

« Accessibilité aux infrastructures de base et atteinte des OMD par le Cameroun »

BATOMEN Francis KAMENI Justin  
fbatomen@yahoo.fr  
CAMEROUN

### Résumé

La possibilité d'atteinte des Objectifs du Millénaire pour le Développement (OMD) par le Cameroun a été évaluée par l'accessibilité géographique aux infrastructures de base à l'aide des données des deuxième et Troisième Enquête Camerounaise auprès des Ménages (ECAM2 et ECAM3). A l'aide de la classification automatique sur les axes factoriels de l'analyse en composantes principales, une typologie des ménages à quatre classes a été construite. Elle regroupe les ménages camerounais en classes homogènes suivant l'accessibilité aux infrastructures de base en 2001 (données de l'ECAM2), permettant ainsi de mieux percevoir les privations en termes d'infrastructures de base. L'approche data mining est utilisée pour classer de nouveau les ménages au sein des groupes de cette typologie suivant l'accessibilité aux infrastructures de base en 2007 (données de l'ECAM3). La méthode de simulation bootstrap a permis de vérifier si la typologie est restée statique entre 2001 et 2007. Il en découle que les changements observés restent moins importants pour contribuer à l'atteinte des OMD. L'analyse des disparités régionales révèle l'intensification des inégalités entre les deux grandes villes Douala et Yaoundé et la partie septentrionale du pays. On retient qu'il est nécessaire de revoir la politique de développement des infrastructures de base si le Cameroun souhaite atteindre les OMD d'ici 2015. Une lutte efficace contre la pauvreté doit se traduire en une lutte contre la paupérisation du milieu rural qui est le principal grenier du pays. La voie royale c'est l'extension du réseau de routes bitumées ce qui faciliterait le développement des activités plus rémunératrices en milieu rural.

**Mots-clés :** Infrastructure de base, OMD, pauvreté, typologie, data mining, Cameroun

### 1. Introduction

Pour les pays en développement, la lutte contre la pauvreté, premier objectif des OMD est une priorité. Ils sont caractérisés par le manque de services d'infrastructure ce qui a une influence incontestable sur la pauvreté. D'après le Professeur Amartya Sen<sup>1</sup>, la pauvreté peut s'expliquer par l'absence d'accès à une infrastructure de base permettant d'obtenir une scolarisation, d'obtenir des soins de santé ou de vendre des produits au marché. De plus, l'absence d'infrastructure d'eau potable expose les populations aux maladies hydriques et le manque de services d'assainissement adéquats favorise la transmission rapide des maladies. Ainsi, examiner la répartition spatiale des infrastructures permettrait de mieux saisir les conditions de vie des ménages, et l'étude de son évolution donnerait, en quelque sorte, une idée sur la possibilité d'atteinte des OMD dont l'échéance est proche.

De surcroît, il ressort des recommandations de juin 2008 du Groupe de pilotage pour la réalisation des objectifs du Millénaire pour le développement en Afrique<sup>2</sup> que le manque d'infrastructure entrave considérablement la croissance économique, le commerce et la lutte contre la

<sup>1</sup> <http://content.undp.org/go/newsroom/2006/march/sen-20060308.fr.jsessionid=adKjxoEzP09e?lang=fr>

<sup>2</sup> [www.mdgafrica.org](http://www.mdgafrica.org)

pauvreté en Afrique et freine ainsi la réalisation des OMD.

Les indicateurs des OMD, bien qu'ils puissent être calculés au niveau régional, restent très globaux, cachant ainsi certaines inégalités. Certes, ils donnent, entre autres, une idée sur la scolarisation, la santé, l'accès à l'eau potable, mais ils ne permettent pas de mesurer les difficultés, qu'ont les populations à satisfaire leurs besoins vitaux (difficultés pour pouvoir accéder à l'eau potable, à l'établissement scolaire, à l'hôpital, au marché de produits alimentaires, etc.). Or, les difficultés quotidiennes font partie intégrante des conditions de vie. Il ne suffit pas d'avoir la possibilité de s'approvisionner en eau potable ou de se rendre à l'hôpital, phénomènes mesurés par les indicateurs des OMD. La distance parcourue pour atteindre le point d'eau ou l'hôpital est l'une des caractéristiques des inégalités à laquelle il faut s'intéresser si l'on souhaite donner à chaque personne la liberté de vivre de la façon dont elle le souhaite.

Ce travail étudie l'évolution de l'accessibilité aux infrastructures de base au Cameroun entre 2001 et 2007 à l'aide des données de la Deuxième enquête camerounaise auprès des ménages (ECAM2) et celles de la Troisième enquête camerounaise auprès des ménages (ECAM3). L'accessibilité aux infrastructures de base pour un ménage donné est quantifiée par les distances séparant son logement des plus proches. A partir de ces distances collectées en 2001, une typologie des ménages sera construite. Des techniques statistiques adéquates permettront de vérifier si la composition de cette typologie a changé en 2007.

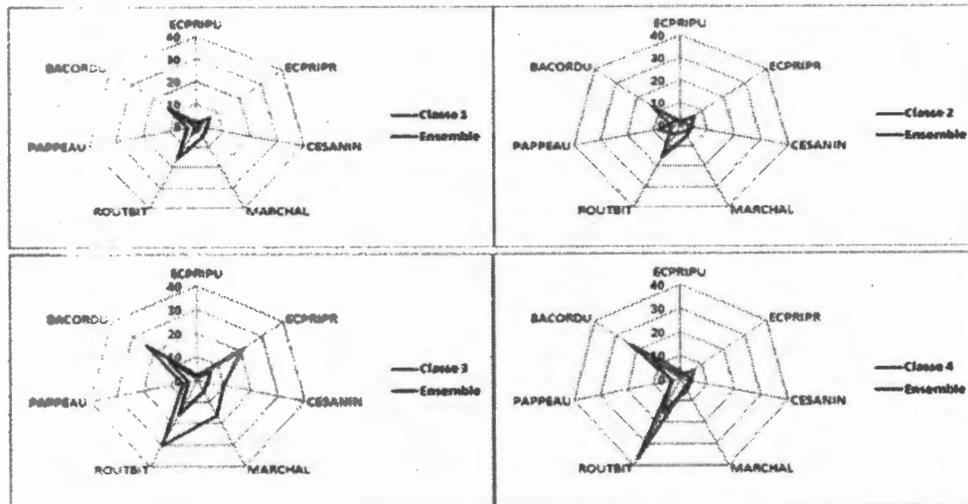
Cette approche permet ainsi de mesurer l'ampleur des changements observés dans l'environnement des ménages, de voir si les efforts gouvernementaux contribuent considérablement à l'amélioration des conditions de vie des ménages en vue de l'atteinte des objectifs du millénaire pour le développement. Ce travail vise aussi à mettre en exergue les disparités régionales, ce qui pourra permettre de faire des recommandations efficaces pour une amélioration rapide des conditions de vie des populations.

## 2. Construction de la typologie

Les infrastructures pour lesquelles les données ont été collectées successivement lors de l'ECAM2 et de l'ECAM3 sont : les écoles primaires publique (ECPRIPIU) et privée (ECPRIPIR), le centre de santé intégré (CESANIN), le marché des produits alimentaires (MARCHAL), la route bitumée (ROUTBIT), le point d'approvisionnement en eau potable à accès public (PAPPEAU) et le point de ramassage d'ordures (BACORDU).

Sur la base des distances séparant les ménages de ces infrastructures de bases collectées en 2001 (ECAM2), une typologie de ménages est construite. La technique adoptée est la classification sur les facteurs d'une Analyse en Composantes Principales, vu le nombre important de ménages (Baccini et Besse, 2005). Une classification par la méthode de réallocation dynamique est d'abord réalisée sur les ménages repérés par leurs composantes factorielles. Cette étape a permis de construire une typologie très composite (1000 classes). La classification ascendante hiérarchique est ensuite appliquée pour réduire ce nombre de classes. Au final, elle permet de constituer, de façon optimale, une typologie à 4 classes (figure 7 en annexe). La combinaison de ces deux méthodes de classification prend la dénomination de classification mixte ou conjointe (Nakache et Confais, 2005).

La figure ci-après donne les caractéristiques de chacune des 4 classes de la typologie retenue en termes de distance moyenne (en kilomètre) séparant les ménages d'une classe donnée à chacune des infrastructures de base.



**Figure 1 :** Description de la typologie (Distances moyennes en kilomètre entre les ménages de chaque classe et les infrastructures de base)

- Classe 1 : Elle regroupe les ménages dont la distance qui les sépare de chaque infrastructure est inférieure à la moyenne nationale. Il s'agit des ménages qui ont le moins de difficultés à accéder aux infrastructures de base ;
- Classe 2 : Ce sont les ménages dont le point d'approvisionnement en eau potable le plus proche et le centre de santé intégré le plus proche sont à des distances supérieures aux moyennes nationales. Quant aux restes infrastructures, elles sont à des distances inférieures à aux moyennes nationales ;
- Classe 3 : Elle regroupe les ménages qui sont à des distances supérieures à la moyenne nationale de toutes les infrastructures de bases. En dehors du point d'approvisionnement en eau potable et de l'école primaire publique, les autres infrastructures se situent à des distances de plus du double des moyennes nationales. C'est la classe des ménages les plus privés des infrastructures de base ;
- Classe 4 : Il s'agit des ménages les plus pénalisés en termes d'accès à la route bitumée et à l'école primaire publique. En plus de cela ces ménages sont à une distance du point de ramassages des ordures de près du double la moyenne nationale.

En 2001, moins de la moitié des ménages du Cameroun éprouvent moins de difficultés à accéder aux infrastructures de base (43%), 18% sont seulement très éloigné du point d'approvisionnement en eau potable et du centre de santé intégré (Tableau 2). Près de 20% des ménages très éloignés de toutes les infrastructures de base sont à plus 3 km de l'école primaire publique et à plus de 20 km du de la route bitumée ou du point de ramassage d'ordures. De tels résultats témoignent la précarité de l'environnement des ménages en 2001. Face aux politiques gouvernementaux de lutte contre la pauvreté, il en nécessaire, 6 ans après, de vérifier si conditions de vie des ménages ont changées.

### 3. Comment classer les ménages enquêtés en 2007 d'après la typologie créée ?

Après la construction de la typologie des ménages en 2001 suivant les difficultés qu'ils peuvent avoir à accéder aux infrastructures de base, il est nécessaire de classer les ménages-échantillons de 2007 d'après cette typologie.

Les techniques utilisées ci-dessus pour créer la typologie sont descriptives. On a plutôt besoin des techniques prédictives pour pouvoir déterminer la classe de chaque nouvel ménage. Dans cette situation, la typologie devient le facteur à expliquer et les distances utilisées pour son élaboration les prédicteurs. Ainsi, la variable prédite est qualitative à 4 modalités et les variables à expliquer quantitatives. Pour une telle modélisation, plusieurs techniques sont envisageables : la régression logistique multinomiale<sup>3</sup>, l'analyse discriminante décisionnelle<sup>4</sup>, les arbres de décision<sup>5</sup>, les réseaux de neurones<sup>6</sup> et les Séparateurs à Vaste Marge (SVM)<sup>7</sup> (Tufféry, 2007 ; Besse 2008). Il faut choisir parmi ces méthodes la plus robuste en termes de qualité de prédiction. Pour cela nous allons emprunter la technique de sélection de modèles du data mining qui consiste à répartir l'échantillon en deux : un échantillon d'apprentissage et un échantillon-test (Tufféry, 2007 ; Besse, 2008). Le modèle est estimé et paramétré sur l'échantillon d'apprentissage, le plus souvent par la méthode de validation croisée<sup>8</sup>. Ensuite, l'erreur de prédiction utilisée pour la comparaison des modèles est plutôt estimée sur l'échantillon-test qui ne participe pas à la phase d'apprentissage. Le modèle retenu par cette méthode présente une bonne qualité prédictive.

L'approche data mining adoptée a été implémenté avec le logiciel R. Le tableau 1 contient les erreurs de classement estimées sur l'échantillon-test. Il faut relever que l'échantillon a été réparti en 80% pour l'apprentissage et 20% pour le test.

*Tableau 1: Erreur de prédiction des différents modèles*

Modèle	Erreur de prédiction sur l'échantillon test
Régression logistique multinomiale	0,00090
Analyse discriminante décisionnelle	0,06667
Arbre de décision	0,18153
Réseau de neurones	0,00270
Séparateurs à Vaste Marge	0,02793

La plus petite erreur de classement sur l'échantillon-test est obtenue avec la régression logistique multinomial. En effet, en utilisant ce modèle pour prédire les classes des nouveaux ménages, on peut mal classer seulement 9 ménages sur 10000. Il est donc retenu pour prédire les classes des ménages échantillonnés en 2007.

Le tableau 2 donne la nouvelle répartition des ménages dans les 4 classes de la typologie. En

<sup>3</sup> [http://en.wikipedia.org/wiki/Multinomial\\_logit](http://en.wikipedia.org/wiki/Multinomial_logit)

<sup>4</sup> [http://fr.wikipedia.org/wiki/Analyse\\_discriminante\\_lin%C3%A9aire](http://fr.wikipedia.org/wiki/Analyse_discriminante_lin%C3%A9aire).

<sup>5</sup> [http://fr.wikipedia.org/wiki/Arbre\\_de\\_d%C3%A9cision](http://fr.wikipedia.org/wiki/Arbre_de_d%C3%A9cision)

<sup>6</sup> [http://fr.wikipedia.org/wiki/Neuronal\\_network](http://fr.wikipedia.org/wiki/Neuronal_network)

<sup>7</sup> [http://en.wikipedia.org/wiki/Support\\_vector\\_machine](http://en.wikipedia.org/wiki/Support_vector_machine)

<sup>8</sup> [http://en.wikipedia.org/wiki/Cross-validation\\_\(statistics\)](http://en.wikipedia.org/wiki/Cross-validation_(statistics))

effet, la *classe 1* regroupe 44,96% des ménages et la *classe 2*, 24,32%. Quant aux deux dernières classes (*classe 3* et *classe 4*) elles comportent respectivement 12,94% et 12,78%. On constate que la répartition des ménages n'est pas restée constante entre 2001 et 2007. L'objectif de ce travail étant de vérifier si les actions en matière de fourniture d'infrastructures de base entre 2001 et 2007 sont conséquentes face aux besoins existants, il est nécessaire de vérifier si les variations de proportions constatées sont significatives.

La principale approche ayant permis de comparer la composition de la typologie en 2007 à celle de 2001 c'est la méthode de simulation Bootstrap (Fox, 2002). Les différentes proportions ont été calculées sur 10000 échantillons issus d'un tirage avec remise dans l'échantillon initial (Saporta, 2006).

**Tableau 2:** Répartition des ménages du Cameroun en 2007 suivant les classes de la typologie

	Année 2001	Année 2007					
	Proportion (%)	Proportion (%)	Ecart-type	Biais	Erreur quadratique moyen	Borne inférieure (%)	Borne supérieure (%)
Classe1	43,00	44,96	0,5771	-0,0001	0,3331	43,83	46,10
Classe2	18,08	24,32	0,5137	0,0048	0,2639	23,31	25,35
Classe3	19,78	17,94	0,4894	0,0027	0,2395	16,99	18,91
Classe4	19,14	12,78	0,4263	-0,0075	0,1818	11,96	13,61

Pour chacune des 4 classes de la typologie l'intervalle de confiance à 95% de la proportion de ménages en 2007 estimé ne contient pas la proportion de 2001, ainsi les variations constatées sont significatives au seuil de 5%. De surcroît les résultats des tests de comparaison de proportions (Annexe ) viennent confirmer le changement significatif de la composition de la typologie entre 2001 et 2007.

En comparant la composition de la typologie en 2001 à celle de 2007, il ressort qu'en général les efforts menés en termes de fourniture d'infrastructures de base ont eu un impact perceptible. En effet, la proportion des ménages les plus proches des infrastructures de base (*Classe 1*) s'est accrue de près de 2 points entre ces deux années tandis que celle des ménages les plus éloignées de ces infrastructures (*classe 3*) a baissé de près de 2 points. La proportion des ménages les plus pénalisés en termes d'accès à la route bitumée, l'école primaire publique et au point de ramassage d'ordures (*classes 4*) a baissé de 6 points tandis que celle des ménages de la classe 2 s'est accrue 6 points.

#### 4. Disparités régionales

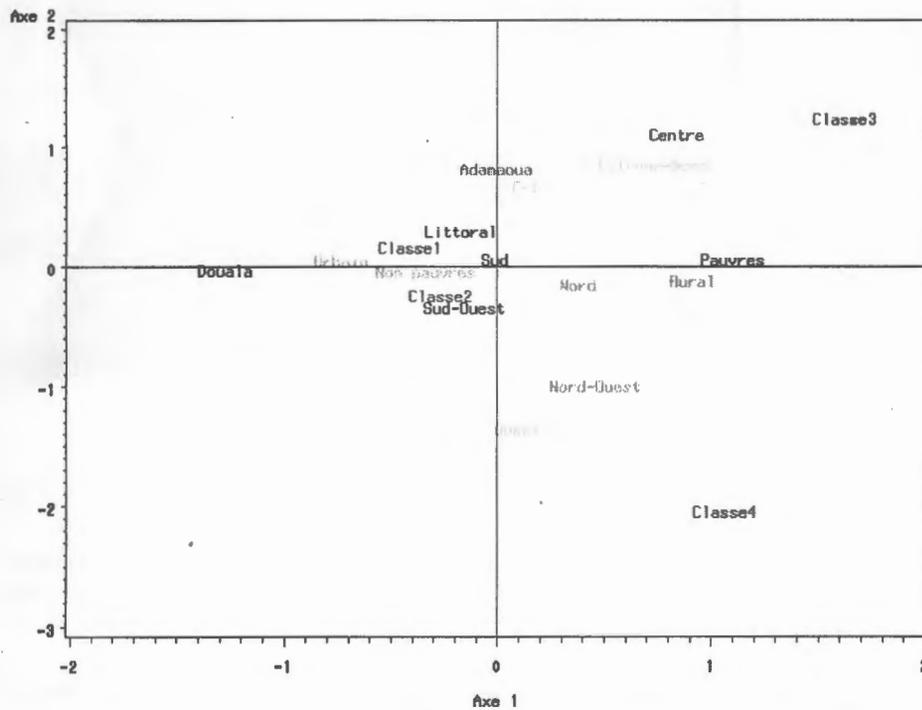
Les disparités régionales sont examinées à l'aide de l'Analyse Factorielles des Correspondances Multiples (AFCM) portant sur le niveau des vie (pauvre ou non pauvre), la région d'enquête (12 au total), le milieu de résidence (urbain et rural) et la typologie en 2007 (*classe 1* à *classe 4*).

Le Cameroun compte 10 régions, mais dans le cadre des ECAM, les villes de Douala et

Yaoundé sont considérées comme régions d'enquête à part entière. Elles sont donc exclues respectivement des régions du Centre et du Littoral (Figure 3).

Les résultats de l'AFCM (Figure 2) montrent que les deux métropoles Douala et Yaoundé sont moins pauvres et les plus pourvues d'infrastructures de base que toutes les autres régions du Cameroun. Pauvreté et manque d'infrastructure sont les caractéristiques propres aux régions du Centre (moins Yaoundé), du Nord et de l'Extrême-Nord qui, en grande partie, sont rurales.

Entre 2001 et 2007 les conditions de vie se sont considérablement dégradées dans la partie septentrionale du pays (formée des régions de l'Extrême-Nord, du Nord et de l'Adamaoua), à l'Est et dans le Nord-ouest (Figure 3). Il s'agit des régions où le manque d'infrastructure est le plus criard.



**Figure 2:** Premier plan factoriel de l'AFCM sur les données de l'ECAM3 (2007)

Dans les deux métropoles le taux de pauvreté a baissé de près de la moitié, il s'est accru de 14 points dans le Nord et de 10 points dans l'Extrême-Nord qui est restée toujours la région la plus pauvre du Cameroun.

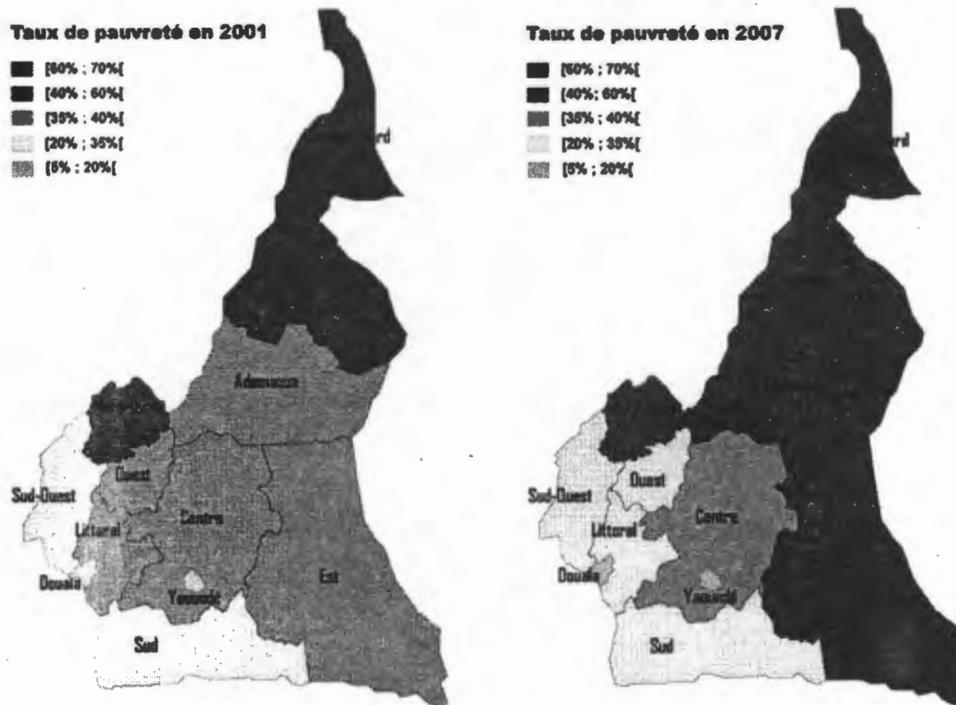


Figure 3: Taux de pauvreté en 2001 et 2007 par région d'enquête

## 5. Discussion

Au Cameroun le principal moyen de locomotion pour se rendre à toute infrastructure la plus proche c'est la marche à pied (d'après les données des deux ECAM). Ainsi la distance traduit très bien les difficultés qu'on les ménages à accéder aux infrastructures de base et représente ainsi un bon indicateur de mesure des inégalités ou en d'autres termes des privations.

Malgré la légère amélioration de l'environnement des ménages le taux de pauvreté est presque resté statique entre 2001 et 2007, passant respectivement de 40,2% à 39,9% (INS, 2008). Cela témoigne d'une intensification du nombre de pauvres au Cameroun étant donné que la taille de la population va croissante.

Les écarts entre les deux métropoles et les autres régions sont très importants. Le milieu rural qui est caractérisé par l'extrême pauvreté s'est davantage appauvri entre ces deux années. L'incidence de la pauvreté y est passée de 52,1% à 55,0%. Cette situation est attribuée en grande partie à l'absence d'infrastructures qui ne favorise la pratique de grandes activités, par exemple l'agriculture et l'élevage de grande taille.

Entre 2001 et 2007, la pauvreté s'est accrue dans les régions les plus pauvres (Extrême-Nord, Nord, Adamaoua et Est) où l'on observe le criard problème d'accessibilité aux infrastructures

de base. En effet, toutes ces régions sont géographiquement très éloignées de Douala et Yaoundé, et les moyens de transport qui les lient à ces deux métropoles sont encore très coûteux en temps de déplacement à cause des infrastructures routières et ferroviaires non modernisées.

Au vu des différentes statistiques, le pessimisme s'installe sur la possibilité d'atteinte des OMD par le Cameroun. Si la politique de développement des infrastructures de base est maintenue, la proportion des ménages accédant facilement aux infrastructures de base serait toujours inférieure à 50% d'ici 2015, ainsi plus de la moitié de la population sera toujours privée de routes bitumées et confrontée au problème de gestion des ordures, près de 48% seront toujours très éloignées des points d'approvisionnement en eau potable. Le taux de pauvreté se situera ainsi autour de 39%, très au-delà du taux de 25,2% escompté en 2015 (INS, 2008) d'après les OMD.

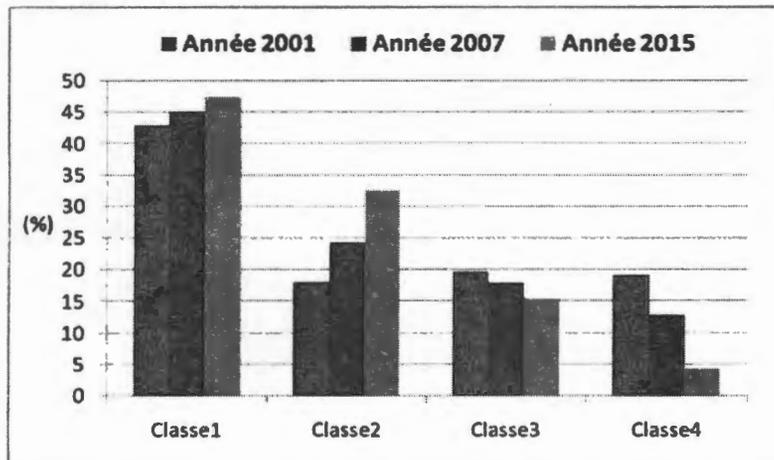


Figure 4: Structure de la typologie de l'accessibilité aux infrastructures de base en 2001, 2007 et estimation pour 2015

## 6. Conclusion et recommandations

Les données des deux Enquête Camerounaises auprès des Ménages montrent que l'accessibilité géographique aux infrastructures de bases traduit bel et bien les conditions de vie des ménages.

Le développement du Cameroun évolue à deux vitesses, d'un côté on a les deux métropoles où les infrastructures de base sont à quelques pas des ménages et de l'autre côté le reste du pays où une part importante de ménages est privée de ces infrastructures. Dans

ce deuxième groupe la partie septentrionale pays est la plus défavorisées et les conditions de vie ne cessent de s'y détériorer.

Parlant de la liberté de se mouvoir, bon nombre d'habitants sont encore très éloignés des routes bitumées et cette situation risque de perdurer dans la mesure où les politiques actuelles de modernisation des infrastructures routières du Cameroun continuent de privilégier les capitales

politique et économique (Yaoundé et Douala respectivement). Il est grand temps de penser à la modernisation des infrastructures routières du milieu rural, ce qui pourra faciliter l'agriculture intensive, et contribuer à la lutte contre l'exode rural.

Pour une amélioration considérable des conditions de vie des ménages, il serait nécessaire d'accroître la fourniture des infrastructures de base dans la partie septentrionale du pays. Par exemple :

- en reliant chaque que arrondissement à son chef-lieu de département par une route bitumée. Cette modernisation favorisera les déplacements des populations vers les marchés, des enseignants et des élèves vers leurs établissements.
- en s'assurant d'une répartition spatiale équitable des établissements scolaires et des centres de santé ;
- dans cette partie très aride du pays, il serait nécessaire d'intensifier l'implantation des forages et songer déjà à des systèmes modernes d'irrigation pour favoriser l'agriculture en toute saison

La collecte et le traitement des ordures doit être étendue à travers le territoire national. Toutes ces actions doivent être accompagnées par d'une gestion efficace des infrastructures existantes.

On retient que la planification des investissements au Cameroun est à revoir compte tenu des disparités dégagées, Il serait nécessaire d'élaborer un plan d'investissement pour chaque région en prenant en compte ses spécificités.

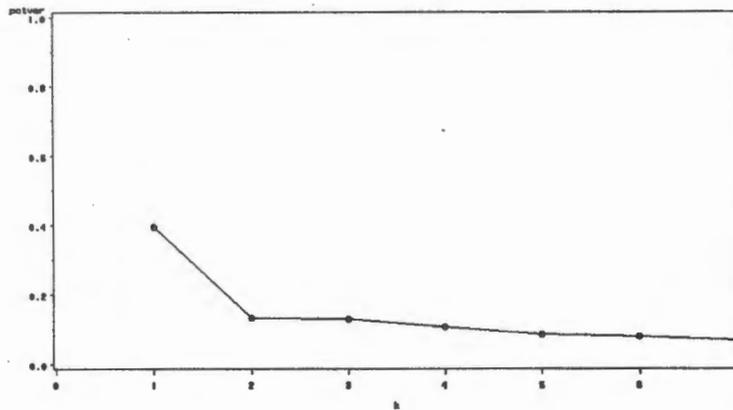
**BIBLIOGRAPHIE**

- Baccini A. & Besse P. (2005). *Data mining I. Exploration Statistique*. Publication du Laboratoire de Statistique et Probabilités de l'Université PAUL SABATIER, Toulouse III, France, 142 p.
- Baccini A. & Besse P. (1999). *Statistique descriptive multidimensionnelle*. Publication du Laboratoire de Statistique et Probabilités de l'Université Paul Sabatier, Toulouse 3, France, 94 p.
- Besse P. (2008). *Apprentissage Statistique & Data mining*. Institut de Mathématiques de Toulouse, Laboratoire de Statistique et Probabilités, Institut National des Sciences Appliquées de Toulouse, France, 124 p.
- Fox J. (2002). *Bootstrapping regression models. Appendix to an R and S-PLUS companion to applied regression*.
- INS (2008). *Tendance, profil et déterminants de la pauvreté au Cameroun entre 2001-2007*. Troisième Enquête Camerounaise auprès des Ménages.
- Nakache J.P. et Confais J. (2005). *Approche pragmatique de la classification*. Paris, France, Editions TECHNIP, 262 p.
- Palm R. (2002). *Utilisation du bootstrap pour les problèmes statistiques liés à l'estimation des paramètres*. BASE 6, p 143-153.
- Paradis E. (2005) *R pour les débutants*. Institut des Sciences de l'Évolution, Université Montpellier II France, 77 p.
- Rapacchi B. (2004). *Une introduction au bootstrap*. Grenoble, France, Centre Interuniversitaire de Calcul, 69 p.
- R Development Core Team (2008). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
- Saporta G. (2006). *Probabilités Analyse de données et Statistique. 2ième édition révisée et augmentée*. Paris, France, Editions TECHNIP, 622 p.
- Tufféry S. (2007). *Data mining et statistique décisionnelle. L'intelligence des données. Nouvelle édition revue et augmentée*. Paris, France, Editions TECHNIP, 533 p.
- Volle M. (1981). *Analyse de données. 2ième édition*. Paris, France, Economica, 319 p.

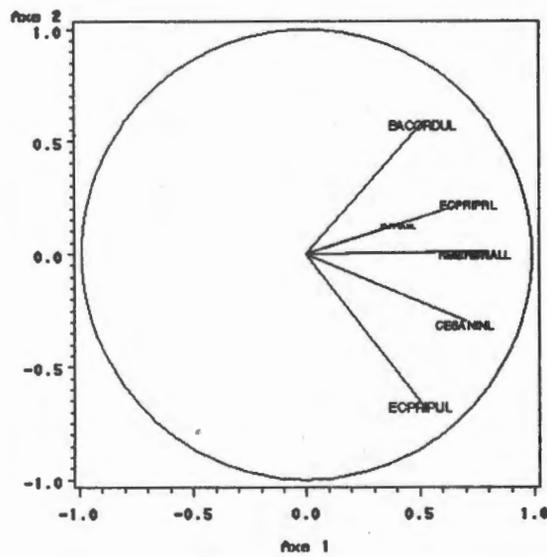
**Sites internet**

- <http://content.undp.org/go/newsroom/2006/march/sen-20060308.fr.jsessionid=adKjxoEzP09e?lang=fr>
- [http://www.waternunc.com/fr/banqmd08\\_2002.htm](http://www.waternunc.com/fr/banqmd08_2002.htm)
- [http://www.ugcdf.org/francais/local\\_development/](http://www.ugcdf.org/francais/local_development/)
- [www.mdgafrica.org](http://www.mdgafrica.org)
- [http://en.wikipedia.org/wiki/Multinomial\\_logit](http://en.wikipedia.org/wiki/Multinomial_logit)
- [http://fr.wikipedia.org/wiki/Analyse\\_discriminante\\_lin%C3%A9aire](http://fr.wikipedia.org/wiki/Analyse_discriminante_lin%C3%A9aire)
- [http://fr.wikipedia.org/wiki/Arbre\\_de\\_d%C3%A9cision](http://fr.wikipedia.org/wiki/Arbre_de_d%C3%A9cision)
- [http://fr.wikipedia.org/wiki/Neuronal\\_network](http://fr.wikipedia.org/wiki/Neuronal_network)
- [http://en.wikipedia.org/wiki/Support\\_vector\\_machine](http://en.wikipedia.org/wiki/Support_vector_machine)
- [http://en.wikipedia.org/wiki/Cross-validation\\_\(statistics\)](http://en.wikipedia.org/wiki/Cross-validation_(statistics))

**Annexe 1 : Résultats l'Analyse en Composantes Principales et de la classification**



**Figure 5: Eboulis des valeurs propres**



**Figure 6: Plan factoriel des variables**

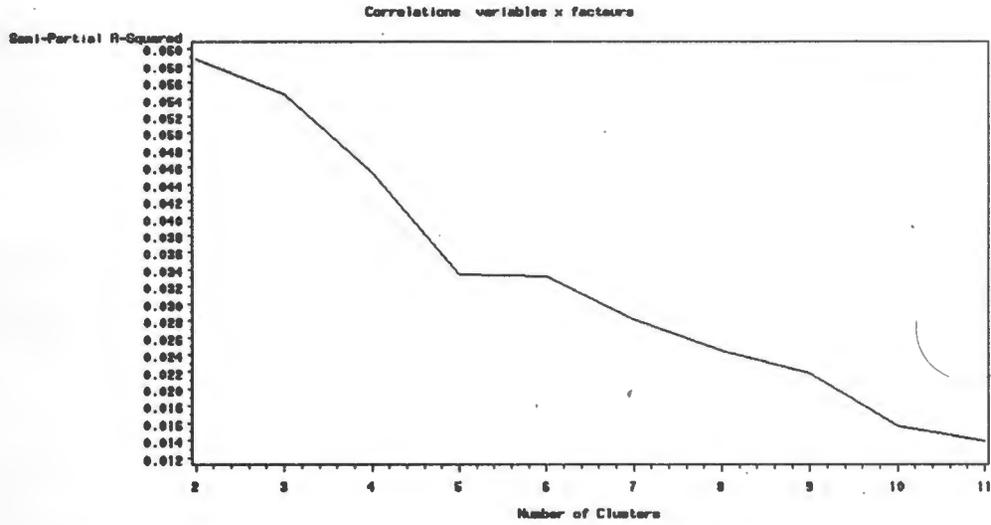


Figure 7: R-carré semi-partiel en fonction du nombre de classes

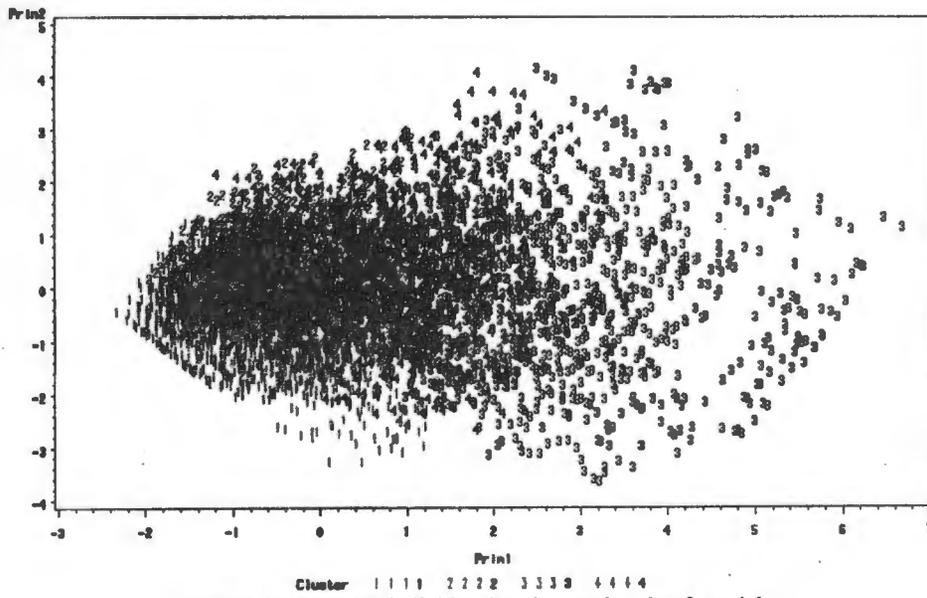


Figure 8: Nuage des individus dans le premier plan factoriel

**Annexe 2 : Résultats de modélisation**

Régression logistique multinomial

Call:

multinom(formula = CLASSE ~ ., data = Ec2appt)

Coefficients:

(Intercept) ECPRI PUL ECPRI PRL CESANINL MARCHALL ROUTBITL PAPPEAUL  
BACORDUL

Classe2 -2309.341 619.2459 -267.7745 1145.7523 177.9072 61.65639 1488.5149 -281.3852

Classe3 -14246.574 2231.2871 1538.6846 2628.6428 2219.6406 1178.47816 789.0530 109.8863

Classe4 -6425.373 2727.7475 -161.1489 535.9288 538.6893 1155.94906 992.9126 316.8467

Std. Errors:

(Intercept) ECPRI PUL ECPRI PRL CESANINL MARCHALL ROUTBITL PAPPEAUL  
BACORDUL

Classe2 139.1716 39.36994 20.09977 69.80953 27.20276 46.52774 97.97658 27.03834

Classe3 103.1876 586.07919 954.10257 248.04150 476.59832 271.37306 262.89892 151.06309

Classe4 281.1640 170.33194 175.80444 132.76315 204.40877 108.79955 193.78680 56.83154

Residual Deviance: 6.157786

AIC: 54.15779

Matrice de confusion :

	Classe1	Classe2	Classe3	Classe4
Classe1	1177	0	0	0
Classe2	0	406	0	0
Classe3	0	0	313	0
Classe4	2	0	0	322

Analyse discriminant décisionnelle

Matrice de confusion

fit.knn	Classe1	Classe2	Classe3	Classe4
Classe1	1157	38	8	21
Classe2	11	352	4	11
Classe3	2	5	286	15
Classe4	7	11	15	277

Arbre de décision

Classification tree:

tree(formula = CLASSE ~ ., data = Ec2appt, na.action = na.omit)

Variables actually used in tree construction:

[1] "ROUTBITL" "PAPPEAUL" "CESANINL" "ECPRI PUL" "MARCHALL" "ECPRI PRL"

Number of terminal nodes: 13

Residual mean deviance: 0.909 = 7962 / 8759

Misclassification error rate: 0.1811 = 1589 / 8772

Matrice de confusion

pred	Classe1	Classe2	Classe3	Classe4
Classe1	1124	117	8	64
Classe2	29	237	10	25
Classe3	13	16	248	27
Classe4	11	36	47	208

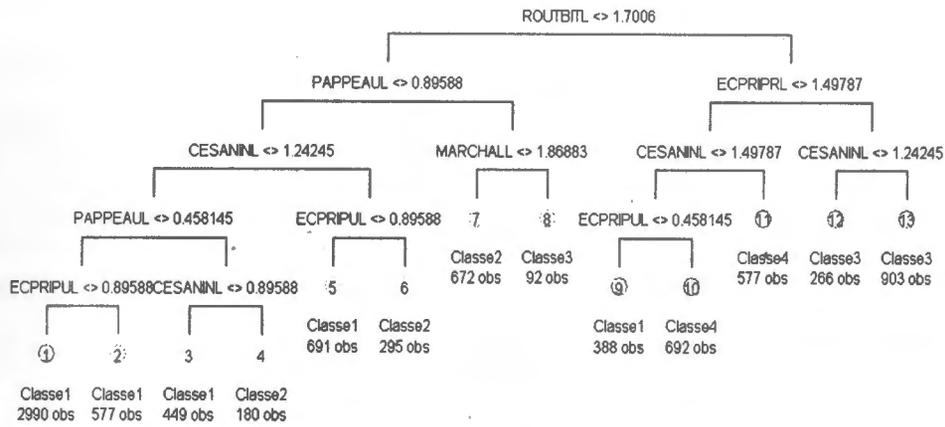


Figure 9: Arbre de classification

Réseau de neurones

a 7-5-4 network with 64 weights  
options were - softmax modelling

b->h1 i1->h1 i2->h1 i3->h1 i4->h1 i5->h1 i6->h1 i7->h1  
-3.68 2.50 0.07 -0.77 0.37 1.30 -0.61 0.72  
b->h2 i1->h2 i2->h2 i3->h2 i4->h2 i5->h2 i6->h2 i7->h2  
8.88 -0.60 -1.77 -2.28 -1.78 0.01 0.20 0.25  
b->h3 i1->h3 i2->h3 i3->h3 i4->h3 i5->h3 i6->h3 i7->h3  
130.23 -55.80 -30.92 69.09 -49.58 -8.94 100.98 -32.43  
b->h4 i1->h4 i2->h4 i3->h4 i4->h4 i5->h4 i6->h4 i7->h4  
-12.13 5.26 -0.33 1.11 1.05 2.20 2.12 0.60  
b->h5 i1->h5 i2->h5 i3->h5 i4->h5 i5->h5 i6->h5 i7->h5  
-1120.04 302.15 -126.67 550.54 90.61 34.80 707.46 -132.04  
b->o1 h1->o1 h2->o1 h3->o1 h4->o1 h5->o1  
233.78 8.37 348.29 -11.49 -696.03 -264.63  
b->o2 h1->o2 h2->o2 h3->o2 h4->o2 h5->o2  
-25.08 -781.36 388.09 8.99 218.61 258.97  
b->o3 h1->o3 h2->o3 h3->o3 h4->o3 h5->o3  
260.63 477.35 -979.81 -4.28 214.12 -1.15  
b->o4 h1->o4 h2->o4 h3->o4 h4->o4 h5->o4  
-468.09 296.31 244.25 6.27 263.91 6.64

## Matrice de confusion

	Classe1	Classe2	Classe3	Classe4
Classe1	1177	0	0	0
Classe2	1	405	0	0
Classe3	0	0	311	2
Classe4	1	0	2	321

## Séparateurs à Vastes Marges

```
svm(formula = CLASSE ~ ., data = Ec2appt)
```

## Parameters:

SVM-Type: C-classification

SVM-Kernel: radial

cost: 1

gamma: 0.1428571

Number of Support Vectors: 1608

( 468 459 399 282 )

Number of Classes: 4

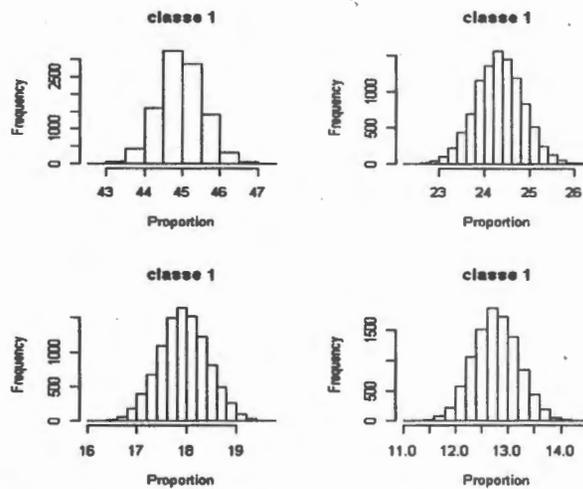
## Levels:

Classe1 Classe2 Classe3 Classe4

## Matrice de confusion

```
pred.svm
```

	Classe1	Classe2	Classe3	Classe4
Classe1	1163	7	1	6
Classe2	12	389	1	4
Classe3	5	2	301	5
Classe4	7	2	10	305

**Annexe 3 : Résultats de la simulation (Bootstrap)****Figure 10:** Histogramme des proportions sur l'échantillon Bootstrap**Annexe 4 : Résultats test de comparaison de proportions****Classe 1 :**

Proportion binomiale pour CLASSE = Classe1	
Proportion	0.4496
ASE	0.0002
95% Limite de conf inf.	0.4491
95% Limite de conf sup.	0.4501
Test de H0 : Proportion = 0.43	
ASE sous H0	0.0002
Z	79.9321
Pr > Z unilatérale	<.0001
Pr >  Z  bilatéral	<.0001

**Classe 2 :**

Proportion binomiale pour CLASSE = Classe2	
Proportion	0.2432
ASE	0.0002
95% Limite de conf inf.	0.2428
95% Limite de conf sup.	0.2436

Test de H0 : Proportion = 0.1808  
 ASE sous H0 0.0002  
 Z 327.2360  
 Pr > Z unilatérale <.0001  
 Pr > |Z| bilatéral <.0001

Classe 3 :

Proportion binomiale  
 pour CLASSE = Classe3

-----  
 Proportion 0.1794  
 ASE 0.0002  
 95% Limite de conf inf. 0.1790  
 95% Limite de conf sup. 0.1798

Test de H0 : Proportion = 0.1978  
 ASE sous H0 0.0002  
 Z -93.2956  
 Pr < Z unilatérale <.0001  
 Pr > |Z| bilatéral <.0001

Classe 4 :

Proportion binomiale  
 pour CLASSE = Classe4

-----  
 Proportion 0.1278  
 ASE 0.0002  
 95% Limite de conf inf. 0.1275  
 95% Limite de conf sup. 0.1281

Test de H0 : Proportion = 0.1914  
 ASE sous H0 0.0002  
 Z -326.2470  
 Pr < Z unilatérale <.0001  
 Pr > |Z| bilatéral <.0001