. (1976). Etude de la décroissance des valeurs propres dans une analyse en composantes principales: Comparaison avec le modele du baton brise.
- GELADI, P. et KOWALSKI, B. R. (1986). Partial least-squares regression: A tutorial.
- KHAIREH, B. A. et FARAH, H. (2012). Towards pre-elimination of malaria in the republic of djibouti.
- LAWSON, A. et ZHOU, H. (2005). Spatial statistical modeling of disease outbreaks with particular reference to the uk foot and mouth disease (fmd) epidemic of 2001.
- LOWE, R., BAILEY, T., STEPHENSON, D., GRAHAM, R., COELHO, C., CARVALHO, M. S. et BRACELLOS, C. (2011). Spatio-temporal modeling of climate sensitive disease risk : Towards an early warning system for dengue in brazil.
- MASON, T., MACKAY, F., HOOVER, R., BLOT, W. et FRAUMENI, J. (1975). Atlas of cancer mortality for us counties, 1950-1969.
- MEYER, S., HELD, L. et HÖHLE., M. (2017). Spatio-temporal analysis of epidemic phenomena using the r package surveillance.
- MOUHADDACH, O., BEN-DAOUD, M. et VANWAMBEKE., S. (2015). Analyse spatio-temporelle de la fièvre typhoid au niveau de la ville de meknes (maroc).
- PEBESMA, E. (2012). Spacetime: Spatio-temporal data in r.
- SANITAIRE (2007-2016). *Annuaire Statistiques de Djibouti 2007-2016*.
- URBANISME (2015). *Schéma Directeur d'Aménagement et d'Urbanisme*.
- VASEY, M. W. et THAYER, J. F. (1987). The continuing problem of false positives in repeated measures anova in psychophysiology: A multivariate solution.
- WOLD, S., RUHE, A., WOLD, H. et DUNN, W. J. (1984). The collinearity problem in linear regression. the partial least squares (pls) approach to generalized inverses.
- WOLD, S., SJOSTROM, M. et ERIKSSON, L. (2001). Pls-regression: a basic tool of chemometrics.