

**REPUBLIQUE DU CAMEROUN**  
**Paix – Travail – Patrie**

-----  
UNIVERSITE DE YAOUNDE I  
Faculté des Arts, Lettres et  
Sciences Humaines  
B.P. 755  
Département de Géographie



**REPUBLIC OF CAMEROON**  
**Peace – Work – Fatherland**

-----  
THE UNIVERSITY OF YAOUNDE I  
Faculty of Arts, Letters and Social  
Sciences  
P.O. Box 755  
Department of Geography

**MASTER PROFESSIONNEL EN CARTOGRAPHIE, SIG ET  
TELEDETECTION APPLIQUEES A LA GESTION DURABLE DES TERRITOIRES**

**STATISTIQUES APPLIQUEES**

**Surveillance des maladies épidémiologiques dans les  
quartiers de la ville de Yaoundé**

*Rédigé et présenté par:*

**NDJOUNGUEP Juscar**

**05L164**

**TANKEU TCHAMOU Brice Landry**

**IDJOLOM O. CUNEGONDE**

**Enseignant**

**M.**

**Année académique 2012-2013**

## TABLE DES MATIERES

TABLE DES MATIERES .....	1
<b>INTRODUCTION</b> .....	<b>2</b>
1. Contexte de l'étude.....	2
2. Objectif de l'étude .....	2
<b>I. METHODOLOGIE</b> .....	<b>2</b>
I.1. Matériel et données.....	2
I.2. Traitement des données .....	2
I.3. Choix du type d'analyse .....	2
I.4. Choix des logiciels.....	3
I.5. Choix de l'échelle de spatialisation des données.....	3
I.6. La spécification des paramètres de l'analyse .....	3
I.6. 1. Définition des variables actives de l'analyse.....	3
I.6.3. Options d'analyse .....	4
<b>III. RESULTATS</b> .....	<b>4</b>
III.1. Analyse univariée et bivariée .....	4
III.1.1. Analyse univariées .....	4
III.1.2. Analyse bivariée.....	5
III.2. Analyse en Composantes Principales avec la cartographie et l'interprétation des plans factoriels pertinents.....	5
III.2.1. Analyse en Composantes Principales .....	5
1. Les données sont-elles factorisables ? .....	6
2. Evaluation du nombre de facteur à retenir .....	7
III.2.2. Interprétation des résultats .....	8
III.2.3 Cartographie et l'interprétation des plans factoriels pertinents .....	10
IV. La classification ascendante hiérarchisée .....	13
<b>RECOMMANDATIONS</b> .....	<b>18</b>
<b>CONCLUSION</b> .....	<b>18</b>
<b>BIBLIOGRAPHIE</b> .....	<b>19</b>

# INTRODUCTION

## 1. Contexte de l'étude

Dans les pays en voie de développement, il se pose un problème accru de la surveillance des épidémies. Les évaluations sont de faites à macro-échelle et ne peuvent permettre une intervention efficace afin de réduire la progression des maladies. Un meilleur suivi de l'évolution de ces dernières, pour qu'elle soit efficace, doit être fait à micro-échelle et doit aussi tenir compte des paramètres physiques du milieu. Car il existe des micros biotopes qui ont des spécificités particulières et des influences particulières sur la santé de la population. Ce qui permettra de prendre en compte les spécificités locales de chaque zone à étudier.

## 2. Objectif de l'étude

L'objectif est de cartographier la répartition spatiale de quelques épidémies enregistrées dans les districts de santé de la ville de Yaoundé. Ceci permettra de détecter les zones vulnérables selon les pathologies dominantes, à travers le taux des malades, afin de prendre une décision efficace.

## I. METHODOLOGIE

### *I.1. Matériel et données*

Dans le cadre de notre étude, nous disposons des données cliniques concernant un certain nombre de cas de maladies dites épidémiologiques enregistrées à l'hôpital central de Yaoundé. A travers les logiciels Excel, SPSS et ArGis, nous allons traiter et analyser ces données.

### *I.2. Traitement des données*

Le traitement des données nous permettra d'avoir les relaxations entre les variables ainsi que leur répartition spatiale. Pour cela, nous allons utiliser les logiciels d'analyse de données et un logiciel SIG.

### *I.3. Choix du type d'analyse*

- Analyse statistique descriptive (uni-variée et bi-variée)
- Analyse en Composantes Principales avec la cartographie et l'interprétation des plans factoriels ;
- L'analyse des variations observées dans le passage de 2 à 4 classes est également attendue.
- Le calcul des interactions spatiales et de l'auto corrélation spatiales

#### ***1.4. Choix des logiciels***

Comme logiciel d'analyse des données, nous allons utiliser SPSS pour les différentes analyses (univariées, bivariées, multi variées) afin d'obtenir les tableaux qui seront traité dans Excel. La logiciel Excel nous a permis de traiter les tableaux statistiques obtenues des analyses statistiques de SPSS. Quant au logiciel ArcGis, il nous a permis de spatialiser les données traitées sur SPSS.

#### ***1.5. Choix de l'échelle de spatialisation des données***

La cartographie qui se fera à une micro échelle permettra aux agents en charge de la santé de localiser directement dans chaque district de santé les quartiers les plus vulnérables à une pathologie donnée. A partir de cela, les interventions pour assainir le milieu ou pour trouver d'autres solutions seront précises et rapides.

#### ***1.6. La spécification des paramètres de l'analyse***

##### ***1.6. 1. Définition des variables actives de l'analyse***

Afin d'afficher la boîte de dialogue principale de la procédure FACTOR

- **Sélection des variables**

Il s'agit de sélectionner les variables numériques choisies pour l'ACP parmi celles figurant dans la liste source en les transférant dans la liste des **Variables** à l'aide du bouton. On obtient alors le listage de la structure initiale, la matrice des corrélations variables-facteurs et les statistiques concernant la structure finale.

- **Statistiques**

Nous allons utiliser les caractéristiques univariées qui consistent à l'affichage du nombre d'observations valide, de la moyenne et de l'écart-type pour chaque variable et la structure initiale c'est-à-dire la communauté de la solution initiale, valeurs propres et pourcentage d'inertie expliquée.

- **Matrice des corrélations**

On choisira la **corrélation** comme indicateur statistique: c'est la matrice des coefficients de corrélation pour les variables actives. L'**Indice KMO et test de Bartlett** pour la mesure de la qualité d'échantillonnage et test de sphéricité de Bartlett. Et l'**inverse** de la matrice des corrélations.

- **Extraction des facteurs**

Elle se fera par Analyse en composantes principales. Il s'agit de la méthode d'extraction par défaut qui permettra d'obtenir les Coordonnées factorielles (matrice des corrélations variables-facteurs), communautés et valeurs propres pour la structure factorielle.

- **Rotation des facteurs**

Afin de sélectionner une procédure de rotation des facteurs, aucune rotation n'est possible s'il n'y a qu'un seul facteur extrait. La standardisation de Kaiser est utilisée avec n'importe laquelle des méthodes de rotation. On choisira la rotation orthogonale selon la méthode Varimax.

### ***1.6.2. Coordonnées factorielles***

Afin de sauvegarder les coordonnées factorielles pour les réutiliser dans d'autres analyses, nous allons utiliser la méthode de Bartlett. Cette option permet d'afficher la matrice des coefficients ainsi que la matrice de variance-covariance des coordonnées factorielles.

### ***1.6.3. Options d'analyse***

Pour changer le traitement, nous allons Exclure toute observation incomplète. Seules les observations ayant des valeurs valides pour l'ensemble des variables actives sont retenues pour l'analyse.

On retiendra qu'un quartier ayant pour coordonnées négative sur l'axe 1 dispose d'une faible proportion des malades tandis qu'un quartier ayant une coordonnée positive dispose d'une forte proportion de malade. Pour l'axe 2, un quartier ayant une coordonnée négative dispose d'une faible proportion de la population et une forte proportion de malades tandis qu'un quartier ayant une coordonnée positive dispose d'une forte proportion de population et une forte proportion de malades

## **III. RESULTATS**

### **III.1. Analyse univariée et bivariée**

#### ***III.1.1. Analyse univariées***

Dans l'analyse univariée, le logiciel fournit la moyenne (mean), l'écart-type (standard déviation), le minimum et le maximum de chaque variable. Le résultat obtenu est présenté dans le tableau suivant :

**Tableau 1 : Analyse univariée des différentes variables**

	Tuberculose	Méningite	AVC	Paludisme	Tétanos	VIH	Accès palustre	Anémie
Moyenne	88,738	87,786	98,595	99,476	88,738	87,881	89,214	86,762
Ecart type	14,6823	16,2711	18,7212	16,6664	14,6823	16,3342	15,4023	15,8507
Minimum	57,0	56,0	54,0	56,0	57,0	56,0	52,0	53,0
Maximum	135,0	142,0	167,0	139,0	135,0	142,0	132,0	133,0

De ce résultat, on note une grande homogénéité dans les différentes variables considérées : même ordre de grandeur pour les moyennes, les écarts-types, les minima et les maxima.

### III.1.2. Analyse bivariée

#### Corrélation entre les variables

	Tétanos	VIH	Tuberculose	Paludisme	Méningite	AVC	Accès palustre	Anémie
Tétanos	1	,928**	1,000	,892	,928	,865	,941	,971
VIH	,928**	1	,928	,788	,999	,775	,867	,917
Tuberculose	1,000	,928	1	,892	,928	,865	,941	,971
Paludisme	,892	,788	,892	1	,789	,911	,915	,850
Méningite	,928	,999	,928	,789	1	,777	,867	,917
AVC	,865	,775	,865	,911	,777	1	,897	,864
Accès palustre	,941	,867	,941	,915	,867	,897	1	,948
Anémie	,971	,917	,971	,850	,917	,864	,948	1

\*\* . Correlation is significant at the 0.01 level (2-tailed).

## III.2. Analyse en Composantes Principales avec la cartographie et l'interprétation des plans factoriels pertinents.

### III.2.1. Analyse en Composantes Principales

L'Analyse en Composantes Principales permet d'analyser des tableaux de données numériques quantitatives pour en réduire la dimensionnalité des principaux facteurs d'interaction entre variables et en représenter graphiquement les interrelations. La mise en œuvre d'une Analyse en Composantes principales (ACP) peut être effectuée au moyen de la procédure d'Analyse Factorielle de SPSS.

#### Descriptive Statistics

	Mean	Std. Deviation	Analysis N
Pop_totale	5707,786	5501,9879	84
Pop_masc	4819,929	4646,1481	84
Pop_femin	86,786	15,2900	84
Paludisme	99,476	16,6664	84
Tétanos	88,738	14,6823	84
VIH	87,881	16,3342	84
Tuberculose	88,738	14,6823	84
Méningite	87,786	16,2711	84
AVC	98,595	18,7212	84
Accès palustre	89,214	15,4023	84
Anémie	86,762	15,8507	84

## 1. Les données sont-elles factorisables ?

Pour répondre à cette question, dans un premier temps, il convient d'observer la matrice des corrélations « Correlation Matrix ». Si plusieurs variables sont corrélées (> 0.5), la factorisation est possible. Si non, la factorisation n'a pas de sens et n'est donc pas conseillée.

**Correlation Matrix<sup>a, b</sup>**

	Pop_totale	Pop_masc	Pop_femin	Paludisme	Tétanos	VIH	Tuberculose	Méningite	AVC	Accès palustre	Anémie
Correlation Pop_totale	1,000	1,000	-,233	-,390	-,327	-,241	-,327	-,239	-,373	-,351	-,287
Pop_masc	1,000	1,000	-,233	-,390	-,327	-,241	-,327	-,239	-,373	-,351	-,287
Pop_femin	-,233	-,233	1,000	,839	,762	,705	,762	,709	,801	,838	,765
Paludisme	-,390	-,390	,839	1,000	,892	,788	,892	,789	,911	,915	,850
Tétanos	-,327	-,327	,762	,892	1,000	,928	1,000	,928	,865	,941	,971
VIH	-,241	-,241	,705	,788	,928	1,000	,928	,999	,775	,867	,917
Tuberculose	-,327	-,327	,762	,892	1,000	,928	1,000	,928	,865	,941	,971
Méningite	-,239	-,239	,709	,789	,928	,999	,928	1,000	,777	,867	,917
AVC	-,373	-,373	,801	,911	,865	,775	,865	,777	1,000	,897	,864
Accès palustre	-,351	-,351	,838	,915	,941	,867	,941	,867	,897	1,000	,948
Anémie	-,287	-,287	,765	,850	,971	,917	,971	,917	,864	,948	1,000

a. Determinant = .000

b. This matrix is not positive definite.

Dans notre exemple, plusieurs variables sont corrélées entre elles :

Dans un deuxième temps, il faut observer l'indice de KMO (Kaiser-Meyer-Olkin) qui doit tendre vers 1. Si ce n'est pas le cas, la factorisation n'est pas conseillée. Pour juger l'indice de KMO, nous avons utilisé l'échelle suivante :

- 0,50 et moins est misérable
- entre 0,60 et 0,70, c'est médiocre
- entre 0,70 et 0,80 c'est moyen
- entre 0,80 et 0,90 c'est méritoire
- et plus 0,9 c'est merveilleux.

### Tableau : Indice KMO et Bartlett

#### KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,810
Bartlett's Test of Sphericity	Approx. Chi-Square	1212,128
	df	28
	Sig.	,000

Dans notre cas le KMO est compris entre 0,70 et 0,80 donc l'échelle est moyenne donc nous pouvons continuer l'analyse.

Enfin, nous avons utilisé le test de sphéricité de Bartlett qui stipule que la signification (Sig.) tend vers 0.000, c'est très significatif, inférieur à 0.05 significatif, entre 0.05 et 0.10

acceptable et au-dessus de 0.10, on rejette. Dans notre cas il tend vers 0.000 donc il est très significatif.

## 2. Evaluation du nombre de facteur à retenir

Trois règles sont applicables :

- **1<sup>ère</sup> règle** : la règle de Kaiser qui veut qu'on ne retienne que les facteurs aux valeurs propres supérieures à 1.
- **2<sup>ème</sup> règle** : on choisit le nombre d'axe en fonction de la restitution minimale d'information que l'on souhaite. Par exemple, on veut que le modèle restitue au moins 80% de l'information.

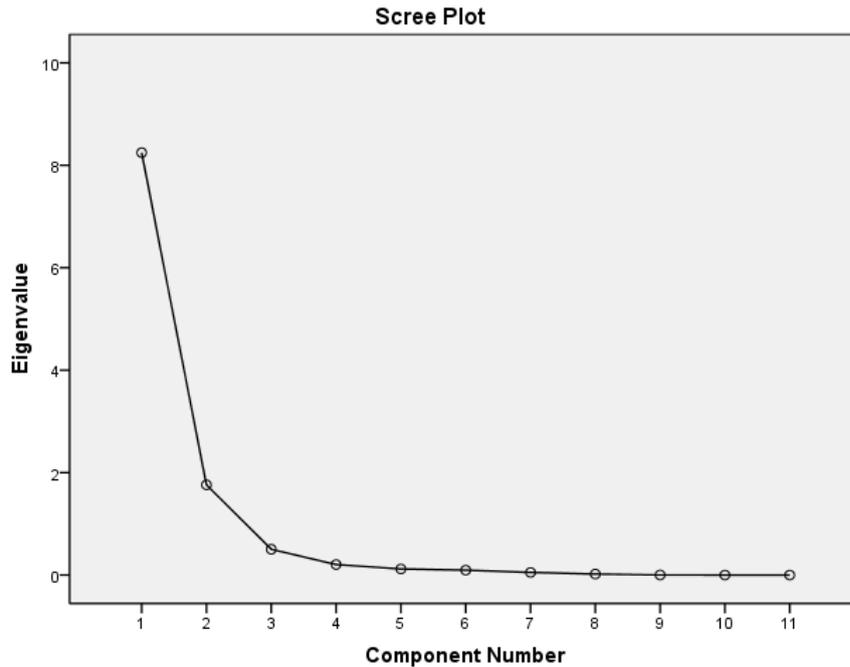
Pour ces deux premières règles, on examine le tableau « Total Variance Explained ».

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	8,247	74,975	74,975	8,247	74,975	74,975	7,787	70,790	70,790
2	1,761	16,005	90,979	1,761	16,005	90,979	2,221	20,189	90,979
3	,503	4,573	95,552						
4	,204	1,859	97,411						
5	,119	1,078	98,490						
6	,095	,864	99,354						
7	,051	,465	99,819						
8	,019	,174	99,993						
9	,001	,007	100,000						
10	2,961E-9	2,691E-8	100,000						
11	2,167E-16	1,970E-15	100,000						

Extraction Method: Principal Component Analysis.

Dans ce tableau, on voit que 2 facteurs seulement sont fortement corrélés entre eux

**3<sup>ème</sup> règle** : le « Scree-test » ou test du coude. On observe le graphique des valeurs propres et on ne retient que les valeurs qui se trouvent à gauche du point d'inflexion. Graphiquement, on part des composants qui apportent le moins d'information (qui se trouvent à droite), on relie par une droite les points presque alignés et on ne retient que les axes qui sont au-dessus de cette ligne.



### ***III.2.2. Interprétation des résultats***

Il s'agit ici de donner un sens à un axe grâce à une recherche lexicale (ou recherche de mots) à partir des coordonnées des variables et des individus. Ce sont les éléments extrêmes qui concourent à l'élaboration des axes. Après l'analyse, nous avons obtenu 2 axes. L'axe 1 à une forte corrélation avec la population féminine, le paludisme, le tétanos, le VIH, la tuberculose, la Méningite, l'AVC, l'accès palustre et l'anémie donc cet axe peut être nommé : proportion des malades divers dans la population féminine. L'axe 2 à deux coordonnées négative (paludisme et AVC) et 9 coordonnées positives, ce qui signifie une faible proportion des malades de paludisme et AVC, une forte proportion de la population et des maladies diverses (tétanos, VIH, tuberculose, méningite, accès palustre et anémie).

**Component Matrix<sup>a</sup>**

	Component	
	1	2
Pop_totale	-,416	,907
Pop_masc	-,416	,907
Pop_femin	,836	,117
Paludisme	,934	-,020
Tétanos	,977	,091
VIH	,924	,178
Tuberculose	,977	,091
Méningite	,925	,181
AVC	,919	-,009
Accès palustre	,970	,052
Anémie	,963	,129

Extraction Method: Principal Component Analysis.

a. 2 components extracted.

Dans notre exemple, ce sont les variables « Paludisme, VIH, tuberculose, Méningite » et « anémie » qui concourent le plus à la construction de l'axe 1. Et l'axe 2 est constitué des variables « populations totale, population masculine et population féminine ». Si la recherche lexicale à partir des variables ne donne rien, il faut alors donner un sens à l'axe en s'appuyant là aussi aux individus qui ont les coordonnées extrêmes.

Des axes peuvent très bien ne pas avoir de sens, auquel cas il ne faut pas les retenir. Si la variance expliquée est trop faible, on peut choisir d'exclure certaines variables. Pour choisir les variables à éliminer, on observe leur qualité de représentation : plus la valeur associée à la ligne « Extraction » est faible, moins la variable explique la variance. Le tableau suivant nous présente la ligne d'extraction avec de fortes valeurs.

**Communalities**

	Initial	Extraction
Pop_totale	1,000	,996
Pop_masc	1,000	,996
Pop_femin	1,000	,713
Paludisme	1,000	,873
Tétanos	1,000	,962
VIH	1,000	,886
Tuberculose	1,000	,962
Méningite	1,000	,888
AVC	1,000	,844
Accès palustre	1,000	,943
Anémie	1,000	,944

Extraction Method: Principal Component Analysis.

Il faut également tenir compte du positionnement de chaque variable sur chaque l'axe : les variables à éliminer sont les variables qui sont

- Soit proches du centre sur l'ensemble des axes retenus.
- Soit au milieu d'un quart de cercle sur les axes retenus.
- Soit les variables qui forment un axe à elles toute seule.

### ***III.2.3 Cartographie et l'interprétation des plans factoriels pertinents***

Pour la cartographie, nous avons retenu deux axes factoriels :

- L'axe 1 : Proportion des malades et population par quartier

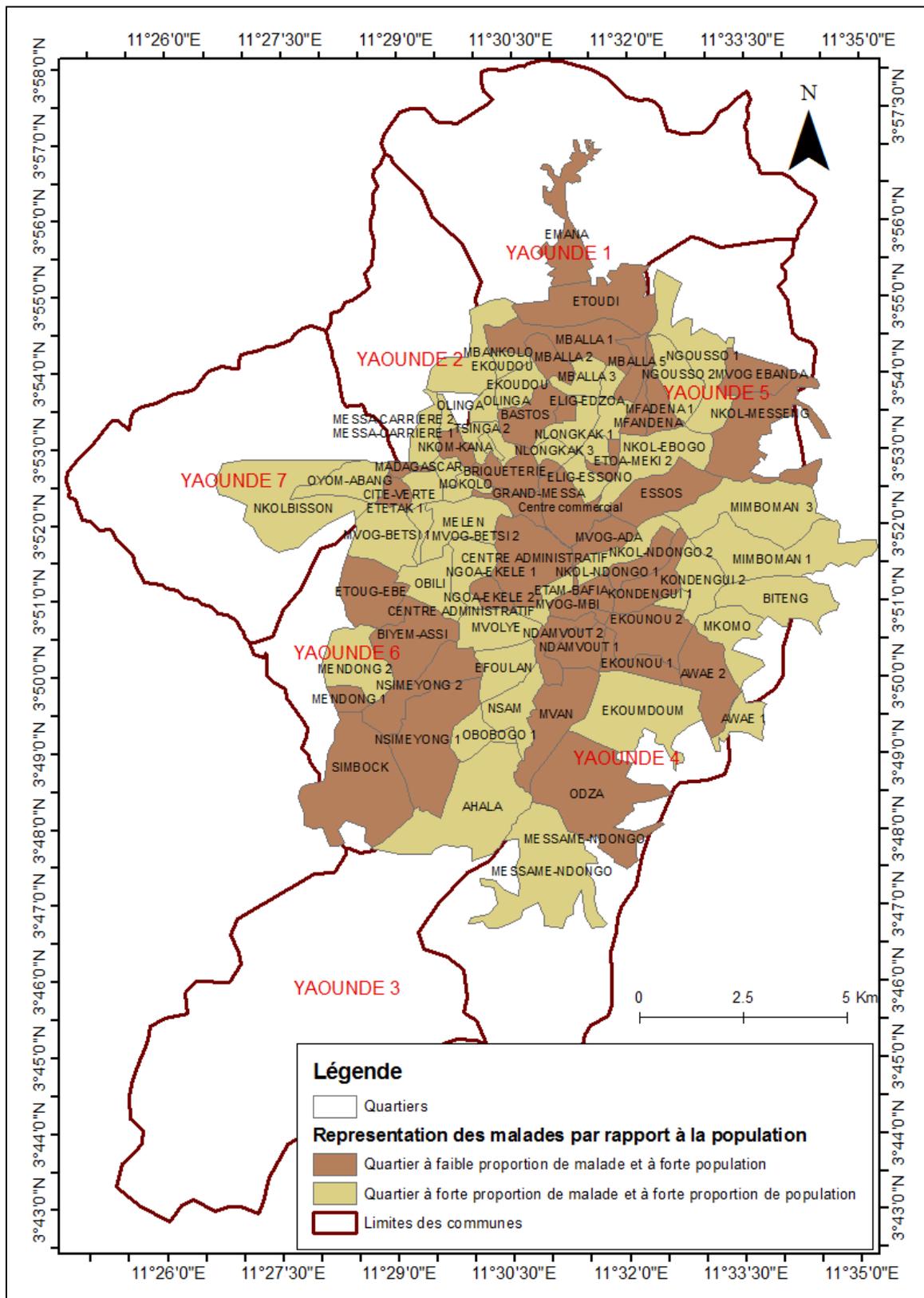
Qui comprend :

- Les quartiers à faible proportion des malades et à forte proportion de population
- Les quartiers à forte proportion de malades et à forte proportion de population.
- L'axe 2 : Proportion des populations par rapport au nombre de malade

Qui comprend :

- Les quartiers à forte proportion de population et à faible proportion des malades
- Les quartiers à forte proportion des malades et à forte proportion des malades.

Ces différents résultats ont été spatialisés sur le fond de la carte des quartiers de Yaoundé (figure 1 et 2).



**Figure 1 : Axe 1 (faible proportion de la population et forte quantité de malades)**

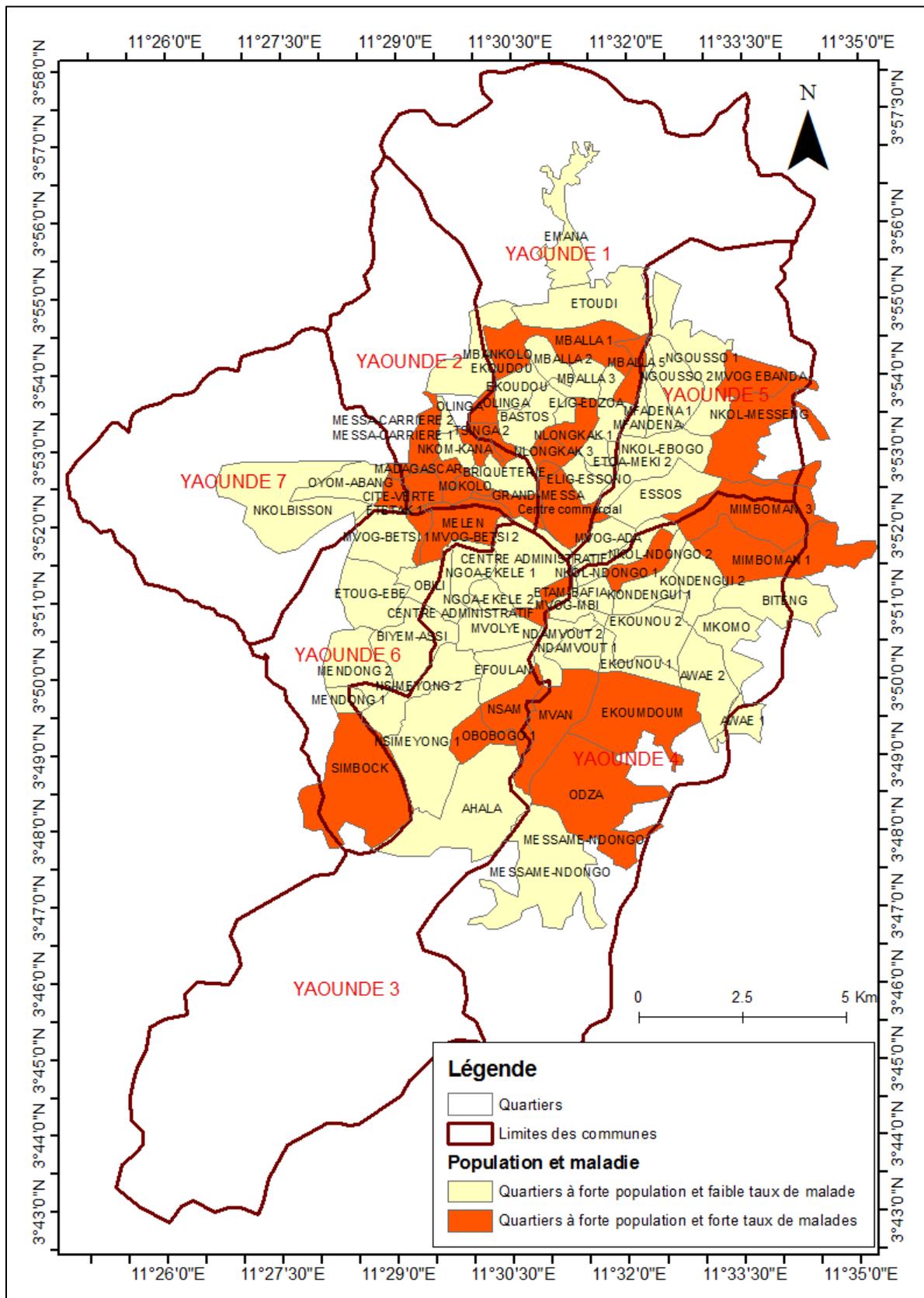


Figure 2 : Axe 2 (faible proportion de malade de paludisme et d'AVC, forte proportion de la population et forte proportion des cas des autres maladies).

#### IV. La classification ascendante hiérarchisée

L'objectif est de regrouper les quartiers en groupe homogène, ce qui permettra d'obtenir la répartition des quartiers ayant la même proportion des cas de malades ou de population.

Ceci nous a permis d'obtenir 3 classes d'informations réparties sur 3 clusters : Tableau suivant.

**Tableau** : Classes de la classification ascendante hiérarchisée

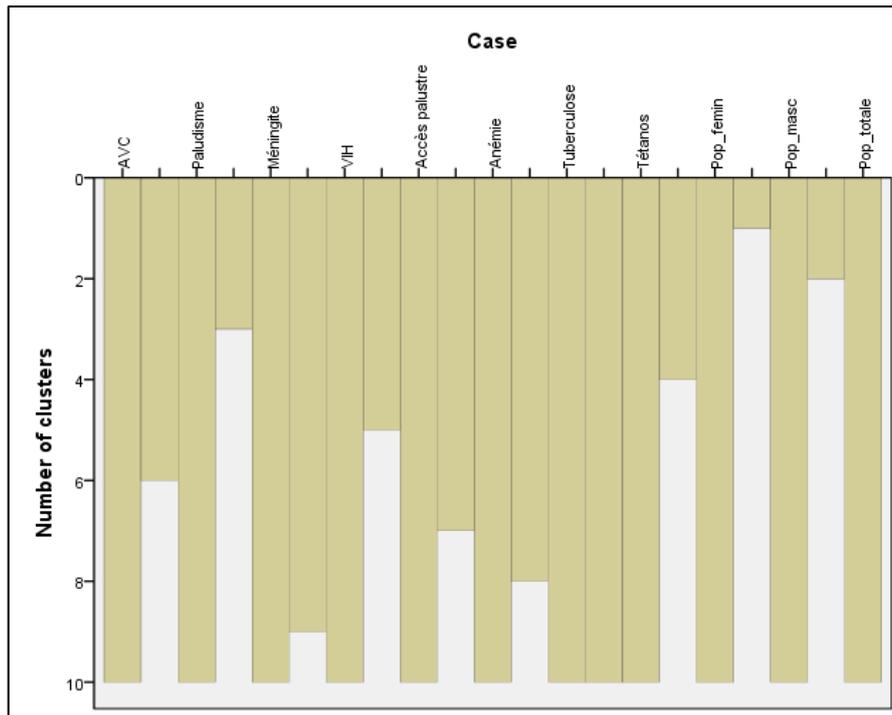
Cluster Membership			
Case	4 Clusters	3 Clusters	2 Clusters
Pop_totale	1	1	1
Pop_masc	2	2	1
Pop_femin	3	3	2
Paludisme	4	3	2
Tétanos	3	3	2
VIH	3	3	2
Tuberculose	3	3	2
Méningite	3	3	2
AVC	4	3	2
Accès palustre	3	3	2
Anémie	3	3	2

D'après le tableau, le cluster 2 comprend 2 classes (1 et 2). La classe 1 est composée de population totale et de population masculine et la classe 2 les différentes maladies. Le cluster 3 comprend 3 classes (1, 2 et 3), la classe 1 correspond à la population totale, la classe 2, population masculine et la classe 3 les différentes maladies. Le cluster 4 comprend 4 classes (1, 2, 3 et 4) où la classe 1 représente la population totale, la classe 2 la population féminine, la classe 3 le paludisme, le tétanos, la tuberculose, la méningite, l'accès palustre et l'AVC.

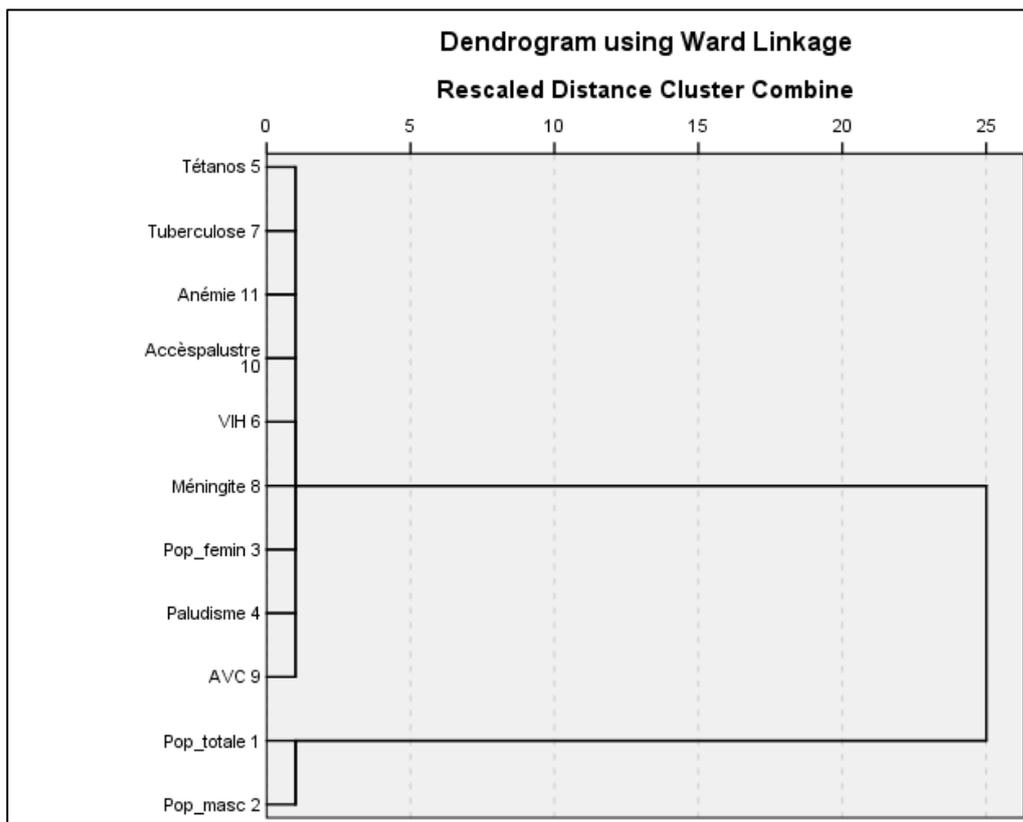
Ainsi on peut regrouper ce résultat en 3 thèmes :

- Populations et maladies dans les quartiers de la ville de Yaoundé (figure 5)
- Proportion des populations masculine et totale et répartition des maladies épidémiques dans la ville de Yaoundé (figure 6)
- Populations et localisation des quartiers ayant une certaine proportion des maladies épidémiologique (figure 7).

Affichage des variables en fonction du nombre de cluster et du dendrogramme en utilisant les liens entre les variables.



**Figure 3:** Diagramme des classes de la classification ascendante hiérarchisée



**Figure 4:** Dendrogramme de la CAH

## Cartographie des différentes classes

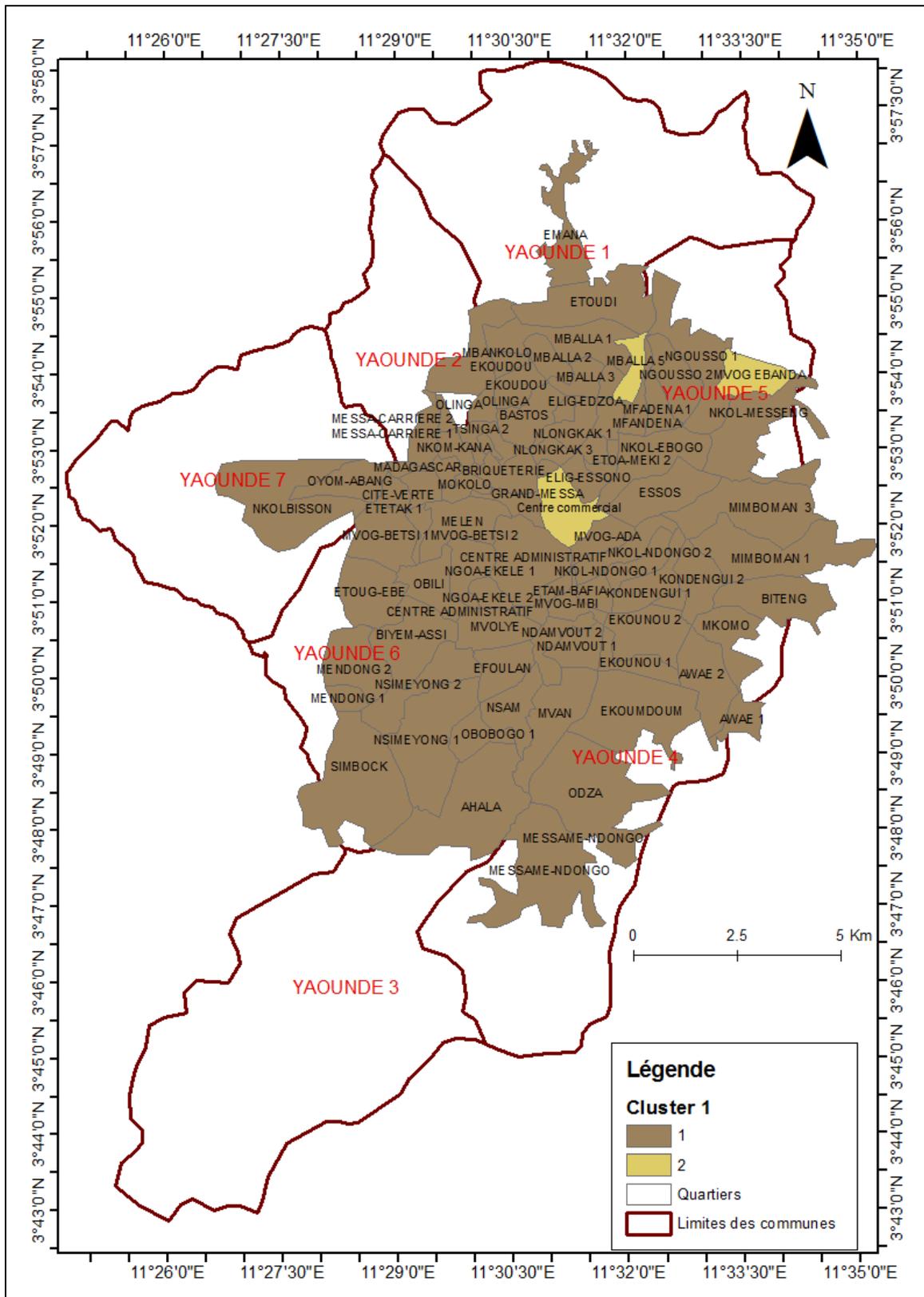


Figure 5 : Populations et maladies dans les quartiers de la ville de Yaoundé

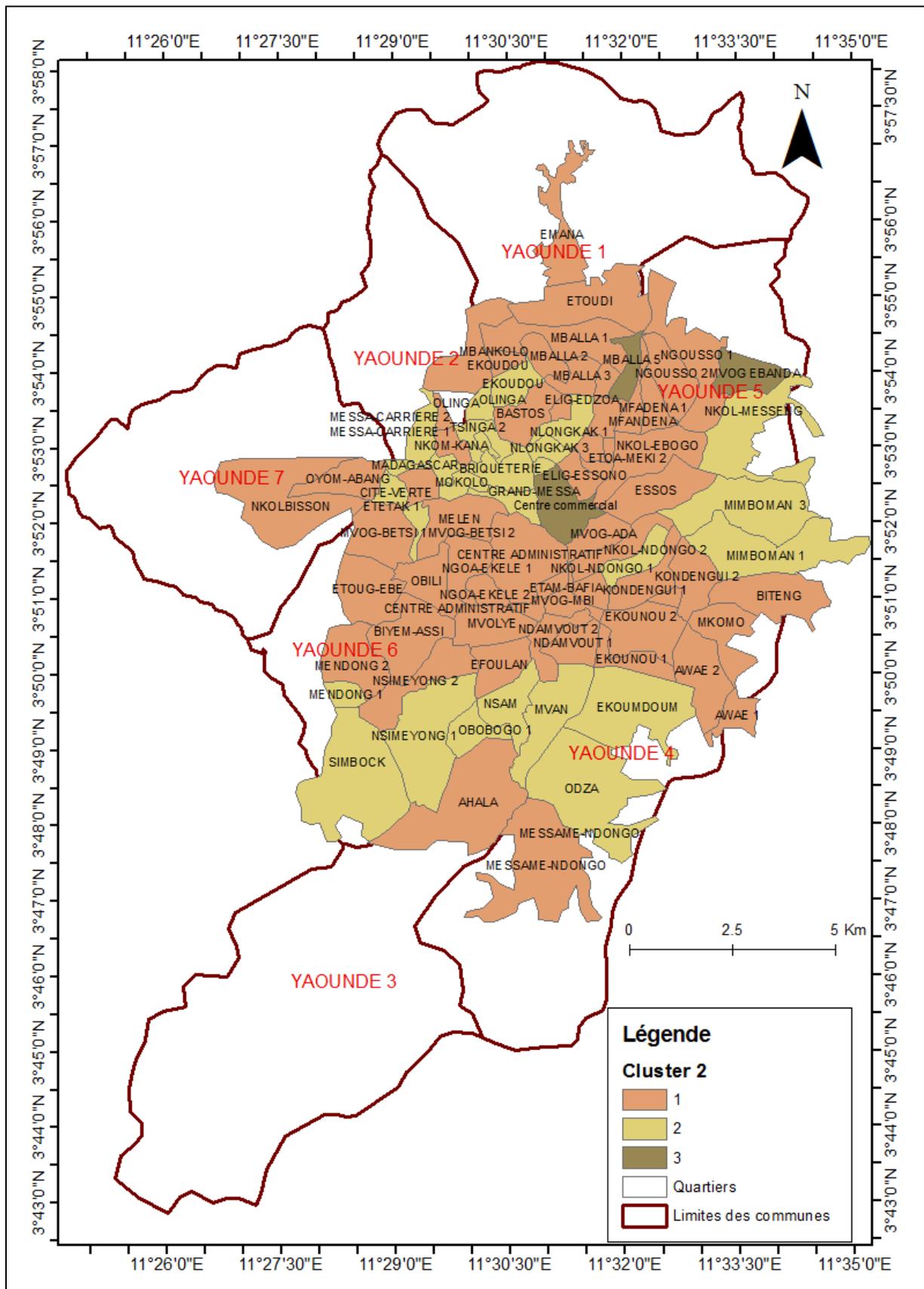


Figure 6 : Proportion des populations masculine et totale et répartition des maladies épidémiques dans la ville de Yaoundé

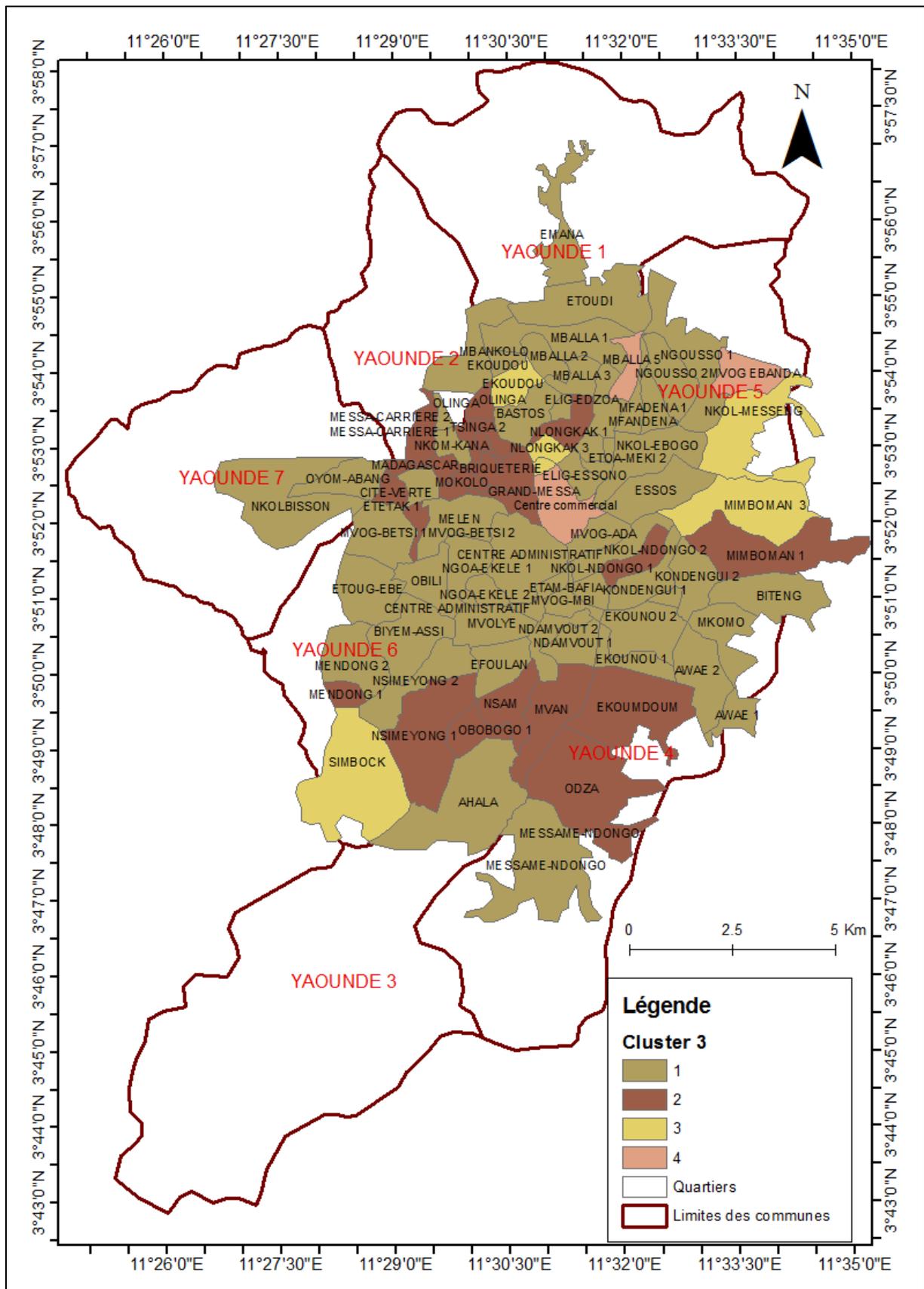


Figure 7: Populations et localisation des quartiers ayant une certaine proportion des maladies épidémiologique

La classification ascendante hiérarchisée nous a permis d'obtenir les classes homogènes des variables de notre étude. Elle nous montre la répartition des quartiers par classe homogène. La figure qui présente 4 classes permet de mieux ressortir les variables ayant les caractéristiques homogènes. Ce sont les quartiers où sévissent les mêmes caractéristiques socio-économiques et environnementales. Ces derniers concourent à la stagnation ou l'augmentation des cas de malades.

## **RECOMMANDATIONS**

L'analyse des données de la surveillance épidémiologique est un exercice minutieux. Elle demande la mise à jour permanente des données de manière à permettre à chaque fois d'observer l'état des lieux dans une entité donnée. D'où la nécessité des responsables de la santé de mettre à la disposition des décideurs des données fiables afin qu'elles contribuent à la réduction voire même l'éradication de la maladie.

## **CONCLUSION**

En somme, notre travail consistait à analyser les données sur la surveillance épidémiologiques dans les quartiers de la ville Yaoundé. Pour cela, nous avons effectué les analyses univariées, bivariées et multivariées.

Le résultat de l'analyse univariée nous présente une forte présence des malades dans les quartiers. L'analyse bivariée nous montre qu'il y a une forte corrélation entre la population et la densité des malades dans les quartiers.

L'analyse multivariée a concerné l'ACP et la CAH. La première nous a permis de spatialiser la proportion des malades dans les quartiers à travers deux axes factoriels qui nous a donné deux cartes. La CAH nous a permis de regrouper les variables en groupes homogènes en 2, 3 et 4 classes. C'est ainsi que nous avons pu quantifier la surveillance épidémiologique dans les quartiers de la ville de Yaoundé.

## **BIBLIOGRAPHIE**

**Maryse Raffestin(2005)**, Pratiquer une analyse en composantes principales, 7pages.

**Ludovic LE MOAL (2002)**, L'Analyse en Composantes Principales, 7pages.

**WHO/CDS/CSR/ISR, (2000)**, Normes recommandées par l'OMS pour la Surveillance, Deuxième édition, 172p.

**Dominique DESBOIS, (2000)**, Une introduction à l'Analyse en Composantes Principales avec SPSS pour Windows, 18p.