

BURKINA FASO

Unité – Progrès – Justice

MINISTRE DE L'ENSEIGNEMENT
SUPERIEUR, DE LA RECHERCHE
SCIENTIFIQUE ET DE L'INNOVATION

MINISTRE DE LA SANTE

Université Polytechnique de
Bobo Dioulasso

Secrétariat Général

Unité de Formation et Recherche en
Sciences et Techniques

Centre MURAZ

Licence de Statistiques et Informatique
(LSI)



*Rapport de stage de licence en Statistique et Informatique
effectué au Centre MURAZ du 02 mai au 02 septembre 2016*

Sous le thème :

***Développement et validation d'un plan de sondage à base aréolaire
dans la commune urbaine de Bobo Dioulasso :***

***Définition d'une base de sondage par
modélisation spatiale***

Présenté par :

SOUBEIGA Armel

Maitre de stage :

Dr. Serge Aymar SOMDA
Méthodologiste-Biostatisticien
Centre MURAZ

Encadré par :

Pr. Jean de Dieu ZABSONRE
Maître de conférences en Mathématiques
Université Polytechnique de Bobo-Dioulasso

4. Le krigeage	28
4.1 La variance de krigeage	28
4.2 Validation croisée	31
4.3 La grille d'estimation.....	32
IV. Présentation des résultats	32
1. Résultat de krigeage	32
2. Résultats de la validation croisée	35
3. Cartographie des lieudits avec les itinéraires associés	37
V. Discussions et conclusion	38
1. Discussions.....	38
1.1 Avantages et limites.....	38
1.2 Perspectives	38
2. Conclusion.....	39
Références bibliographiques	40
Annexes : Code R	a
Annexes A : Krigeage	a
1. Analyse variographique avec R.....	a
2. Le Krigeage avec R	b
3. La validation croisée avec R.....	c
Annexe B : Cartographie avec R.....	d

***Définition
d'une base de
sondage par
modélisation
spatiale***

Mai 2016

Année universitaire : 2015—2016

Dédicaces

Je dédie ce rapport à :

Mes chers parents, que nulle dédicace ne puisse exprimer mes sincères sentiments, pour leur patience illimitée, leur encouragement continu, leur aide, en témoignage de mon profond amour et respect pour leurs grands sacrifices.

Mon père,

« L'épaule solide, l'œil attentif, compréhensif et la personne la plus digne de mon estime et de mon respect »

Ma mère,

« Tu m'as donné la vie, la tendresse et le courage pour réussir ; tout ce que je pourrais t'offrir ne pourra jamais exprimer l'amour et la reconnaissance que je te porte »

Mes chers frères : Adolph, Aristide, Anicet et Alex,

Pour leur grand amour et leur soutien qu'ils trouvent ici l'expression de ma haute gratitude

Mes chers amis qui sans leur encouragement ce travail n'aura vu le jour, et à toute ma famille et à tous ceux que j'aime ...!

Goubeïga Armel

Remerciements

Nulle œuvre n'est exaltante que celle réalisée avec le soutien moral, intellectuel et financier des personnes qui nous sont proches.

Je tiens à remercier particulièrement mon directeur de mémoire le Pr. Jean de dieu ZABSONRE, directeur de l'Institut Universitaire de Technologie et enseignant-chercheur à l'Université Polytechnique de Bobo-Dioulasso, pour sa disponibilité et son intérêt pour ce travail. Sa gentillesse, sa modestie, sa riche expérience et ses précieuses directives tout au long de la réalisation de ce travail m'ont inspiré une grande admiration à son égard.

Je remercie également mon directeur de stage Dr. Serge M. A. SOMDA, enseignant-chercheur à l'Université Polytechnique de Bobo-Dioulasso et responsable de l'Equipe d'Appui Méthodologique et de Formation (AMF) du Centre MURAZ de m'avoir accueilli dans son équipe et pour toutes les connaissances statistiques qu'il m'a apportées tout au long de ma période de stage. Sa rigueur scientifique, sa disponibilité et ses qualités humaines m'ont profondément touché.

Le directeur de l'Unité de Formation et de Recherche en Science et Technologie, Pr. Sado TRAORE, son directeur adjoint Dr. Lassina OUATTARA, à Dr. Adama de S. OUEDRAOGO, Coordonnateur de la filière Statistiques-Informatique, à tous les professeurs et missionnaires de la Licence en Statistiques et Informatique, ainsi que tout le personnel de l'Unité de Formation et de Recherche en Science et Technologie reçoivent ici mes sincères remerciements pour nous avoir offert un cadre d'apprentissage adéquat.

Un grand merci et une réelle reconnaissance à toute l'équipe de l'AMF et à l'ensemble du personnel du Centre MURAZ pour sa bonne humeur et sa convivialité. Je remercie particulièrement Mr. John Emmanuel SAWADOGO, géographe au Centre MURAZ et Mr. Alassane Dicko, Informaticien à la Direction Régionale du Ministère de l'Habitat et de l'Urbanisme pour les aides et les conseils.

Mes remerciements les plus chaleureux vont à tous mes camarades de la Licence en Statistiques et Informatique pour l'ambiance agréable tout au long de ce cycle de formation et en particulier à Ouattara Bakary de m'avoir apporté son aide.

Je tiens à remercier sincèrement les membres du jury qui me font le grand honneur d'évaluer ce travail.

Avant-propos

Ce manuel est le fruit d'un stage de fin de cycle effectué durant huit (08) mois dont quatre (04) mois en temps plein allant du 02 mai au 02 septembre 2016 au sein de l'Equipe d'Appui Méthodologique et de Formation (AMF) du Centre MURAZ. Au cours de ce stage, il était question de réfléchir sur des méthodes de conception d'un système de sondage dans la commune urbaine de Bobo-Dioulasso.

En effet, la formation en Licence de Statistiques et Informatique (LSI) est un cursus de BAC + 3 et est validée par une licence professionnelle à l'issue de la présentation d'un rapport de stage obligatoire soutenu devant un jury.

La LSI est une filière professionnalisante de l'Université Polytechnique de Bobo-Dioulasso (UPB) et rattachée à l'Unité de Formation et de Recherche en Sciences Techniques (UFR/ST). Les étudiants reçoivent une formation solide basée en statistique et en informatique qui les prépare aux différents métiers dans les domaines de la statistique et de l'aide à la décision. Cette formation leurs permet :

- ✓ d'organiser la collecte des données
- ✓ d'analyser, résumer, segmenter des vastes ensembles de données ;
- ✓ de décrire, traiter, synthétiser des résultats d'enquêtes ;
- ✓ d'analyser, décomposer, désaisonnaliser, modéliser des séries chronologiques ;
- ✓ d'estimer et tester les effets d'un ensemble de facteurs ;
- ✓ de concevoir et planifier un sondage.

Ces statisticiens doivent être capables d'aborder avec compétence les problèmes que l'on rencontre dans les applications (analyse du problème, acquisition des données, description des données, modélisation, mise en œuvre, interprétation et aide à la décision).

Résumé

Les méthodes statistiques permettent de recueillir différentes informations sur différents domaines au sein d'une population. Ces informations sont produites soit par recensement, soit par sondage. La méthode des sondages est une approche différentielle qui consiste à faire l'étude sur un échantillon et à extrapoler les résultats sur toute la population. De ce fait, plusieurs approches méthodologiques de sondage sont utilisées telles que les méthodes probabilistes et les méthodes empiriques.

Toutefois, de nos jours, de nombreuses études et projets nécessitent de bases de données instantanées et dans des lieux bien spécifiques. La nécessité est donc souvent de disposer de données actualisées, dans les contraintes de temps et dans des moindres coûts. Egalement de disposer de base de sondage valide, fiable, représentative et surtout minimisant les biais. Ce qui n'est pas le cas des méthodes de collecte de données que proposent les instituts nationaux de statistique (méthodes probabilistes) et des méthodes empiriques qu'utilisent des entreprises nationales, les petits instituts ou départements de statistique.

L'objectif de ce travail était de combiner ces deux approches (probabilistes et empiriques) afin de proposer une approche technique et scientifique dans la réalisation d'un système de sondage qui allie démarche scientifique valide, facilité de mise en œuvre et à moindre coût dans la ville de Bobo-Dioulasso. En outre, l'objectif est de contribuer à renforcer les capacités du Centre Muraz à travers son équipe AMF dans la collecte de l'information et sa politique de recherches médicales.

Pour ce faire, l'étape présente était de proposer une méthodologie de construction d'une base de sondage aréolaire de la ville de Bobo-Dioulasso. Nous avons utilisé des outils statistiques pour les analyses univariées, bivariées et spatiales des variables qui ont retenu notre attention. Il s'agit, entre autres, des histogrammes, des courbes de densité, l'indice de Moran d'autocorrélation spatiale, l'ellipse de déviation standard (nuée variographique), des tests de normalité de Kolmogorov-Smirnov. En ce qui concerne, la modélisation, nous avons fait recours aux notions de fonction aléatoire, processus stationnaire et intrinsèque, de variogramme et d'analyse variographique. Pour l'interpolation spatiale, le krigeage ordinaire par surface de réponse s'est avéré être le mieux adapté à notre situation.

Les résultats obtenus sont entre autres une carte d'interpolation de la densité de population sur toute l'étendue de la ville, une fonction d'estimation de la densité sur 3 mille km² à tout endroit de la ville. Egalement, une cartographie d'un certain nombre de lieux dits avec les itinéraires associés et une base de sondage aréolaire dont la validation a fait l'objet d'une autre étude.

Abstract

The objective of this work was to combine two approaches surveys, (probabilists and empirical) in order to propose a technical and scientific approach in the realization of a system of survey which combines valid scientific approach, simplicity of implementation and lower costs in the town of Bobo-Dioulasso. Moreover, the objective is to contribute to reinforce the capacities of the Centre Muraz through its AMF team in the collection of information and its policy of medical research.

To do this, the present step was to propose a methodology for the construction of an area data surveys in the town of Bobo-Dioulasso. We used statistical tools for the univariate, bivariate and spatial analysis of the variables that caught our attention. These include histograms, density curves, spatial autocorrelation Moran index, standard deviation ellipse (variographic cloud), Kolmogorov-Smirnov normality tests. In terms of modeling, we have used the notions of random function, stationary and intrinsic process, variogram and variographic analysis. For spatial interpolation, the ordinary kriging per response surface proved to be the best suited to our situation.

Liste des sigles

AMF	Appui Méthodologique et Formation
LSI	Licence en Statistique et Informatique
UPB	Université Polytechnique de Bobo-Dioulasso
INSD	Institut Nationale de Statistique et Démographique
UFR/ST	Unité de Formation et de Recherche en Sciences et Techniques
UP	Unité Primaire
US	Unité Secondaire
ZD	Zone de Dénombrement
BNDT	Base Nationale des Données Topographique
UES	Unité Environnement Santé
EPS	Etablissement Publique de Santé
CHU	Centre Hospitalier Universitaire
CMA	Centre Médical avec Antenne chirurgical
CSPS	Centre Social et de Promotion Sociale

Liste des tableaux

Tableau 1 : Répartition des ménages en fonction du type d'habitat	13
Tableau 2 : Répartition des ménages pour un habitat à plusieurs logements	13

Liste des graphiques et Cartes

Figure 1 : Présentation de la zone d'étude	4
Figure 2 : Organigramme Scientifique.....	7
Figure 3 : Répartition spatiale des lieux-dits observés.....	10
Figure 4 : Méthodes de dénombrement des parcelles d'habitation par lieudits.....	11
Figure 5 : Observation des Lieudits sur Google Earth.....	12
Figure 6 : Exemple de photos de quelques lieudits échantillonnées.....	15
Figure 7 : Spatialisation des lieudits en fonction de la densité	16
Figure 8 : Histogramme des densités	17
Figure 9 : Histogramme des densités après correction.....	17
Figure 10 : La droite de Henry	18
Figure 11 : Nuée variographique omnidirectionnelle	19
Figure 12 : Le Scattergramme.....	20
Figure 13 : Exemple du processus variographie	23
Figure 14 : Nombre de paires de lieudits pour chaque classe	24
Figure 15 : Exemple du corrélogramme d'estimation à X_0 avec 9 sites	26
Figure 16 : Le variogramme.....	27
Figure 18 : Processus de conception de la grille	32
Figure 17 : La grille d'interpolation	33
Figure 19 : Estimation de la densité de population sur l'étendue de la ville	33
Figure 20 : Cartographie des erreurs d'estimation et histogramme des erreurs standardisées	35
Figure 21 : Cartographie des lieudits avec les itinéraires associés	37

Définition et Concepts

Cette partie donne une définition des mots et termes consacrés que nous avons utilisés dans ce rapport de recherche. Le lecteur pourra donc aborder dans les chapitres suivants le vocabulaire avec plus de facilité.

Un arrondissement : est une subdivision administrative de la commune et qui regroupe des secteurs.

Une commune : est une collectivité territoriale de base. Son territoire est organisé en secteurs et/ou en villages. C'est une agglomération d'habitations définie par l'administration. Elle relève de l'autorité administrative de la Mairie.

Commune Urbaine : est toute entité territoriale comprenant au moins une agglomération permanente de 25000 habitants et dont les activités économiques permettent de générer des ressources budgétaires propres annuelles d'au moins 25 millions de francs CFA.

Un lieu-dit (lieudit) : est tout endroit remarquable, de faible étendue auquel est associé un nom propre. Le lieu-dit peut être un monument, une pharmacie, un centre de santé, une mairie.

Un secteur : c'est une subdivision d'une commune urbaine. Il regroupe les quartiers. Un quartier est un ensemble d'habitations, géographiquement délimité, défini comme tel par la population.

Le ménage ordinaire : c'est l'unité socio-économique de base au sein de laquelle un ou plusieurs membres, apparentés ou non, vivent dans la même maison ou concession, mettent en commun leurs ressources et satisfont en commun à l'essentiel de leurs besoins alimentaires et autres besoins vitaux, sous l'autorité de l'un d'entre eux appelé chef de ménage.

Le ménage collectif : c'est un groupe de personnes généralement sans lien de parenté entre elles, habitant ensemble dans des conditions spéciales, c'est-à-dire utilisant principalement les installations qu'un établissement met à leur disposition pour subvenir à leurs besoins essentiels (repas, logement, soins). Des exemples de ménages collectifs sont: casernes militaires, prisons, couvents, orphelinats, internats scolaires, etc. (INSD 2006)

Immeuble à appartements: C'est un bâtiment de type moderne d'un ou de plusieurs niveaux (étages) divisés en appartements. Chaque appartement peut abriter au moins un ménage.

Villa : c'est une maison individuelle comportant un séjour (salon), une ou plusieurs chambres ainsi que les commodités intérieures. Elles sont ordinairement destinées à l'hébergement d'un seul ménage.

Bâtiment à plusieurs logements (non immeuble): construction sans niveau formée de plusieurs logements séparés et pouvant abriter plusieurs ménages.

Zones d'habitation : ce sont les zones destinées au logement des ménages. Elles comprennent les zones loties et les zones non loties.

Zones administratives : ce sont des zones destinées à abriter des structures de l'administration, notamment les mairies, les directions régionales de ministères, les infrastructures privées, etc.

Zones commerciales : ce sont des zones destinées à l'exercice de l'activité de vente de biens et services. On y trouve des marchés, des supermarchés, des salons de coiffure, des restaurants, des séries de boutiques, etc.

I. Introduction

La méthode des sondages est utilisée au Burkina Faso aussi bien que dans les pays en voies de développement pour produire l'information sur différents domaines : enquêtes démographiques, enquêtes socio-économiques ou enquête épidémiologique, etc. L'INSD est la structure de l'Etat burkinabè qui s'occupe de la production des données. Ces données produites sont principalement démographiques à échelle temporelle de cinq à dix ans et à échelle régionale, provinciale, communale et par secteur. Cependant, la croissance rapide de la population, couplée de l'étalement urbain pose des problèmes sociodémographiques, de prise en charge sanitaire et économique importants. En effet, de plus en plus d'études nécessitent de données récentes et actualisées, au sein de périmètre bien spécifiques. L'absence de ces bases de sondage¹ constitue une barrière dans le processus de la quête d'information sur la population et donc dans la suivie de son évolutions. Cela rend également laborieux l'élaboration des plans de sondage fiables et rigoureux (probabilistes). Il devient impératif d'avoir de nouvelles méthodes de sondage ou de collecte d'informations. Toutefois, la construction d'un système de sondage à base aréolaire pose de nombreuses questions dont les suivantes : comment et à travers quelle fonction mathématique peut-on estimer ou modéliser les caractéristiques de la population d'une agglomération? Comment construire une base aréolaire des ménages qui allie démarche scientifique valide et facile de mise en œuvre à moindre coût dans la commune urbaine de Bobo-Dioulasso ?

Dans les lignes qui vont suivre, ces questions seront abordées. Dans un premier temps, nous présenterons et justifierons le sujet, la problématique et ses objectifs. Dans un deuxième temps, nous présenterons le Centre MURAZ, ses missions, ses objectifs, ses unités en particulier l'unité AMF. Pour poursuivre, nous montrerons les techniques et les méthodes utilisées aux cours de ce stage. Avant de conclure, nous présenterons les résultats et les méthodologies mises en place.

1. Contexte

La province du Houet créée le 15 Septembre 1983, est l'une des 45 provinces que compte le Burkina Faso. Située à l'Ouest, elle couvre une superficie de 11 540 km², soit 4,21% du pays. Elle fait partie des provinces les plus arrosées du Burkina Faso. Son chef-lieu Bobo-Dioulasso, latitude 11° 11' 00" nord et longitude 4° 17' 00" sud, est la deuxième ville du pays et se situe à 365 km de la capitale Ouagadougou. De superficie 136,78km², la commune de Bobo-Dioulasso comprend une zone urbaine et une zone périurbaine. Elle est organisée en 7 arrondissements regroupant 33 secteurs disposés en spirale, dont les secteurs les plus anciennement construits étant situés au centre (1 à 21) et les nouveaux en périphérie (22 à 33).

Bobo-Dioulasso connaît une croissance rapide de sa population depuis les années 1975. En effet, la population est passée de 100 793 habitants en 1975 à 309 771 habitants en 1996 (RGPH, 1996). Elle a été évaluée à 489 967 habitants en 2006 (RGPH, 2006). Cette croissance est imputable à l'effet conjugué du croît naturel et du phénomène migratoire. En outre, l'espace consacré à la ville ne fait que s'accroître. Il est passé de 84 ha en 1897 à 4 600 ha en 1995². Entre 1995 et 2006, l'extension de l'espace urbain de Bobo-Dioulasso est passée à 136 780 ha qualifiée de lotissement de nouvelles zones sur de grandes étendues.

¹ Liste exhaustive et mise à jour de tous les individus de la population sans omission ni double décompte.

²Socio-économique des villes africaines (Yves-A. Fauré et Pascal Labazée IRD Paris).

2. Justification

L'évolution imprévisible de la population constitue ainsi un problème de développement local. Conscient des problèmes de pressions sur les services sociaux et sanitaires qui ne cessent de croître, le besoin est donc d'avoir des connaissances sur l'évolution exacte de la population et ses caractéristiques afin de mieux prévoir les politiques de bonnes gouvernances. En sondages de population, deux méthodes sont généralement utilisées: les méthodes probabilistes et les méthodes empiriques.

- La méthode probabiliste ou la méthode utilisée par l'INSD

L'INSD, c'est l'organe officiel de l'Etat en matière d'information statistique. Les bases de données sont mises à jours chaque dix (10) ans à l'issu d'un recensement. Le recensement est une opération statistique de dénombrement de la population. Elle nécessite des moyens colossaux tant au niveau humains, matériels que financiers. Le budget du recensement de 2016 est estimé à 19,8 milliards³.

L'INSD réalise aussi souvent des enquêtes dans le cadre du suivi des conditions de vie des ménages à travers des sondages à méthodes probabilistes. Par définition, les sondages probabilistes, sont ceux pour lesquels chaque individu de la population a une probabilité connue, et fixée par avance d'appartenir à l'échantillon. En effet, l'INSD a une cartographie exhaustive de la population qu'elle réalise lors des recensements. Cette cartographie permet un découpage de l'ensemble du territoire en zones de dénombrement (ZD) qui sont utilisées dans les sondages. Il utilise le plus souvent un plan de sondage à deux degrés : au premier degré, sont tirées les unités primaires (UP) encore appelées ZD et au second degré, sont tirées les unités secondaires (US), appelées ménages, avec une probabilité égale pour toutes les US de l'UP. Cependant, ces méthodes ne sont applicables que lorsqu'il existe une liste exhaustive de toute la population. Ensuite, elles peuvent être fastidieuses si l'échantillon et la population sont tous deux de grande taille.

- Les méthodes empiriques

Les méthodes empiriques ou encore méthodes à choix raisonné sont les plus utilisées. On fait appel à elles lorsque nous n'avons pas de base de sondage. Ce sont des méthodes non probabilistes car la probabilité d'être enquêtée de chaque unité statistique n'est pas connue à priori. Elle dépend en fait de l'enquêteur. Les enquêteurs sont laissés libres d'interroger tout sujet, pourvu que la condition de pourcentage (appelle poids) de chaque strate soit remplie. Les méthodes empiriques n'ont pas de fondement théorique, donc on ne peut pas calculer la précision (erreur, intervalle de confiance ...), ils fournissent habituellement des échantillons biaisés.

3. Présentation du sujet

Le présent rapport a été réalisé à partir d'un projet de recherche dont le thème est : « Développement et validation d'un plan de sondage à base aréolaire dans la commune urbaine de Bobo Dioulasso ». On appelle sondage aréolaire un sondage où l'unité d'échantillonnage est constituée par une fraction du territoire occupé par la population étudiée.

³<http://countryoffice.unfpa.org/burkinafaso/>

Notre tâche consistait à réaliser et produire la partie cartographique et la modélisation spatiale et interpolation par krigeage de la population dont le thème est : « **Définition d'une base de sondage par modélisation spatiale** ».

Ce rapport concerne la première phase de la réalisation du dossier. Les résultats finaux sont consultables à travers la direction du Centre MURAZ à l'Equipe d'Appui Méthodologique et de Formation.

4 Objectifs de la recherche

Les objectifs visés dans ce travail de recherche consistent avant tout à :

- réaliser la cartographie de la répartition des lieudits retenus pour la construction de la base ;
- faire ressortir les grands traits de la répartition spatiale de la population de Bobo-Dioulasso;
- localiser la population et faire l'analyse spatiale de ces résultats cartographiques;
- produire une interpolation de la densité de population observée dans les zones échantillonnées par une modélisation spatiale.

5. La zone d'étude

Les données ont été collectées dans la ville de Bobo-Dioulasso, plus précisément dans la limite parcellée de la ville. Cependant, l'interpolation ou l'estimation de la population est faite sur l'étendue du centre urbain. Ce centre urbain ou encore appelé zone d'étude est la partie de la commune exclu les villages rattachés.

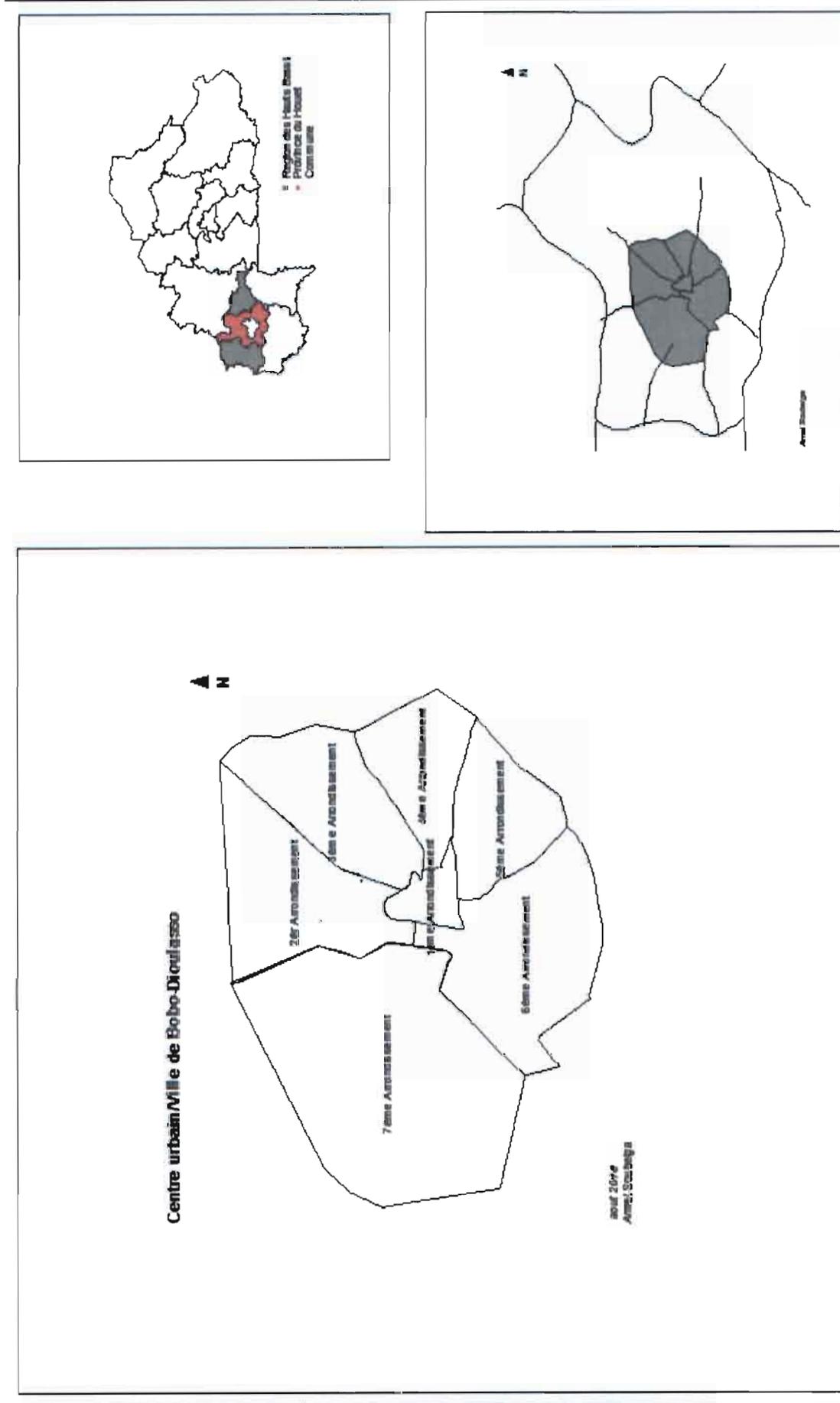


Figure 1 : Présentation de la zone d'étude

II. Présentation de la structure d'accueil : Centre MURAZ

Le Centre MURAZ est un Etablissement Public de Santé (EPS) de l'Etat depuis 2001. Il est placé sous tutelle technique du Ministère de la Santé et sous tutelle financière du Ministère de l'Economie et des Finances.

1. Historique

Créé en 1939 par le médecin colonel français Gaston MURAZ, le Centre MURAZ est l'institution la plus ancienne dans le paysage burkinabè de la recherche scientifique. Il avait comme mission à sa création de lutter contre la trypanosomiase humaine encore appelée maladie du sommeil qui sévissait en Afrique Occidentale Française (AOF) sous la désignation de « Service Général Autonome de la Maladie du Sommeil (SGAMS) ». C'est en 1956 que SGAMS prendra le nom de « Centre MURAZ » en l'honneur de Gaston MURAZ. Au fil des années, sa mission s'élargit et le centre s'est intéressé à la lutte contre certaines grandes endémies qui constituaient les principales préoccupations sanitaires de l'époque.

De 1960 à l'an 2000, le Centre a été intégré et était en même temps le siège de l'Organisation de Coopération et de Coordination pour la lutte contre les grandes Endémies en Afrique de l'Ouest (OCCGE) qui regroupait huit (8) pays d'Afrique de l'Ouest.

À la dissolution de l'OCCGE en l'an 2000 pour la création de l'Organisation Ouest Africaine de la Santé (OOAS), le Centre MURAZ est devenu une institution publique burkinabè de recherche pour la santé.

Ces transformations institutionnelles ont été imprimées par les pouvoirs publics pour garantir au Centre MURAZ les conditions d'une pleine contribution par la recherche au développement sanitaire du Burkina Faso. Aujourd'hui, il s'intéresse essentiellement à des maladies comme le paludisme, les IST/VIH/SIDA et maladies associées (la tuberculose), la fièvre jaune et la méningite.

En décembre 2015, le Centre MURAZ s'est vu un statut particulier. Ce statut permet une meilleure gestion des projets de recherche et donne la possibilité à ses chercheurs de poursuivre leur carrière au niveau du Conseil Africain et Malgache de l'Enseignement Supérieur (CAMES).

2. Missions et objectifs

Les missions essentielles du Centre MURAZ sont :

- La recherche : mission première et aussi la plus connue du centre.
- La formation : participation effective à la formation des personnels de santé à travers des stages pour paramédicaux, doctorants (mémoires, thèses), post-doctorants.
- L'expertise : renforcement des capacités des compétences nationales (en priorité) de sorte à disposer sur le plan technique d'experts de haut niveau. De plus, le centre est sollicité par des organismes tels que l'OMS, L'UNICEF, la Coopération française, le PNUD dans des domaines aussi variés que possible comme l'animation de séminaires, de formation, la planification et l'évaluation des programmes, le contrôle de qualité de laboratoire, etc.

Les principaux objectifs du Centre MURAZ sont entre autres :

- Promouvoir et harmoniser les recherches fondamentales, appliquées ou opérationnelles sur les maladies transmissibles et lutte contre leur transmission.
- Collaborer avec les autres services nationaux pour recueillir et diffuser l'information sanitaire pertinente et la mettre à la disposition des autorités sanitaires nationales et participer à la formation des personnels de santé.
- participer dans la mesure de ses moyens à la coordination de la lutte contre les maladies transmissibles relevant de ses domaines de compétence avec les autres pays de la sous-région et tout particulièrement avec les Etats voisins.

3. Organisation et formation

Etablissement Public à caractère Administratif (EPA) depuis 2001 et établissement Public de Santé (EPS) en 2008, le Centre MURAZ reçoit une délégation de pouvoir du Ministre en charge de la Santé et du Ministre en charge des Finances. Il rassemble aujourd'hui près de 200 agents dont 44 chercheurs. Pour accomplir ses missions. Il peut selon ses besoins :

- créer des directions, des unités de recherche et/ou de production ;
- assurer des prestations de services à but lucratif ;
- exploiter des brevets et licences ;
- conclure des conventions/accords de coopération avec d'autres institutions de recherche au niveau national, régional et/ou international.

Un nouvel organigramme scientifique est mis en place depuis 2015. Il coiffe trois départements de recherche :

- le Département des Sciences Biomédicales (DSB) ;
- le Département de Recherche Clinique (DRC) ;
- le Département de Santé Publique (DSP).

Chaque département est constitué d'unités de recherches.

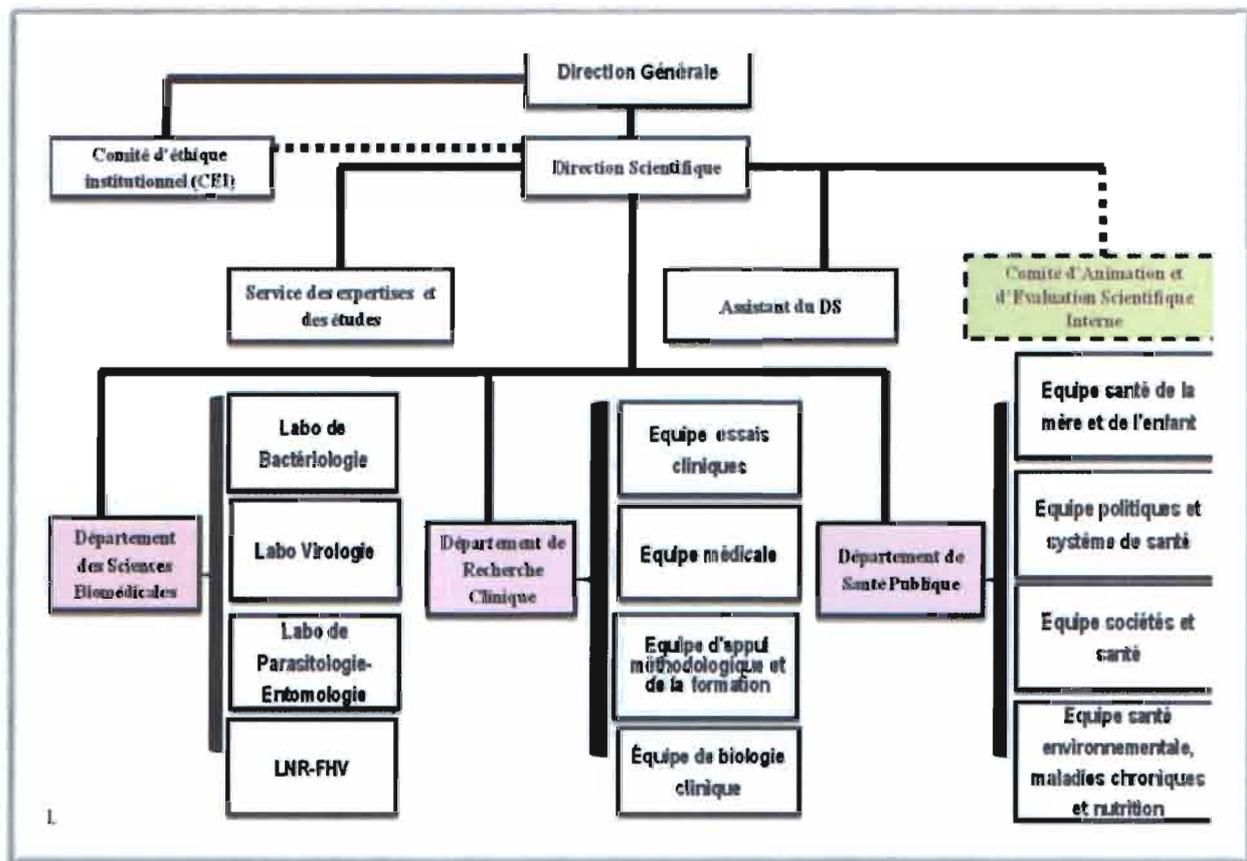


Figure 2 : Organigramme Scientifique

Ces 3 grands départements sont chargés de la coordination de la recherche, de la prospective, de l'animation et de l'évaluation au niveau des unités de recherche et seront de ce fait, l'interface entre les unités de Recherche et la Direction Scientifique.

4. Présentation de l'Equipe d'Appui Méthodologique et Formation (AMF)

Créer en 2004, l'AMF a d'abord pris le nom de Centre de Calcul (CC) et en 2014 l'Unité de Formation et d'Appui Méthodologique (UFAM). C'est dans cette unité que nous avons effectué notre stage de fin de cycle.

L'AMF a été créé pour répondre aux missions de recherche, de formation, d'appui méthodologique, d'expertise scientifique et technique dans une perspective d'excellence et pour répondre au mieux aux priorités nationales et internationales au bénéfice de la santé des populations. Ses objectifs sont les suivants :

- d'assurer une masse critique d'expertise méthodologique hautement spécialisée et multidisciplinaire aux chercheurs du Centre MURAZ et de ses institutions partenaires;
- d'aider les chercheurs à concevoir des projets de recherche pertinents, à effectuer des analyses Biostatistiques, à gérer les données, à valoriser la diffusion des résultats, et à veiller à ce que les études respectent toutes les normes réglementaires pertinentes ;
- apporter son expertise scientifique et technique en méthodologie de la recherche ainsi qu'en organisation, gestion et analyse des données quantitatives.
- répondre aux besoins du centre dans le domaine de la statistique.

III. Matériels et méthodes

1. Collecte des données

1.1 Stratégie d'étude

- La recherche documentaire

Elle consiste à la recherche de documents se rapportant à notre thème d'étude. Cela nous a conduits auprès de plusieurs directions régionales des ministères comme le Ministère de l'Urbanisme, le Ministère de la Santé, la Direction Régionale de l'Institut National de la Statistique et de la Démographie (INSD), les mairies des sept (7) arrondissements. Cette recherche nous a informés d'une part sur l'évolution urbaine et celle de la répartition de la population de Bobo-Dioulasso et d'autre part sur différentes méthodes d'acquisition des données statistiques.

- La collecte de données primaires

Il s'agit du repérage de points remarquables dans la ville de Bobo-Dioulasso. Un certain nombre de renseignements ont ensuite été collectés autour de ces points. Les données collectées sur le terrain ont été étayées en triangulant avec celles de la Base Nationale des Données Topographiques (BNDT) telles que, les couches de la trame urbaine, les couches des limites des provinces, communes et régions. Nous avons bénéficié de l'appui des chercheurs de l'Unité Environnement Santé (UES) du Centre MURAZ pour l'accès et l'exploitation de ces données. C'est à partir de cette base de données que toute l'analyse a été menée.

1.2 Stratégie d'observation

- Population et unité type

En statistique, la population ou référentiel statistique est un ensemble « désigné » sur lequel porte l'étude. Dans notre étude, la base aréolaire, est constituée d'une liste de points enquêtés, en d'autres termes, elle est formée des coordonnées géographiques d'un ensemble de lieudits représentatifs de la commune. Ainsi la population cible serait ici l'ensemble des lieudits⁴ de la commune et il faudra toutefois en exclure certains en raison de contraintes opérationnelles et afin de ne garder que la population à observer. Les éléments de cette population observée sont alors les unités type.

- Plan d'échantillonnage ou critères de sélection des lieudits

L'échantillonnage permet ici le passage de la population cible à la population observée. Son objectif est donc de fournir un échantillon qui sera « représentatif »⁵ de l'ensemble de la population.

Cependant une question particulière mérite d'être posée : quel est le nombre de lieudits représentatifs ? Cette question répond à deux soucis. D'une part, s'il était suffisamment important il assurerait une plus grande dispersion de l'échantillon et donc une meilleure représentativité de l'agglomération.

⁴ En fonction de la définition d'un lieudit établie dans la partie Concepts et Définitions. Ceci devrait être une phrase complète.

⁵ Qui représentera la population cible et reproduira aussi fidèlement que possible ces principales caractéristiques.

D'autre part, plus le nombre de lieudit est important, plus les coûts des travaux de dénombrement sont élevés ainsi que le temps de travail consacré. Le choix du nombre de lieudit établi alors un équilibre entre :

- Ce qui est exigé du point de vue de la précision d'un échantillonnage ;
- Ce qui est réalisable du point de vue de l'application pratique (exemple: budget, personnel de terrain et administratif, ressources techniques, contrôle de qualité, contraintes temporelles, gérable, pérennité).

Pour répondre à cette double préoccupation, on s'est appuyé sur la dispersion (la représentativité) des lieudits recensés, qu'on avait observée en groupes de catégories en tenant compte des critères comme la répartition administrative et spatiale, le peuplement ainsi que les caractéristiques de ces différentes catégories.

- La commune urbaine de Bobo-Dioulasso est répartie en sept (7) arrondissements administrés chacun par une mairie. Les mairies constituent des unités types représentatives.
- La commune compte 43 pharmacies suivant une logique de dispersion, selon les critères d'implantation pour l'ouverture d'une officine pharmaceutique définis dans les arrêtés N°2003-147/MS/CAB. Cet arrêté stipulait que la distance minimale entre deux officines voisines est de 500 mètres et une officine pour 30 000 habitants pour les localités de moins de 250 000 habitants. Dans cette logique, les 43 officines de la ville donnaient une distribution spatiale assez représentative.
- La commune de Bobo-Dioulasso dispose d'un Centre Hospitalier Universitaire (CHU) et deux districts (CMA) soutenant les CSPPS. Sur le territoire communal, on compte 30 CSPPS répartis entre les secteurs urbains et les villages. Chaque CSPPS de la ville est responsable d'un ou plusieurs secteurs.

La figure ci-dessous montre alors la dispersion spatiale et la représentativité des lieudits enquêtés, constituant ainsi la population observée.

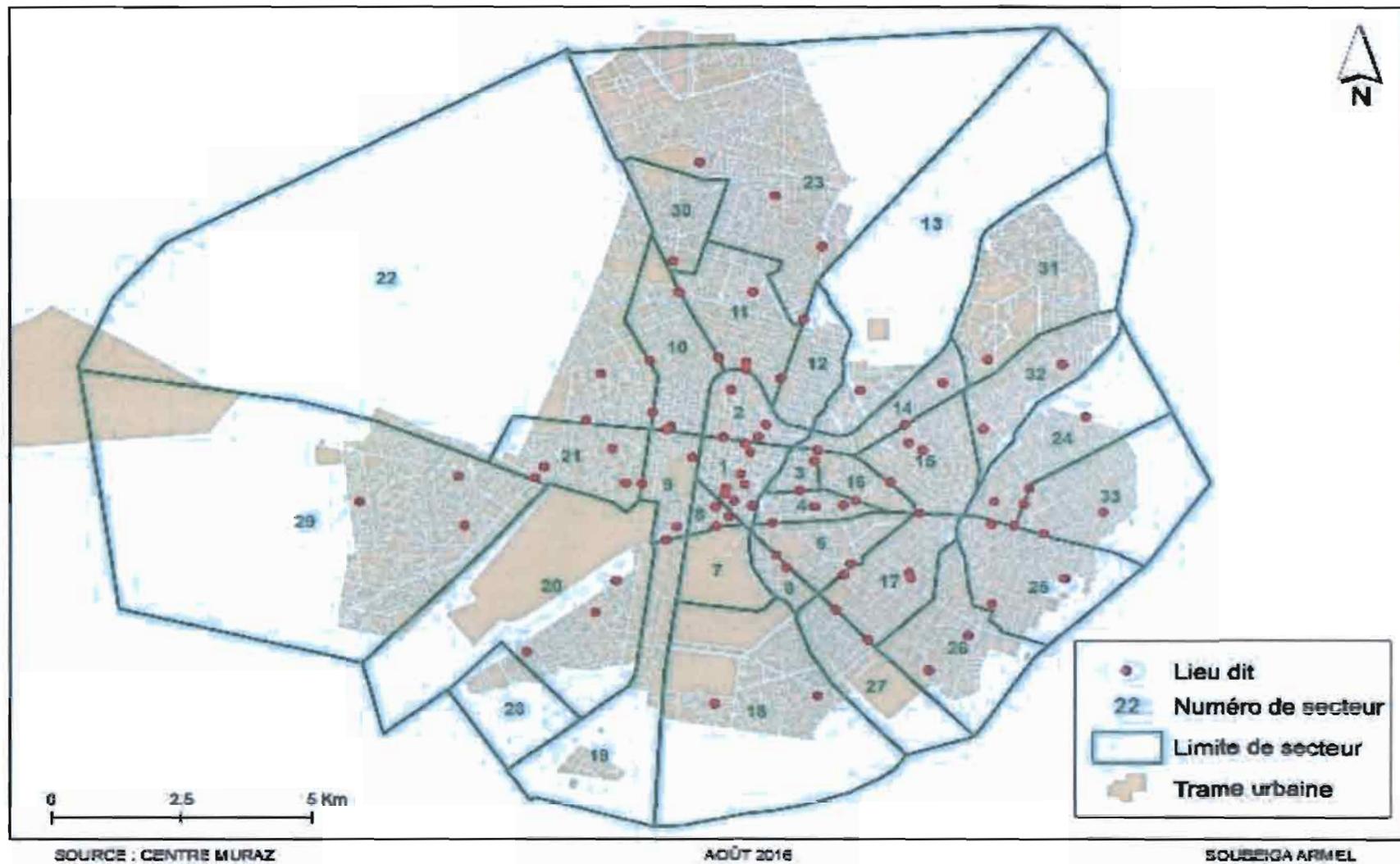


Figure 3 : Répartition spatiale des lieux-dits observés

- Méthode d'observation

La population observée était finalement constituée de 85 lieudits. Ces lieudits sont uniquement observés dans la zone parcellée de la ville, ceci afin de faciliter la méthode appliquée pour l'estimation de la densité de population associée à chaque lieu. En effet, cette méthode consistait, pour chaque unité d'enquête (lieudit) de dénombrer les parcelles d'habitation situées dans un rayon de cent mètres (100 m), c'est-à-dire sur une superficie de 3 ha 14 a, soit 31 415 km² (31 415 000m²) et sur une circonférence de 628 m autour de ce lieu. Ceci a été possible grâce aux outils et aux idées procurées par le département informatique de la Direction Régionale du Ministère de l'Habitat et de l'Urbanisme.



Figure 4 : *Méthodes de dénombrement des parcelles d'habitation par lieudits*

Ensuite une observation spatiale et satellitaire (image satellitaire datant du 06/2016) de la commune à travers le logiciel Google Earth a permis d'identifier parmi ces parcelles d'habitation celles qui étaient construites (sensées être occupées).

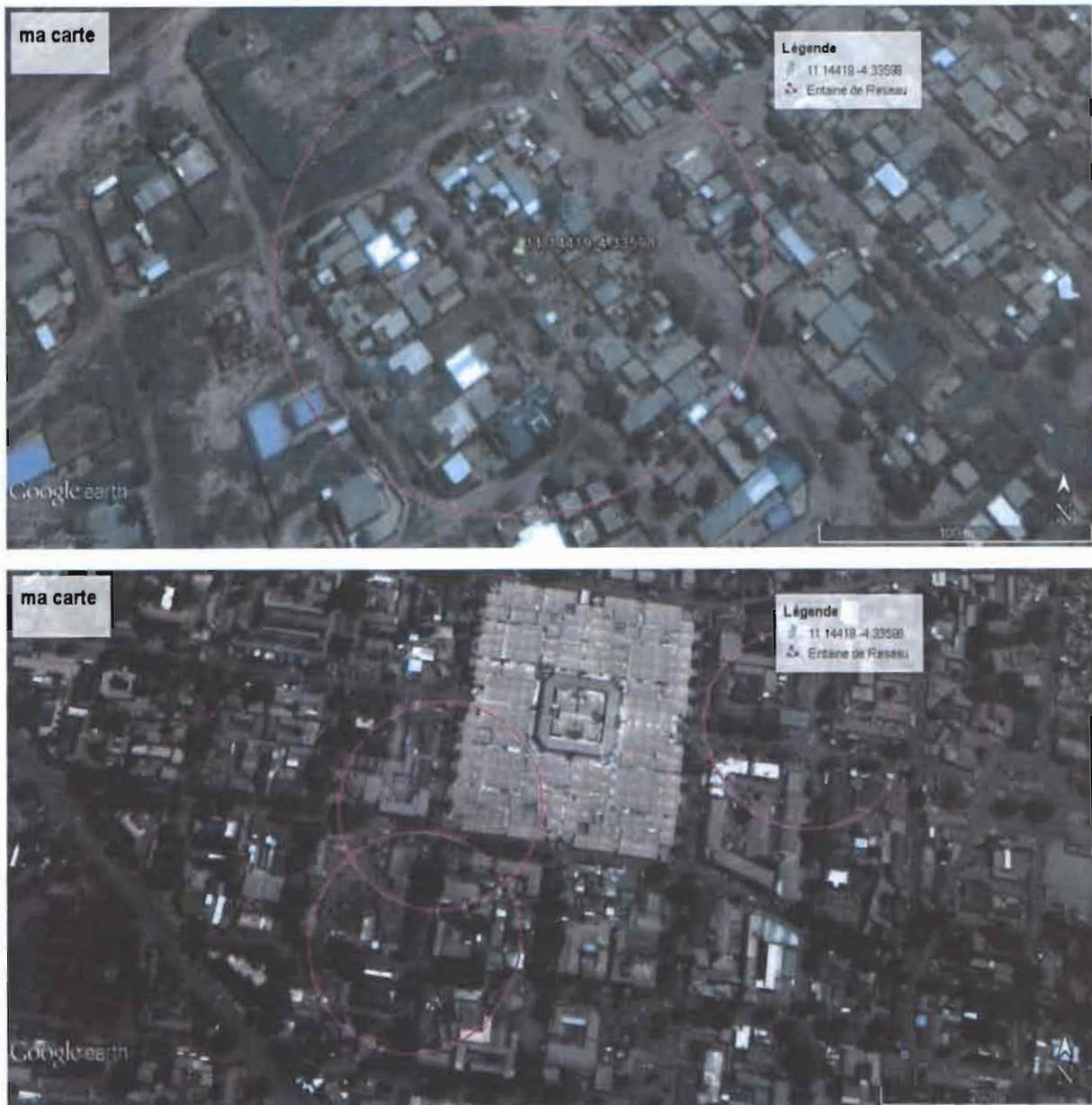


Figure 5: Observation des Lieudits sur Google Earth

- Tailles de parcelles et modes d'occupation

Du point de vue démographique, le concept de ménage est étroitement associé au concept de concession ou d'habitation qui reflète ses caractéristiques socio-économiques et culturelles. En effet, en fonction de chaque type d'habitat est associé un ménage ou groupe de ménages. Les types d'habitats dominant de la ville de Bobo-Dioulasso soit 97,2 % sont composés entre autres de maisons individuelles simples, des bâtiments à plusieurs logements, et des villas. (INSD 2006).

Tableau 1 : Répartition des ménages en fonction du type d'habitat

Type d'habitat	Proportions (%)	Nombre moyen de ménages
Maison individuelle simple	57,4	1
Bâtiment à plusieurs logements	29,2	-
Villa	10,6	1,2
Autres	2,8	1

Les anciens lotissements possèdent des surfaces de parcelles élevées, jusqu'à 1400 m² en moyenne et les surfaces des parcelles pour les zones récemment loties ont une moyenne de 375 m². Dans le cas des concessions abritant plusieurs ménages (Bâtiment à plusieurs logements), il existe souvent un ménage principal, généralement propriétaire de la parcelle, qui décide des investissements sur la concession. Ainsi 60% des concessions sont occupées par un seul ménage, en moyenne 2,5 ménages dans 23 % des concessions et en moyenne 5 ménages dans 17 %. (PS-EAU, PDM 2001)

Tableau 2 : Répartition des ménages pour un habitat à plusieurs logements

Nombre moyen de ménages	Proportion (%)
1	60
2,5	23
5	17

La taille moyenne des ménages est le nombre de personnes qui y vivent habituellement. Au niveau national, cette taille est estimée à 5,9 personnes par ménage et la ville de Bobo-Dioulasso compte 94 947 ménages de taille moyenne de 5,2 personnes par ménage. (INSD 2006).

1.3 Stratégie d'équipements

- Définition opérationnelle des variables et le questionnaire

Une fois les objectifs, la population et la structure d'étude fixés, les variables à étudier sont sélectionnées. Le choix des variables doit être conforme aux objectifs de l'étude. Il s'agit entre autres des statistiques relatives à la situation géographique de chaque lieu-dit (localisation) et la spécificité de chaque lieu (caractéristique).

A. Identification

Identifiant de l'enquêteur

Identifiant du Lieu-dit

Date de collecte

B. Localisation du Lieu-dit

Dans quel arrondissement se trouve le lieu-dit ?

Dans quel secteur se trouve le lieu-dit ?

La Longitude du lieu-dit

La Latitude du lieu-dit

Le type de zone où se trouve le lieu-dit

1. Industrielle
2. Administrative
3. Commerciale
4. Habitation

C. Caractéristiques du Lieu-dit

La nature du lieu-dit

1. Pharmacie
2. CSPPS
3. Marie
4. Autre

Photographie (photo du lieudit)

Le questionnaire a été programmé avec le logiciel CSPro version 6.3 incorporé dans des tables Android TABLE 3. Ceci a permis de faciliter la phase de collecte et de traitement à travers les codages.

- **La collecte et Traitement des données**

Avant de passer à la phase de réalisation, un test de la validité du questionnaire et des méthodes de collecte ont été réalisés sur deux sites (la pharmacie Sifoma et le CSPPS de Colsama) à la date du 22/05/2016. Cette enquête pilote a en effet permis de tester la faisabilité de l'étude, de valider à petite échelle les méthodologies, puis de les adapter à la mise en place de l'étude sur l'ensemble de tous les sites de la ville. La collecte s'est déroulée en deux étapes :

La première étape s'est déroulée en trois (3) jours, du 30/05/2016 au 01/06/2016. Cette étape a permis de couvrir la zone centrale de la ville et la liste s'est arrêtée à 67 lieudits enquêtés. Une deuxième collecte complémentaire sur 17 sites a été réalisée le 05/07/2016 afin d'assurer une meilleure répartition sur toute l'étendue de la trame parcellaire de la ville.

En ce qui concerne le traitement, il est parti du recueil des données d'enquête de la tablette en fichier Excel ainsi qu'aux différentes corrections telles que les fautes d'orthographe des noms, la concordance des données pour aboutir aux analyses.

- Logiciels

Nous avons utilisé les logiciels ci-dessous :

CSPro 6.3 et Excel 2013 : pour la programmation du questionnaire et le traitement des données.

Arc GIS 10 et R 3.1.3 : pour la cartographie et l'analyse spatiale.

- Photographies des zones échantillonnées



Figure 6 : Exemple de photos de quelques lieux échantillonnés

2. Analyse statistique et cartographique

2.1 Analyse cartographique

La carte constitue un graphe particulier de l'analyse des données. Une carte statistique permet l'étude des caractéristiques d'une population traitées dans des portions de l'espace. Les types de cartes proposés ici ont pour but d'offrir une visualisation de la dispersion des sites échantillonnés et la répartition des différentes valeurs associées dans l'espace.

Spatialisation des lieudits en fonction des densités

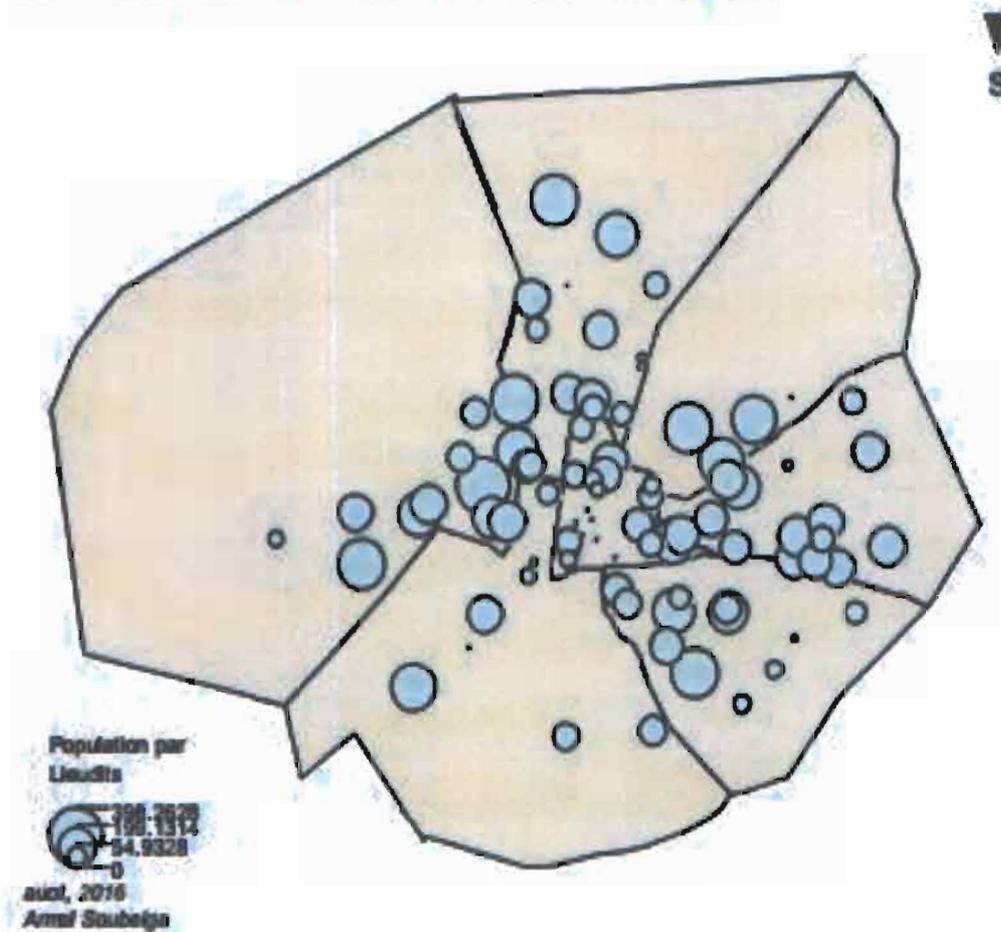


Figure 7 : Spatialisation des lieudits en fonction de la densité

La carte des densités de population par lieudits fait ressortir des auréoles de valeurs décroissantes autour des nœuds de regroupements urbains. L'interprétation de ce fait permet d'émettre l'hypothèse que la distance entre le centre-ville aux nœuds urbains est un élément de compréhension et d'explication des répartitions des densités de population.

2.2 Analyse descriptive

Cette étude exploratoire a pour but d'apprécier la distribution des données dans l'espace, d'appréhender leur degré d'homogénéité, de rechercher et de visualiser les observations atypiques ou tout simplement de se familiariser avec la variable d'intérêt. Il est ici rappelé que la variable d'intérêt étudiée est la densité de population.

- Histogramme

L'étude de l'histogramme permet :

- d'apprécier la variabilité et la stationnarité des données ;
- de détecter d'éventuelles valeurs aberrantes.

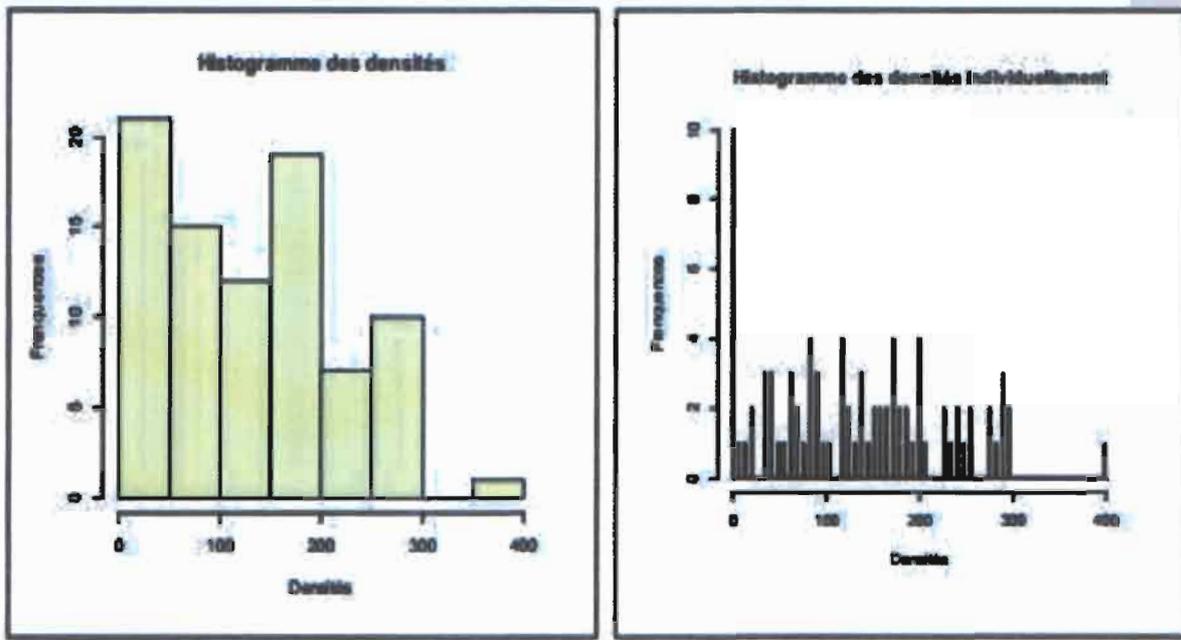


Figure 8 : Histogramme des densités

L'histogramme des densités associé aux zones échantillonnées (lieudits) n'est pas symétrique. Son examen visuel permet de noter une étendue importante d'environ 400. Mais toutefois, on observe des valeurs aberrantes dans les deux extrêmes. En effet, on observe que la fréquence des lieudits dont la densité est nulle est plus de 10%. De plus, la densité autour de la « pharmacie Bethel » est de l'ordre 25% plus grande (400) que les autres. La géostatistique indique que ces valeurs aberrantes ont un poids sur la stationnarité, donc sur les hypothèses de krigeage. Dans le but d'avoir un échantillon stationnaire, ces sites seront exclus de l'étude.

- Histogramme des données corrigées

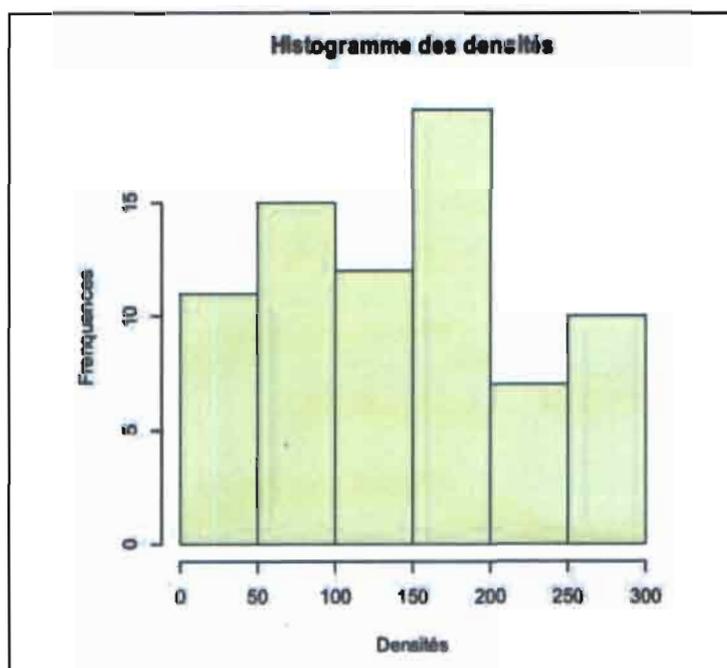


Figure 9 : Histogramme des densités après correction

Après correction, on observe que les données semblent stationnaires (se ressembler) et ont l'allure d'une loi Gaussienne.

- La droite de Henry

La droite de Henry des données ajustées : si les probabilités sont alignées le long de la droite de Henry, nous pouvons supposer que la variable suit une distribution gaussienne. En effet, la loi gaussienne ou normale est l'une des lois de probabilités les plus adaptées pour modéliser des phénomènes naturels issus de plusieurs événements aléatoires.

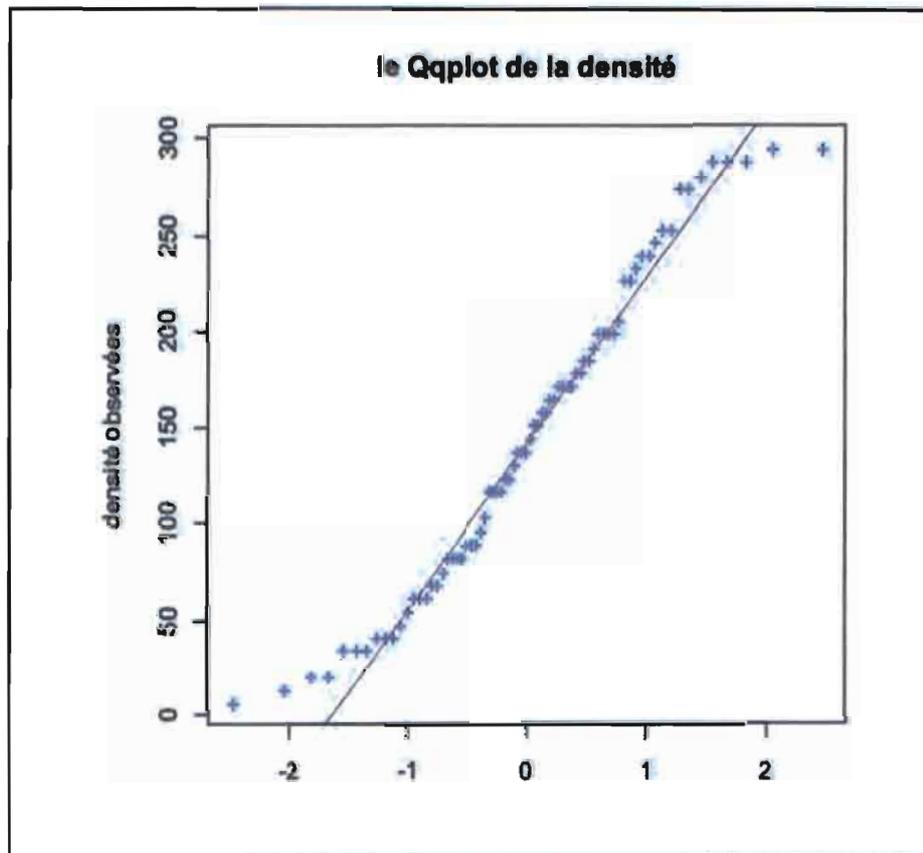


Figure 10 : La droite de Henry

La droite de Henry montre que la distribution des densités de population suit une loi Gaussienne. Cela est confirmé par le test de Kolmogorov-Smirnov de normalité.

```
One - sample Kolmogorov - Smirnov test
data: pop
D = 0.0921, p - value = 0.5571
alternative hypothesis: two - sided
```

Le test n'étant pas significatif, car $p - value > 5\%$, l'hypothèse nulle de distribution normale est alors conservée.

- Commentaire 1

La droite de Henry et le test de normalité de Kolmogorov-Smirnov montrent que la distribution des données ajustées suit une loi Gaussienne. Celle-ci n'est pas une condition obligatoire en géostatistique linéaire, mais toutefois d'après (Cressie 2015), le krigeage a tendance à fournir de meilleures prévisions lorsque les données suivent une loi normale. Cressie recommande alors soit une transformation des données par une fonction mathématique pour rendre la distribution gaussienne, soit discrétiser en zones géographiques afin que les données de chaque zone soient proches les unes des autres.

- Nuée Variographique omnidirectionnelle

Pour étudier le comportement dans l'espace des valeurs observées, le nuage variographique ou nuée variographique est un outil exploratoire synthétique qui permet de visualiser les densités de population le long de toutes les directions de coordonnées.

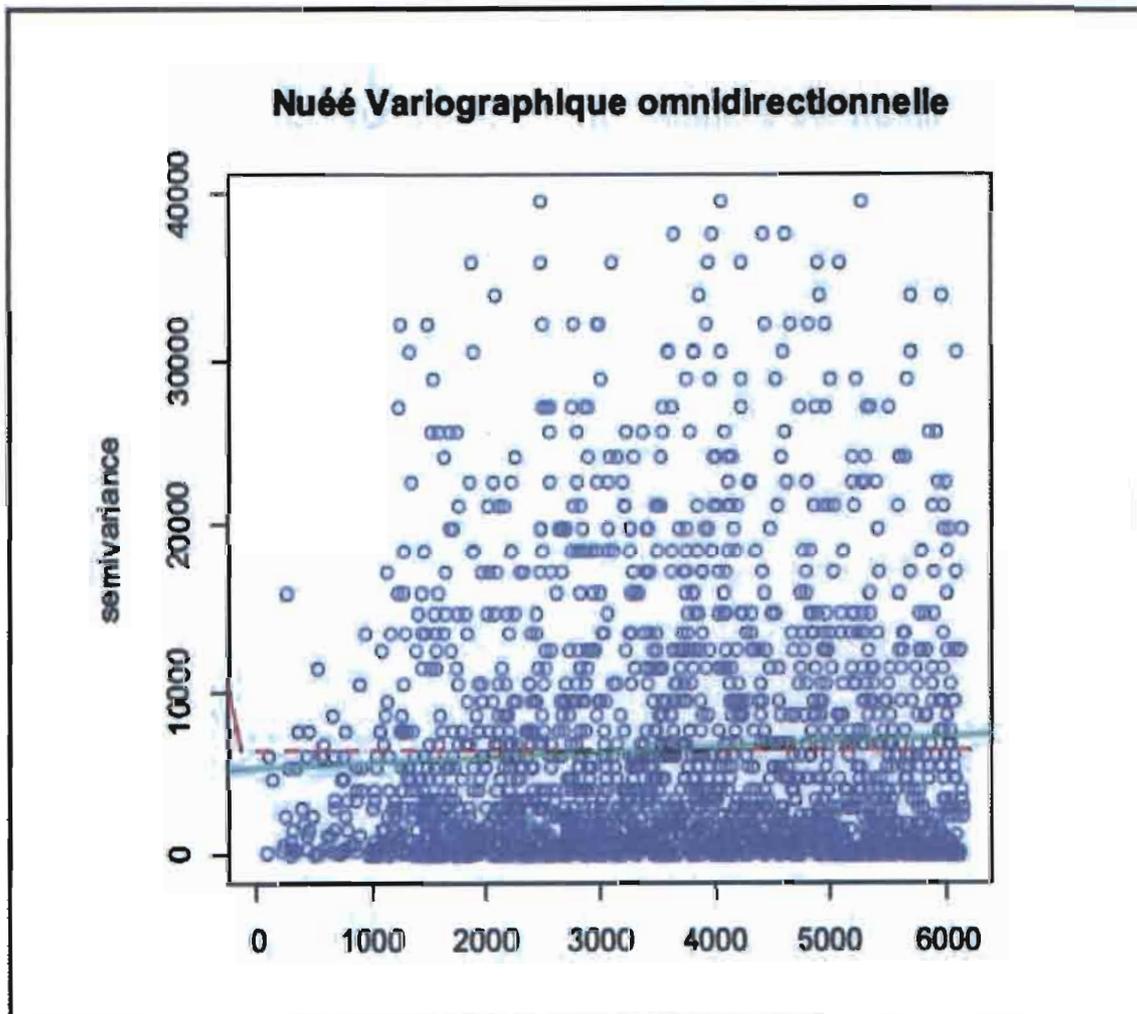


Figure 11 : Nuée variographique omnidirectionnelle

Le tracé de l'évolution des densités en fonction de la longitude et de la latitude est un moyen de détecter la présence d'une dérive spatiale qui, si elle existe, devrait se traduire graphiquement par une croissance ou une décroissance du nuage de points.

En effet, on observe que le nuage de points de la variable d'intérêt est relativement homogène autour de la moyenne (les pointillés en rouge). En revanche, les valeurs ont une légère tendance à augmenter

lorsque l'on se dirige du sud vers le nord. Une hypothèse de stationnarité serait alors adéquate dans toutes les deux directions sud-nord et est-ouest.

- Scattergramme

Pour une distance de séparation h donnée, le scattergramme est le nuage des points formés par les couples de valeurs $(Z(s_i), Z(s_i + h))$; $Z(s_i)$ représente la densité de population au site s_i et h la distance qui sépare un site à un autre. Pour que cet outil soit pertinent, il faut que les observations soient équitablement réparties sur l'ensemble du champ. Plus les valeurs prises aux sites s_i et $s_i + h$ sont semblables, plus le nuage est proche de la première bissectrice. Par contre, plus ces valeurs sont différentes, plus le nuage est diffus.

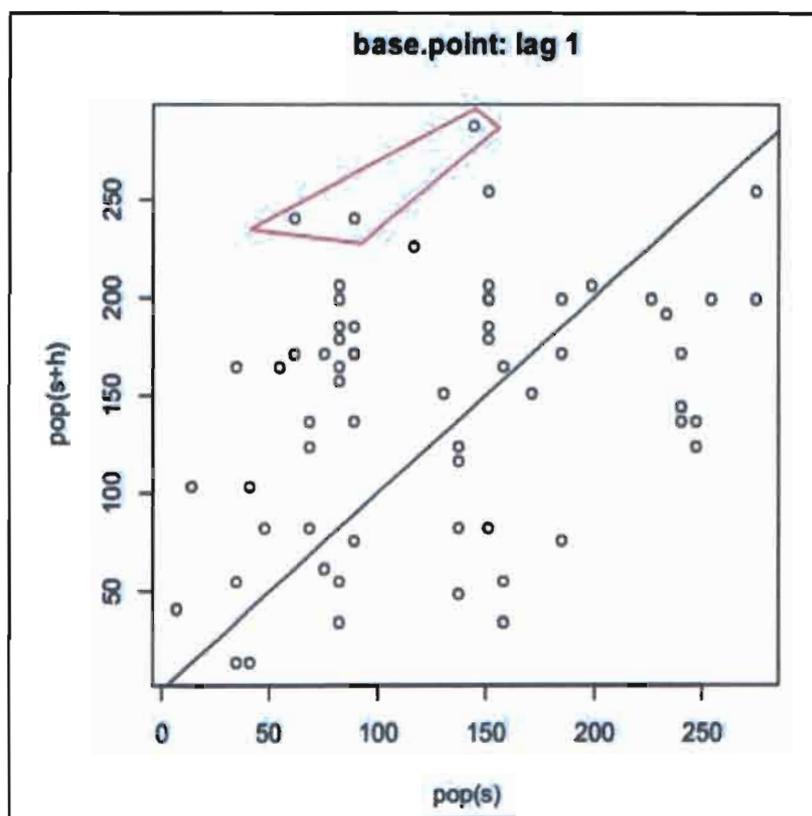


Figure 12 : Le Scattergramme

L'observation du scattergramme montre trois (3) couples de valeurs éloignées de la bissectrice par rapport aux autres. Cela signifie que la variance de la densité entre ces couples de sites est grande. La géostatistique suggère de tenir compte de cette dissemblance lors du krigeage s'il arrive que la fonction de covariance ne l'écarte pas.

- Commentaire 2

L'étude exploratoire nous a permis de nous familiariser avec les données. Elle montre comment la distribution de la densité de population se rapproche d'une loi gaussienne. Enfin, les nuages des valeurs le long des directions de coordonnées (nuée variographique omnidirectionnelle) et les nuages de corrélations différées (Scattergramme) permettent de vérifier la ressemblance des densités dans un certain rayon autour d'une observation.

3. Analyse Spatiale

3.1 Définition

L'interpolation spatiale est une procédure qui consiste à estimer la valeur d'un attribut pour des sites non échantillonnés. Elle s'applique à des variables régionalisées, appelées variables aléatoire, c'est-à-dire des fonctions numériques qui prennent leurs valeurs dans des régions délimitées de l'espace appelées champ. Dans cette étude, la variable régionalisée est la densité de la population mesurée dans la commune urbaine de Bobo-Dioulasso (champ).

En effet, nous utilisons l'interpolation spatiale pour mettre en œuvre des algorithmes mathématiques ou probabilistes afin d'estimer la densité de la population entre les sites échantillonnés (les lieudits). Ainsi, le variogramme est l'outil géostatistique utilisé pour analyser la continuité spatiale du phénomène étudié. Il s'appuie sur une méthode de modélisation probabiliste appelée le Krigeage.

Pour appliquer ce type d'interpolation par Krigeage, il est nécessaire de se familiariser aux données qui seront utilisées.

3.2 Types de données spatiales

Les données spatiales expriment toutes les informations concernant la localisation des objets référencés géographiquement. Pour les besoins de cette étude, nous disposons des coordonnées géographiques (longitude, latitude) des lieudits, des polygones, des ZD et le plan parcellaire de la commune urbaine de Bobo-Dioulasso.

Descriptions :

- Longitude en degrés décimaux, mesuré par rapport au méridien d'origine (Greenwich)
- Latitude en degrés décimaux, mesuré par rapport à l'équateur
- Longitude et Latitude en UTM : les coordonnées en degrés décimaux, projetées en zone NORD 30 pour obtenir des coordonnées métriques afin de pouvoir effectuer les calculs de distances.
- UTM : projection **Transverse Universelle de Mercator** (en anglais Universal Transverse Mercator) est un type de projection conforme de la surface de la Terre. Le système de coordonnées UTM permet de représenter les coordonnées géographiques d'un point du globe terrestre au moyen d'un quadrillage basé sur la projection dite de Mercator. Par exemple, le territoire Burkina Faso est situé sur le fuseau UTM Nord 30 : entre 6 degrés ouest et 0 degré Greenwich.
- Système de référence : le Burkina Faso est situé dans le système World Geodetic System WGS84.
- Points: localisation ponctuelle, GPS et adresse géocodées
- Lignes: ensemble de points ordonnés, connectés par des segments de droites
- Polygones: zones définies par des lignes fermées
- Grilles: ensemble de points ou de cellules rectangulaires organisés en un treillis régulier

3.3 Notation

Z : La variable régionalisée (densité de la population)

D : Le champ (domaine sur lequel la variable régionalisée est définie : la zone d'étude)

s : Un site dans le champ $s \in D$ (un lieu dit)

$Z(s)$: Une valeur prise par la variable régionalisée au site s (la densité en un lieu dit)

h : La distance qui sépare deux sites

$\gamma(h)$: La fonction variographique ou le variogramme

$C(h)$: La fonction de covariance ou le covariogramme

3.4 Hypothèses de base

La fonction de densité de population de la suite des variables $Z(s)$, $s \in D$ est caractérisée par la loi spatiale F qui correspond à la loi de probabilité conjointe de $(Z(s_1), Z(s_2), \dots, Z(s_n))$.

$$F(V_1, V_2, \dots, V_n) = P\{(Z(s_1) < V_1), (Z(s_2) < V_2), \dots, (Z(s_n) < V_n)\}$$

Cependant étant donné que cette fonction est très complexe par l'infinité des combinaisons possibles, la géostatistique linéaire se limite à la fonction de distribution d'ordre un $F(v)$, et d'ordre deux $F_{Z(s_i), Z(s_j)}(V_i, V_j)$.

La première fonction de distribution d'ordre un ($F(v)$) nous permettra de calculer l'espérance de la variable aléatoire Z en un lieu dit s en supposant que :

- L'espérance mathématique ne dépend pas de s

$$E[Z(s)] = m.$$

En d'autres termes,

- L'espérance des écarts est zéro

$$E[Z(s) - Z(s + h)] = 0.$$

La deuxième fonction nous fournira la covariance pour quantifier le degré de ressemblance ou de dissemblance, ainsi que la loi de probabilité entre les valeurs prises aux sites s_i et s_j . Mais le problème qui se pose est que $Z(s)$ n'est observée qu'une seule fois dans une zone précise. Or pour estimer les moments d'ordre deux, il nous faudrait un grand nombre (>30) d'observations. Pour pallier à ce problème, nous posons comme hypothèse que la variable régionalisée est stationnaire. C'est à dire que deux paires de sites distants d'une même distance h ont des caractéristiques semblables. On peut également dire que la variable régionalisée ne dépend pas de sa position dans l'espace, elle garde les mêmes caractéristiques où que l'on se place. Cela nous permet de nous détacher de la localisation et de nous restreindre uniquement à la distance qui sépare les points d'observations.

- La covariance entre $Z(s)$ et $Z(s + h)$ ne dépend que de h (stationnarité du second ordre),

$$Cov[Z(s), Z(s + h)] = C(h),$$

Ou bien

- Le variogramme $\gamma(h)$ ne dépend pas de la localisation s , mais seulement de h ,

Ou

- La variance entre $Z(s)$ et $Z(s + h)$ est minimale pour tout h .

3.5 Analyse variographique

3.5.1 Le variogramme théorique

L'analyse variographique est l'étape préalable au krigeage. Elle permet d'explorer l'éloignement et la structure spatiale des sites échantillonnés afin d'estimer la fonction de la continuité spatiale de la variable d'intérêt. L'outil principal permettant cette analyse est le semi-variogramme qui correspond à l'évolution de la demi-variance en fonction de la distance entre les mesures des différents sites. Il est défini de la manière suivante :

$$\begin{aligned} \gamma(h) &= \frac{1}{2} V[Z(s) - Z(s+h)] \\ &= \frac{1}{2} E \left[(Z(s) - Z(s+h))^2 \right] , \quad s \in D \end{aligned}$$

La formule implique de calculer la différence de l'espérance mathématique au carré entre les valeurs des emplacements couplés. Cette variance devrait être plus petite lorsque les points sont rapprochés jusqu'à s'annuler. Par exemple, l'image ci-dessous illustre l'appariement d'un point (le rouge) avec tous les autres emplacements mesurés. Ce processus se poursuit pour chaque point mesuré.

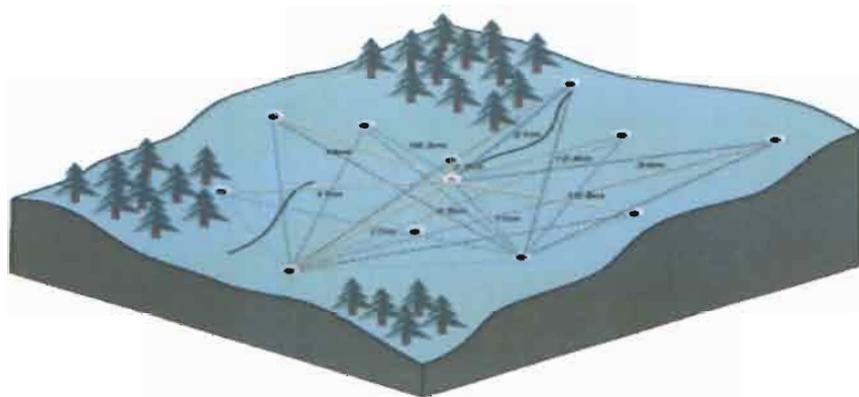


Figure 13 : Exemple du processus variographiques

Pour $h = 0$, $\gamma(0) = V[Z(s) - Z(s)] = 0$

En effet, il existe plusieurs paires de points. Il devient alors impossible de tracer rapidement toutes les paires. Plutôt que de tracer chaque paire individuellement, ces paires seront regroupées dans des groupes de distance.

Par exemple sur R, on a :

```
> pop.vgm
  np  dist      gamma dir.hor dir.ver id
1 35 409.1395 3100.462  0  0 var1
2 99 1062.6616 4592.143  0  0 var1
3 213 1671.3497 5505.615  0  0 var1
```

⁶<http://desktop.arcgis.com/fr/arcmap/10.3/tools/3d-analyst-toolbox/how-kriging-works.htm>, consulté le 08 septembre 2016.

Pour la classe de distance [49509.13 ; 1062.6616 [, on observe 99 paires de sites. Le processus a donné au total quinze (15) classes de distance pour les paires possibles.

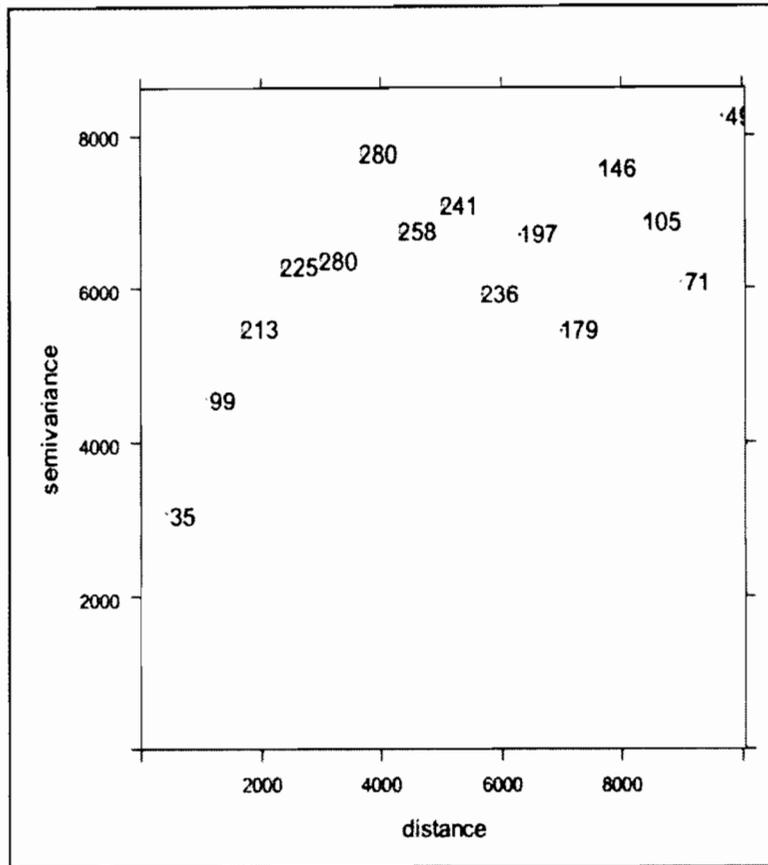


Figure 14 : Nombre de paires de lieudits pour chaque classe de distance (points du variogramme)

3.5.2 Estimation du variogramme

Le variogramme réel d'une fonction aléatoire est généralement inconnu, mais il peut être évalué à partir des données d'échantillonnage. A l'aide des mesures de densité obtenues au niveau des différents lieudits, on estime le variogramme expérimental par la méthode des moments (méthode qui consiste à estimer l'espérance par la moyenne empirique). Cet estimateur est défini de la façon suivante :

$$\hat{\gamma}(h) = \frac{1}{2|N(h)|} \sum_{N(h)} [Z(s) - Z(s+h)]^2$$

Le variogramme empirique représente un diagramme affichant les valeurs de semi-variogramme moyennes omnidirectionnelle (la continuité étant observée dans toutes les directions) sur l'axe des y et la distance (ou décalage) sur l'axe des x.

Note :

Un point du variogramme n'a de poids que lorsqu'il admet un nombre élevé de paires de points. On accorde plus de poids aux premiers points du variogramme pour lesquels h petits. Ce sont ces points qui ont plus d'impact dans les calculs géostatistiques. (Hennequi 2010)

Pour une distance maximale (Cutoff= 10000), on observe 15 classes de distance comprises entre 0 et le Cutoff. À chaque classe de distance, on associe le nombre de points inter distances de cette classe.

On essaie d'avoir $N(h) \geq 0$ pour chaque point expérimental du variogramme. Pour les classes où le nombre de paires $N(h)$ ne vaut pas 30, on en porte peu d'importance.

3.5.3 Le Covariogramme ou l'auto-corrélation

D'après Marcotte (cours de «géostatistique minière»), la nature n'est pas entièrement imprévisible. Deux observations situées l'une près de l'autre devraient en moyenne se ressembler davantage que deux observations éloignées. On ne pourra estimer la densité en un endroit qu'à travers les sites qui lui sont semblables. Le semi-variogramme empirique donne des informations sur l'auto-corrélation spatiale qui se définit comme suit :

$$\begin{aligned}\gamma(h) &= \frac{1}{2} V[Z(s) - Z(s + h)] \\ &= \frac{1}{2} E[(Z(s) - Z(s + h))^2] - \frac{1}{2} E[Z(s) - Z(s + h)]^2.\end{aligned}$$

Or, selon les hypothèses de bases définies sur une variable aléatoire stationnaire,

$$E[Z(s) - Z(s + h)] = 0, \quad \forall s \in D \text{ et } \forall h.$$

D'où

$$\begin{aligned}\gamma(h) &= \frac{1}{2} E[(Z(s) - Z(s + h))^2] \\ &= \frac{1}{2} E[(Z(s))^2 - 2Z(s)Z(s + h) + (Z(s + h))^2] \\ &= \frac{1}{2} [V(Z(s)) + V(Z(s + h)) - 2COV(Z(s), Z(s + h))]\end{aligned}$$

De même, selon les hypothèses de base sur la variance de Z , $V(Z(s)) \cong V(Z(s + h))$.

Par conséquent,

$$\begin{aligned}\gamma(h) &= \sigma^2 - COV(Z(s), Z(s + h)) \\ &= \sigma^2 - C(h)\end{aligned}$$

Avec

$C(h)$: La fonction de covariance ou covariogramme

σ^2 : La variance de la variable d'intérêt.

Le covariogramme quantifie un principe élémentaire de la géographie : le fait que les choses qui sont plus proches se ressemblent davantage que les choses qui sont éloignées. Par conséquent, les paires d'emplacements plus proches devraient avoir des densités similaires.

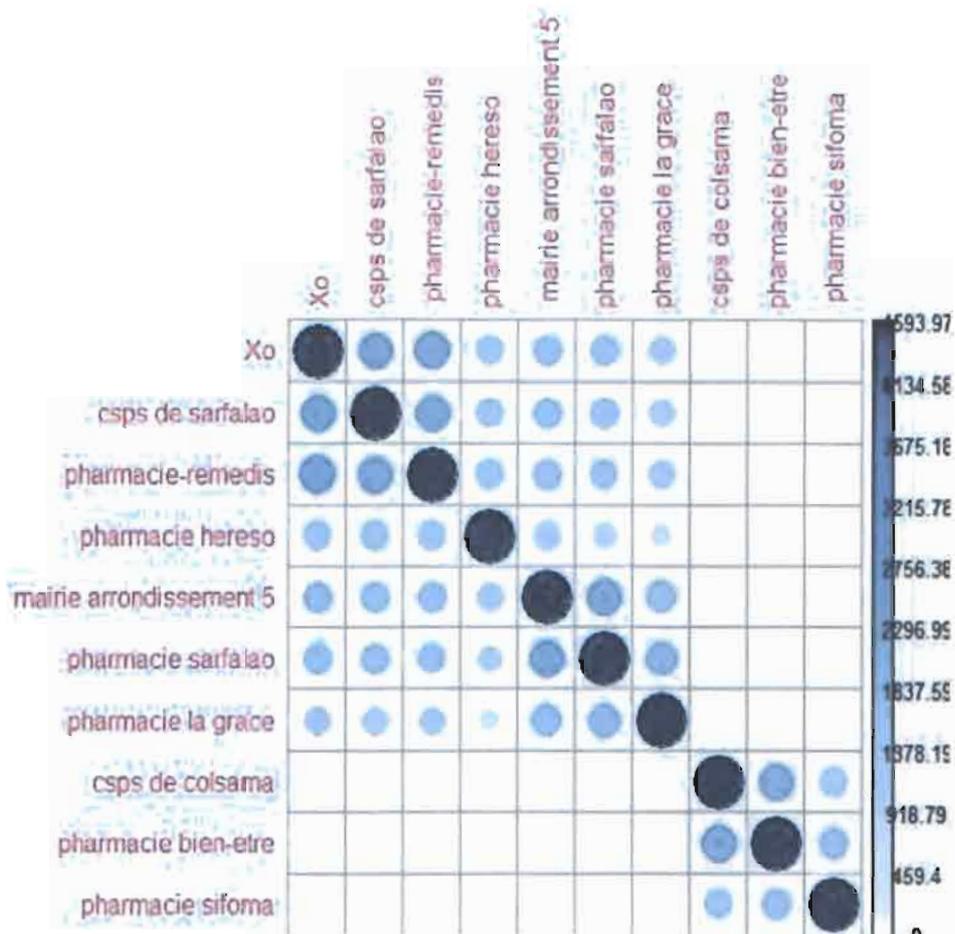


Figure 15 : Exemple du corrélogramme d'estimation à Xo avec 9 sites

Pour l'estimation de la densité au site Xo, le variogramme ci-dessus va prendre en compte les lieudits qui sont plus proche de Xo. En d'autre terme, seul les sites semblables à Xo qui seront pris dans estimation. Il s'agit, allant du plus important (le csps de sarfalao) au site le plus distant tel que pharmacie la grâce. Nous avons été sur le terrain à fin de vérifier ces hypothèses.

3.5.4 Propriétés du variogramme

La fonction variographique est une fonction de distance. Elle synthétise beaucoup d'information concernant la variation conjointe de la densité à différents lieux.

Portée a : c'est la distance au-delà de laquelle deux observations ne se ressemblent plus du tout en moyenne. Elles ne sont plus liées (covariance nulle) linéairement. À cette distance, la valeur du variogramme correspond à la variance de la variable aléatoire.

Effet de pépite c_0 : elle correspond à la variation possible des erreurs (erreurs de localisation, erreurs d'analyse et de précision analytique).

Palier $\sigma^2 = c_0 + c$: variance de la variable aléatoire $V(Z(s))$ et $C(h)$ une constante.

Cutoff : Distance maximale entre deux points au-delà de laquelle le variogramme n'est plus calculé. Elle correspond ici à la distance maximale qu'on puisse observer entre les couples des lieudits échantillonnés.

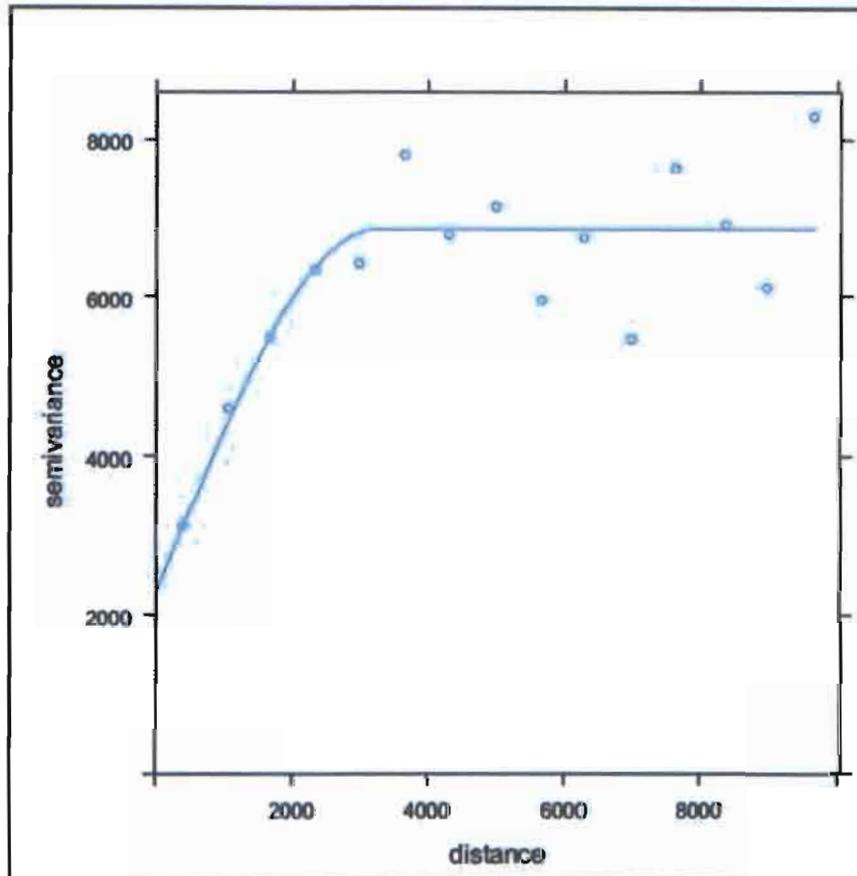


Figure 16 : Le variogramme

3.5.5 Modélisation du variogramme

Une fois le variogramme estimé, il convient de le modéliser par une fonction de référence (modèle exponentiel, sphérique, ...) suivant la disposition des points formant le variogramme empirique. Ce modèle aura alors toutes les propriétés requises pour effectuer un Krigeage sur nos données. C'est une étape clé située entre la description spatiale et la prévision spatiale. Le modèle le plus couramment utilisé (et celui utilisé pour ce stage) est le modèle Sphérique défini de la manière suivante (Casajus 2013) :

$$\gamma(h) = \begin{cases} C_0 + C \left(\frac{3h}{2a} - \frac{1}{2} \frac{h^3}{a^3} \right) & \text{pour } 0 \leq h \leq a \\ C_0 + C & \text{pour } h > a \end{cases}$$

Avec

a : la portée,

C_0 : correspondant à l'effet pépité et

$C_0 + C$: correspondant au palier.

Ce modèle indique une réduction progressive de l'auto-corrélation spatiale (équivalant à une augmentation de semi-variance) jusqu'à une certaine distance au-delà de laquelle l'auto-corrélation est de 0.

4. Le krigage

Cette section expose l'une des techniques de géostatistique d'estimation locale, connue sous le nom de krigage et cokrigage ordinaire. Nous cherchons à estimer la valeur de la densité Z en une zone s_0 quelconque de la commune urbaine à partir des mesures observées $Z(s)$, $i = 1, \dots, n$ (n : nombre de lieux-dits). En d'autres termes, c'est prévoir la densité de population en un site non échantillonné.

Le modèle de base du krigage a la même forme que le modèle de régression classique mais les erreurs sont maintenant supposées dépendantes spatialement. Il s'énonce de la façon suivante :

$$Z^* = \sum_{i=1}^n \lambda_i Z_i,$$

avec

Z^* : Estimateur de Z (la densité de population estimée)

Z_i : Valeur observée au point s_i (v.a.)

λ_i : Les paramètres à estimer.

Parmi les trois (3) types de krigage (ordinaire, simple et universel), seul le krigage ordinaire sera développé ici car il répond aux besoins de notre problématique. En plus de cela, il est également le plus utilisé en pratique car les hypothèses de départ sont moins contraignantes que celle du krigage simple ou universel.

En effet, le krigage ordinaire ne requiert pas la connaissance de l'espérance de la variable régionalisée. Autrement dit, on se place dans l'hypothèse d'une stationnarité intrinsèque. En outre, l'estimateur du krigage ordinaire vérifie les caractéristiques d'un estimateur sans biais :

- Variance minimale
- $\sum_{i=1}^n \lambda_i = 1$
- $E[Z^* - Z] = 0$.

4.1 La variance de krigage

Le krigage est un interpolateur linéaire sans biais. Il prend en compte la géométrie des données, les caractéristiques de la régionalisation et de la variance. Son but est de lisser les données.

En considérant :

$$Z^* = \sum_{i=1}^n \lambda_i Z_i \quad (1)$$

On définit alors l'erreur d'estimation comme suit :

$$e = Z - Z^*$$

La variance de cette erreur correspond à la variance d'estimation et se définit comme suit :

$$V(e) = V(Z - Z^*)$$

$$= E[(Z - Z^*)^2]$$

D'où

$$\sigma_e^2 = V(Z) + V(Z^*) - 2COV(Z, Z^*)$$

En substituant Z^* par son expression en fonction des Z_i dans l'équation (1) et en tenant compte des hypothèses et contraintes définies précédemment sur la robustesse de l'estimateur, on obtient :

$$\sigma_e^2 = V(Z) + \sum_i \sum_j \lambda_i \lambda_j Cov(Z_i, Z_j) - 2 \sum_i Cov(Z_i, Z) \quad (2)$$

Les paramètres λ_i s'obtiennent par la méthode de vraisemblance. Cependant, on a alors un problème de minimisation d'une fonction quadratique sous contrainte d'égalité que l'on résout par la méthode de Lagrange comme suit :

$$\begin{aligned} L(\lambda) &= \sigma_e^2 + 2\mu \left(\sum_{i=1}^n \lambda_i - 1 \right) \\ &= V(Z) + \sum_i \sum_j \lambda_i \lambda_j Cov(Z_i, Z_j) - 2 \sum_i Cov(Z_i, Z) + 2\mu (\sum_{i=1}^n \lambda_i - 1) \end{aligned}$$

Où μ est le multiplicateur de Lagrange. Le minimum est atteint lorsque toutes les dérivées partielles par rapport à chacun des λ_i et par rapport à μ s'annulent (condition du premier ordre). Ceci conduit au système de krigeage ordinaire ou encore l'équation de krigeage :

$$\sum_{j=1}^n \lambda_j Cov(Z_i, Z_j) + \mu = Cov(Z, Z_i) \quad \forall i = 1, \dots, n, \quad (3)$$

avec $\sum_{j=1}^n \lambda_j = 1.$

La variance d'estimation minimale, appelée variance de krigeage, est alors obtenue en substituant les équations de krigeage (3) dans l'expression générale pour la variance d'estimation (2). La variance de krigeage est alors égale à :

$$\sigma_K^2 = V(Z) - \sum_{i=1}^n \lambda_i Cov(Z, Z_i) - \mu \quad (4)$$

Il est intéressant de visualiser le système de krigeage ordinaire et la variance de krigeage ordinaire sous forme matricielle. Soit la configuration suivante :

$$\sum_{j=1}^n \lambda_j Cov(Z_i, Z_j) + \mu = K_0 \lambda_0, \text{ et } Cov(Z, Z_i) = k_0.$$

Ainsi, selon l'équation (3) on peut écrire l'équation matricielle suivante :

$$K_0 \lambda_0 = k_0$$

$$\lambda_0 = k_0 * K_0^{-1}$$

Après que les coefficients soient estimés, selon l'équation (1) on en déduit que :

$$Z^* = \lambda^t Z$$

Où

$$K_0 = \begin{bmatrix} \sigma^2 & Cov(Z_1, Z_2) & \dots & Cov(Z_1, Z_n) & 1 \\ Cov(Z_2, Z_1) & \sigma^2 & \dots & Cov(Z_2, Z_n) & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ Cov(Z_n, Z_1) & Cov(Z_n, Z_2) & \dots & \sigma^2 & 1 \\ 1 & 1 & \dots & 1 & 0 \end{bmatrix}$$

$$k_0 = \begin{bmatrix} Cov(Z_1, Z) \\ Cov(Z_2, Z) \\ \vdots \\ Cov(Z_n, Z) \\ 1 \end{bmatrix}, \quad \lambda_0 = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \\ \mu \end{bmatrix} \quad \text{Et } Z = \begin{bmatrix} Z_1 \\ Z_2 \\ \vdots \\ Z_n \\ 0 \end{bmatrix}$$

Z Correspond à l'échantillon dont on dispose.

La variance d'estimation en krigeage, selon l'équation (4) est définie comme suit :

$$\sigma_K^2 = \sigma^2 - \lambda_0^t * k_0$$

avec

$$\sigma^2 = \text{palier}$$

Note : tous les termes $Cov(Z_i, Z)$ et σ^2 s'obtiennent directement du variogramme comme suit :

$$\gamma(h) = \sigma^2 - C(h) \quad \Rightarrow \quad C(h) = \sigma^2 - \gamma(h)$$

Par exemple :

$$\begin{aligned} Cov(Z_1, Z) &= \text{palier} - \gamma(s_1, s) \\ &= \text{palier} - \gamma(h), \end{aligned}$$

Où h est cependant la distance séparant les sites s_1 et s . C'est pourquoi il est nécessaire de déterminer la matrice de distance entre les différents lieudits échantillonnés.

4.2 Validation croisée

Lors de cette étude, nous avons utilisé la validation croisée dite «leave-one-out». Supposons que l'on ait un échantillon de n observations. Le processus se déroule de la façon suivante :

- On retire l'observation i , notée $Z(s_i)$ de l'échantillon et on lance le modèle avec le nouvel échantillon.
- Puis à l'aide de ce nouveau modèle, on calcule la valeur prédite de l'observation retirée i que l'on note $\hat{Z}(s_i)$.

On réitère ainsi ce processus sur toutes les observations dont on dispose (les 85 lieudits). On obtient alors une série de valeurs prédites que l'on peut comparer aux valeurs observées en calculant les résidus du modèle. Il est ainsi possible d'étudier les erreurs commises sur l'ensemble des données en regardant leur moyenne et leur écart-type. Cependant, l'étude des erreurs n'est pas toujours satisfaisante. En effet, l'amplitude des erreurs commises en un site de données ne dépend pas seulement de la qualité du modèle variographique, mais aussi de l'éloignement par rapport aux autres sites. Ainsi, un site isolé tend à produire une erreur d'estimation élevée. Cet effet peut être corrigé en examinant les erreurs standardisées, c'est-à-dire les erreurs divisées par les écarts-types de krigeage. Celles-ci sont alors ramenées à la même échelle et ne sont plus sensibles à la configuration géométrique des échantillons.

Une fois la validation croisée effectuée, le calcul des critères suivants permet de qualifier le modèle (jugé de son adéquation). (Laffly, s. d.)

- Critère d'adéquation = $\frac{1}{n} \sum_{i=1}^n \left(\frac{\hat{Z}(s_i) - Z(s_i)}{\sigma_i^2} \right)$.

Avec σ_i^2 la variance de krigeage au site i , ainsi le modèle sera «meilleur» et plus performant si ce critère est proche de 0. L'histogramme des erreurs standardisées et la disposition spatiale des erreurs seront examinés pour vérifier si les statistiques précédentes pourraient être biaisées par 1 ou 2 données extrêmes et vérifier si les résidus sont spatialement homogènes.

4.3 La grille d'estimation

Pour obtenir une estimation des densités dans le domaine d'étude, le krigeage est réalisé sur une grille régulière de points qui couvre ce domaine.

Le choix d'une grille d'estimation dépend de l'étude et du domaine d'étude. La taille de la maille de calcul est définie selon le degré de détail souhaité. La grille doit être suffisamment resserrée pour que la carte obtenue représente effectivement les estimations par krigeage. Dans cette étude, le krigeage est effectué sur une grille régulière couvrant toute la superficie du centre urbain (139 km²) sur des pas de 100 mètres.

En effet, les lieudits (ou sites) sont identifiés pour des points de l'espace, mais la densité associée à chaque point correspond à la densité de 100 mètres à la ronde de ce lieudit. De ce fait cette densité sera mesurée pour chaque endroit de la ville.

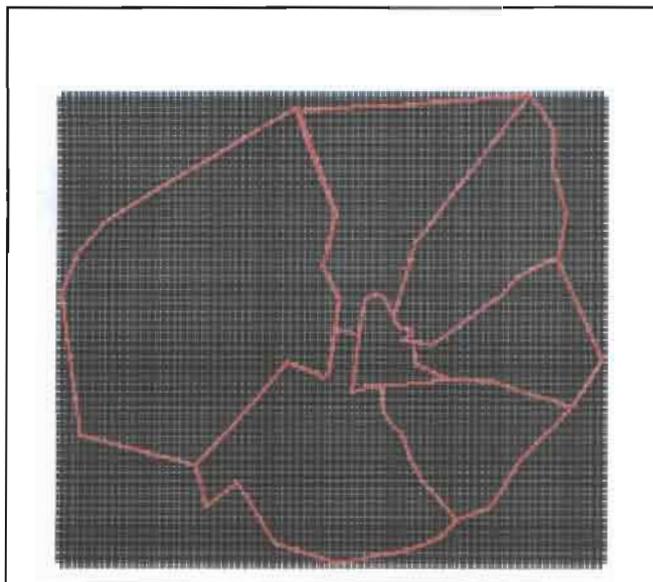
IV. Présentation des résultats

L'objectif de ce chapitre est de présenter les résultats de krigeage obtenus à l'aide du logiciel R et d'en vérifier la validité, ou tout au moins, la cohérence.

1. Résultat de krigeage

Les indicateurs utilisés pour obtenir les différentes densités associées aux lieudits correspondent aux données de la Monographie des Haut-Bassins de l'année 2006. Les scripts du logiciel R utilisés pour obtenir les cartes de krigeage sont tous inspirés de celui présenté en annexe.

- La grille d'interpolation



La zone noire est un ensemble de 37 753 Points générés dans l'espace, séparer chacun de 100 m dont on superpose la trame de la ville

Figure 17 : Processus de conception de la grille

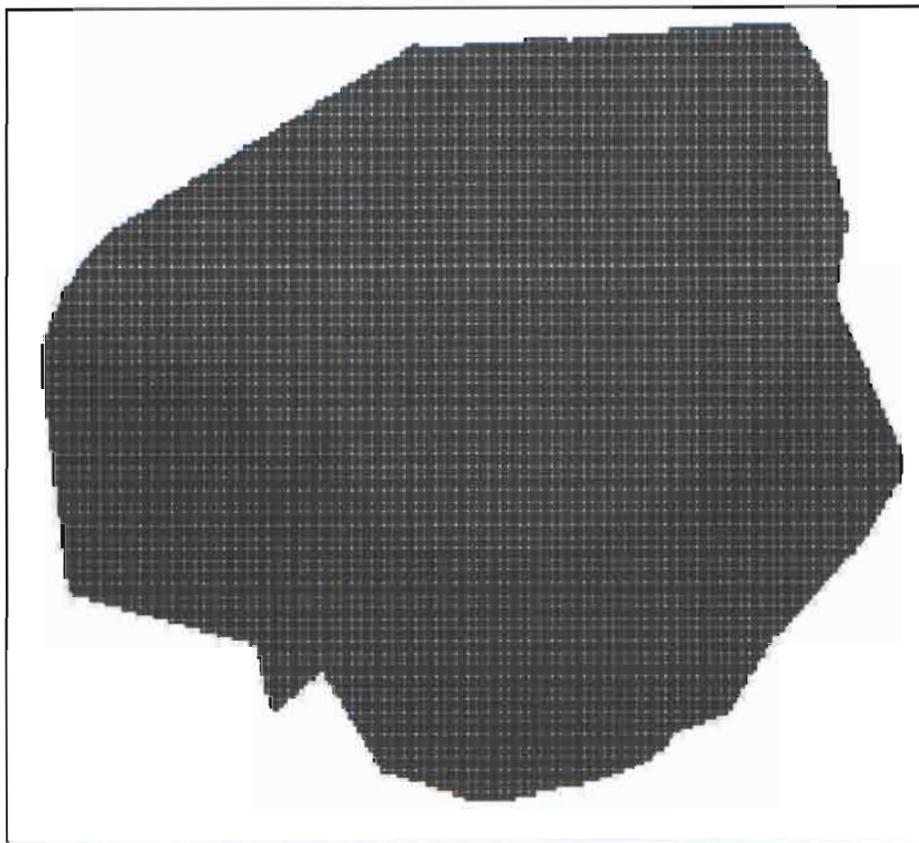


Figure 18 : La grille d'interpolation

- Le résultat du krigeage

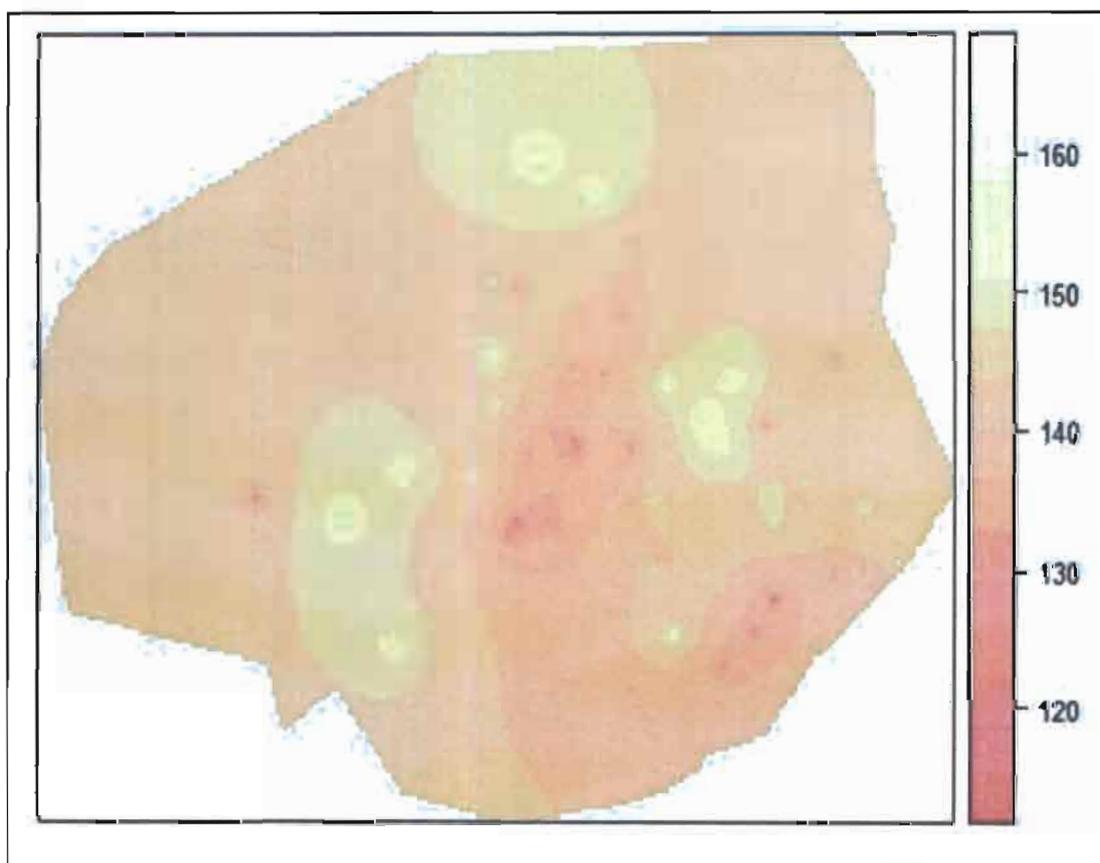


Figure 19 : Estimation de la densité de population sur l'étendue de la ville

Interprétation :

L'observation de la cartographie de krigeage montre que la concentration de la population de Bobo-Dioulasso varie en fonction des zones. Cette densité varie de moins de 120 individus sur 3 hectares à plus de 160 individus par 3 hectares.

En effet, on observe que le centre-ville (du secteur 1 au secteur 7) est le moins concentré. Cette remarque a également attiré notre attention pendant la collecte, plus de 90% des lieudits qui y figurent sont de natures commerciales ou administratives. La zone nord de la ville couvrant le quartier « bobo 2010 » avec les secteurs tels que 10,11, 30 et 23 a une densité allant de 150 par 3 000 km² à tout au plus 160 par 3 000 km². Cette densité est également observable dans les zones couvrant les secteurs 20 à 21 et la zone de « Sarfalao » avec les secteurs comme 14, 15, 16.

- La fonction d'interpolation

Cette fonction permet d'estimer la densité à n'importe quel endroit de la ville. Elle a été programmée à travers le logiciel R, en fonction des processus du système de krigeage. A travers des coordonnées GPS (longitude et latitude) d'un endroit de la ville, cette fonction estime la densité 100 m au tour de ce lieu.

#FONCTION ESTIMATION

```
estimation <- fonction(X0,Y0)
{
#matrice de distance
cords <- base[c("POINT_X","POINT_Y")]
S0 <- c(X0,Y0)
cords <- as.matrix(rbind(S0,cords))
dist <- as.matrix(dist(cords))
#variogramme
pa = 4593.973
po = mean(dist)
ep = 2267.514
dist1 = dist
for(i in 1:length(dist))
{
if(dist1[i] == 0)dist1[i] = 0
if(dist1[i] > 0 & dist1[i] <= po)dist1[i]
= ep + (pa - ep) * (1.5 * (dist1[i]/po) - 0.5 * (dist1[i]/po)^3)
if(dist1[i] > po)dist1[i] = pa
}
#covariance
dist2 = dist1
for(i in 1:length(dist2))
dist2[i] <- (pa - dist2[i])
#calcul des coeficient (K * coef = k0)
#K
K1 <- dist2[-1,-1]
K2 <- cbind(K1,1)
v1 <- rep(1,length(K1[,1]))
```

```

v2 <- -c(v1,0)
K <- -rbind(K2,v2)
#k0
kk <- -dist2[1,]
k0 <- -c(kk[-1],1)
#coef
coef <- -k0% * %solve(K)
coefs <- -coef[-length(coef)]
Z0 <- -sum(coefs% * %pop)
#resulta
print("la densité estimé 100m autour ce point est")
return(Z0)
}
    
```

Note : cette fonction permet l'estimation du nombre d'individus (la densité de la population) situé sur 100 mètres autour de tout lieu identifier de la ville de Bobo-Dioulasso (le centre urbain). En d'autre terme elle donne la densité de la population sur 3 000 km². La fonction estimation est utilisée avec deux paramètres (longitude, latitude).

2. Résultats de la validation croisée

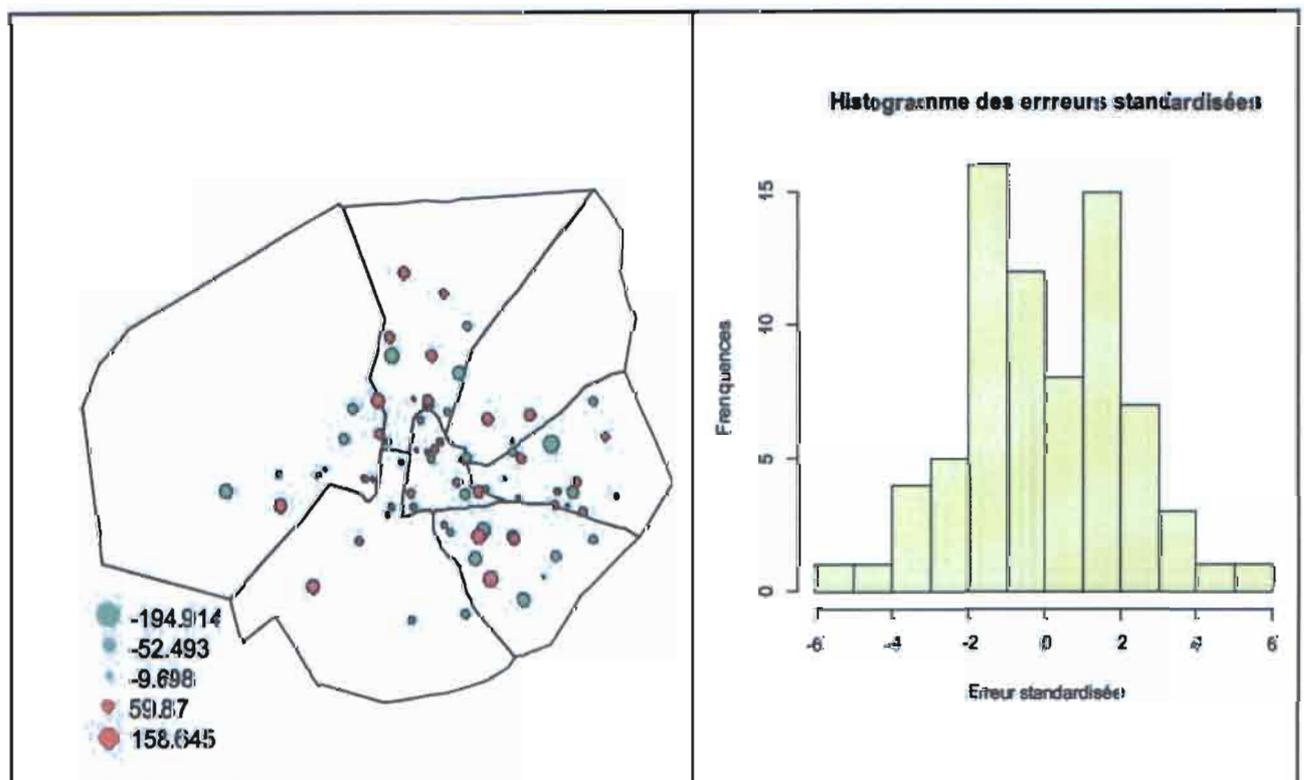


Figure 20 : Cartographie des erreurs d'estimation et histogramme des erreurs standardisées

En effectuant la validation croisée sur le krigeage (cf. figure 20), on note que les erreurs standardisées (les scores) sont majoritairement centrées en zéro. En d'autres termes, nos erreurs suivent une loi normale centrée. De plus, hormis le couple des lieudits « la pharmacie becy » et « le collège évangélique

la charité », on observe que les résidus sont relativement faibles. Pour le site « collège évangélique la charité » qui présente l'erreur la plus élevée s'explique en grande partie par sa situation géographique. En effet ce site est situé à quelque mètre du marigot séparant les secteurs 25 et 32. Cette erreur peut également se rapprocher par d'autres variables auxiliaires, par exemple du fait que le lieu est situé dans une zone accidentée, peuplée de grands bassins et également constitue la zone limitrophe de la partie « lotie » et « non lotie » du secteur 32. Cependant, on admet que ce gradient n'est pas représentatif d'une évolution à échelle régionale mais simplement de la zone dans laquelle elles sont implantées.

De plus le calcul du critère d'adéquation suivant confirme la validité de notre modèle.

$$\text{Critère d'adéquation} = \frac{1}{n} \sum_{i=1}^n \left(\frac{\hat{Z}(s_i) - Z(s_i)}{\sigma_i^2} \right) = -3.202733e - 15$$

3. Cartographie des lieudits avec les itinéraires associés

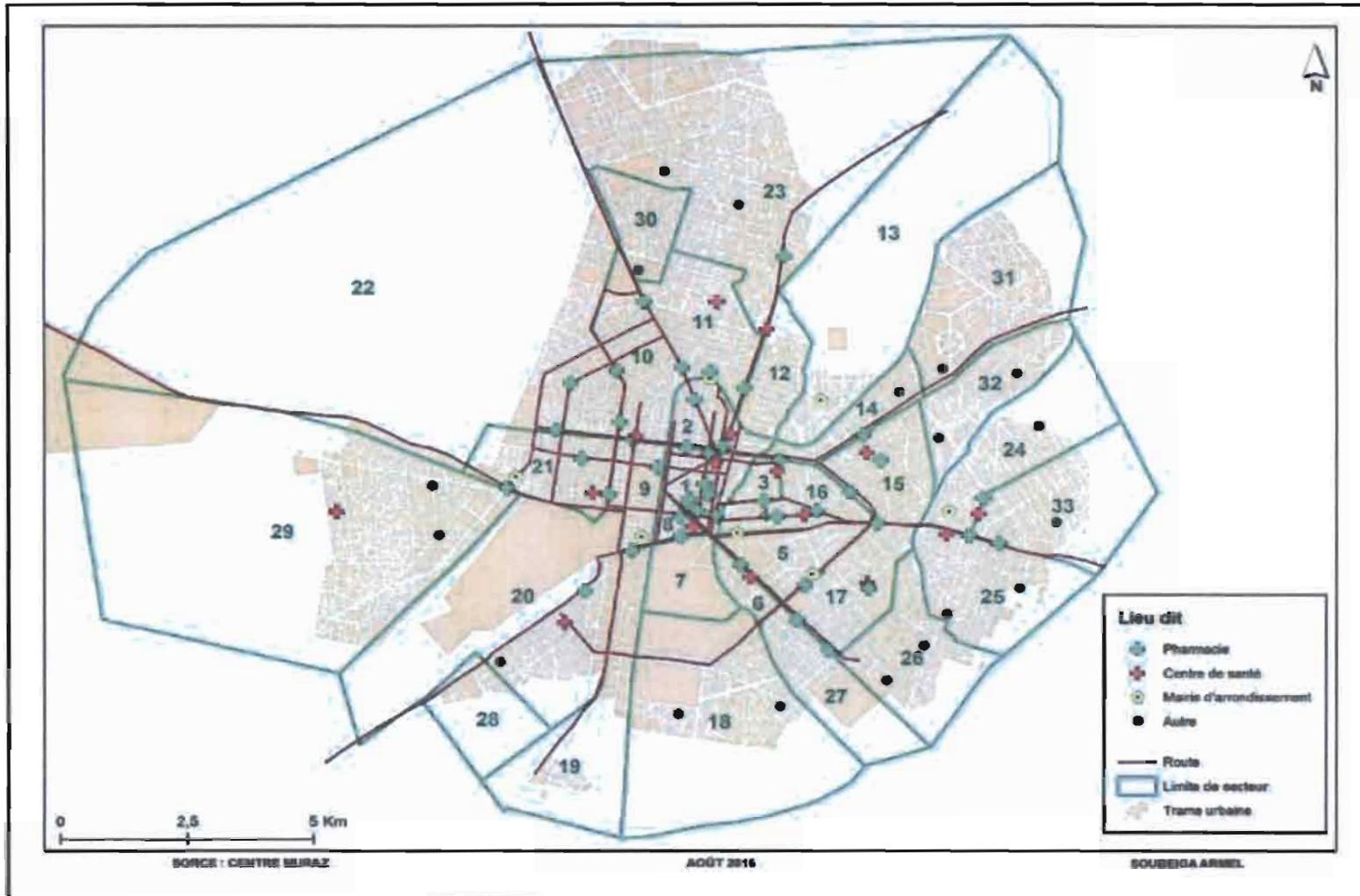


Figure 21 : Cartographie des lieudits avec les itinéraires associés

V. Discussions et conclusion

1. Discussions

1.1 Avantages et limites

A défaut des pourcentages de répartition des ménages en fonction du type d'habitat et des proportions moyennes du nombre d'habitant par ménage à jour, nous avons tenus à utiliser des taux de la monographie de Bobo-Dioulasso de 2006. Cela freiner d'une part la fiabilité des résultats. En plus de cela, nous soulignons quelques limites, en effet, cette étude est principalement inhérentes à la constitution de l'échantillon dont la faible taille joue sur la précision de l'estimateur. La constitution de cet échantillon sur la base de sélection des lieudits ne la rend pas non plus exempt de biais de sélection.

Toutefois, la Base Nationale des Données Topographiques (BNDT) porte un grand intérêt quant aux travaux d'urbanistes et de géostatistiques. Elle fournit des informations sur les habitats et leurs caractéristiques et regroupements spatiales. En effet, des méthodes urbanistes et statistiques peuvent être utilisées pour estimer avec une grande précision le nombre moyen de ménages ou d'individus par type d'habitat dans une agglomération, une zone ou un périmètre donné.

L'approche méthodologique utilisée (krigeage) présente certains avantages, notamment comparée à une approche de simulation géostatistique ou à la modélisation paramétrique ou linéaire en générale. Elle permet d'une part d'éviter de faire des hypothèses fortes sur des distributions d'éventuels résidus et garantit le minimum de variance. Le krigeage réalise l'interpolation spatiale d'une variable régionalisée par calcul de l'espérance mathématique d'une variable aléatoire, utilisant l'interprétation et la modélisation du variogramme expérimental. C'est le meilleur estimateur linéaire non-biaisé. Il permet surtout une présentation visuelle et conviviale des résultats, facilement interprétable par les décideurs.

1.2 Perspectives

Le regard doit se porter davantage sur la méthode que sur les données utilisées. En effet, au sortir d'une telle étude, il pourrait être recommandé d'approfondir cette recherche afin de trouver des solutions idoines validées par des communautés scientifiques. Le présent document présente les résultats préliminaires d'une initiative élaborée à l'intention d'un projet de recherche initié par le Dr. Serge SOMDA, Méthodologiste-Biostatisticien. Ce travail de recherche a fait également l'objet d'une candidature à la conférence internationale de Statistique Appliquée pour le Développement en Afrique (SADA 2016) qui s'est déroulé du 28 novembre au 3 décembre 2016 à Cotonou au Bénin. C'est une conférence de discussions et d'échanges sur les différents aspects de la statistique et de la modélisation stochastique, tant théoriques que pratiques, en relation avec les problèmes du développement en Afrique. Un article scientifique est également en cour de rédaction. Dans le cadre de la collecte électronique de données, nous avons contribué à la mise à jour des données de la plateforme OpenStreetMap (OSM).

Dans nos recherches, moins d'études sont orientées dans ce sens. Cependant, au Canada, des spécialistes de l'urbanisme ont essayé d'aborder la question. Pour ces spécialistes, il ressortait que la méthode géostatistique appliquée à la démographie donne des résultats intéressants dans la mesure où elles permettent d'estimer et de suivre l'évolution d'une population en temps réel. Cette méthode serait

très utile pour tout chercheurs, instituts et tous ceux qui souhaiteraient obtenir des données démographiques actualisées. (Marry 2009)

2. Conclusion

Au terme de l'étude intitulée «Définition d'une base de sondage par modélisation spatiale», nous avons pu mettre à la disposition de l'AMF la première version de la base de données aréolaire de la ville de Bobo-Dioulasso. Cette base lui permettra de réaliser efficacement ces enquêtes postérieures.

D'un point de vue personnel, ce stage m'a permis de connaître de nouvelles méthodes statistiques et géostatistiques, notamment toutes les méthodes liées à l'exploration des modèles. Il m'a également permis de mettre en application mes compétences acquises durant mes trois années d'études post-baccalauréat. Il m'a apporté aussi l'opportunité de découvrir l'environnement dans le domaine de la recherche. J'ai apprécié cette expérience car je l'ai trouvée enrichissante pour mes connaissances personnelles. Le travail que j'ai réalisé m'a permis de me perfectionner dans la programmation sous R. J'ai appris l'utilisation des paquets sous R, l'utilisation de différentes méthodes de modélisations.

Mon maître de stage m'a apporté de nombreuses connaissances et m'a rapidement donné la responsabilité de réaliser mes propres analyses en m'aiguillant au fil du temps. Par son intermédiaire, j'ai pu connaître le rôle et les différentes fonctions du biostatisticien et méthodologie dans un centre de recherche en sciences de la santé. J'ai pu donc appliquer et approfondir mes connaissances acquises lors de ma formation.

Références bibliographiques

• Bibliographie

- [1] Casajus, Nicolas. 2013. « Analyses Spatiales Sous R ». Université du Québec à Rimouski, Département de Biologie, Chimie et Géographie.
- [2] Cressie, Noel. 2015. « Statistics for Spatial Data ».
- [3] Hennequi, Marie. 2010. « Spatialisation des données de modélisation par Krigeage ». Master statistique et Applications, Université de Strasbourg.
- [4] INSD. 2006. « Monographie de la Commune Urbaine de Bobo Dioulasso ». Bobo Dioulasso.
- [5] Laffly, Dominique. s. d. « Initiation à la Statistique Descriptive : la cartographie statistique ». Université de Pau, Laboratoire Société Environnement Territoire.
- [6] Marry, S. 2009. « Densité urbaine et qualité de vie ». En quoi la compréhension des représentations de la densité urbaine est-elle prépondérante dans l'analyse et la conception de formes urbaines propices à une certaine qualité de vie.
- [7] PS-EAU, PDM. 2001. « gestion durable des déchets et de l'assainissement urbain ». Commande du Ministère des Affaires Etrangères, lettre de commande 2001 0024200.

• Webographie

- [1] Centre Muraz : <http://www.centre-muraz.fr>
- [2] Baillargeon S. (2005), Le krigeage : revue de la théorie et application à l'interpolation spatiale de données de précipitations ; http://www.umat.ulaval.ca/theses/S-Baillargeon_05.pdf
- [3] R Development Core Team (2009). R: A language and environment for statistical computing, R. Foundation for Statistical Computing, Vienna, Austria: <http://www.R-project.org>
- [4] Pebesma, E.J. (2004). Multivariable geostatistics in S: the gstat package, *Computers & Geosciences*, 30: 683-691; <http://www.imm.dtu.dk/~erj/gstat/gstat.pdf>
- [5] Pebesma, E.J. (2004). Package 'gstat'. [En ligne] <http://www.r-project.org/doc/packages/gstat/>
- [6] Pebesma, E.J. (1999). Gstat user's manual. [En ligne] <http://www.cnr.it/~erj/gstat/>

Annexes : Code R

Annexes A : Krigeage

Voici un résumé des packages de R permettant d'effectuer une analyse variographique, un krigeage et une validation croisée. Les packages de R utilisés sont : *gstat*, *sp*.

1. Analyse variographique avec R

Dans tout ce que nous faisons avec R, on considère que nous avons des données géographiques. Cependant, lorsque l'on fournit un tableau au logiciel, il faut lui préciser quelles colonnes correspondent aux coordonnées de nos points. Pour cela, on utilise la fonction : `coordinates(...)`

Exemple :

```
> base
```

	Arr_Secteur	Id_lieudit	POINT_X	POINT_Y	Noms	pop
1	5	17	361395.0	1233962	csps de sarfalao	116.7322
2	5	17	361431.6	1233850	pharmacie – remedis	226.5978
3	5	17	361591.7	1235260	pharmacie hereso	130.4654

```
> coordinates(base) <- c("POINT_X", "POINT_Y")
```

```
# ou coordinates(base) <- ~POINT_X + POINT_Y
```

	coordinates	Arr_Secteur	Id_lieudit	Noms	pop
1	(361395,1233962)	5	17	csps de sarfalao	116.7322
2	(361431.6,1233850)	5	17	pharmacie – remedis	226.5978
3	(361591.7,1235260)	5	17	pharmacie hereso	130.4654

1.1 Le variogramme expérimental

Cette étape consiste à calculer et tracer le variogramme estimé à l'aide de nos observations.

La fonction de R permettant de calculer le variogramme expérimental est la suivante :

```
> variogram(formule, data, cutoff...)
```

Exemple :

```
#Variogramme experimental omnidirectionnel
```

```
coordinates(base) <- ~POINT_X + POINT_Y
```

```
pop.vgm <- gstat::variogram(pop~ 1, base, cutoff = 10000)
```

```
> pop.vgm
```

	np	dist	gamma	dir.hor	dir.ver	id
1	35	409.1395	3100.462	0	0	var1
2	99	1062.6616	4592.143	0	0	var1
3	213	1671.3497	5505.615	0	0	var1

```
plot(pop.vgm) (cf. figure 14)
```

1.2 Modélisation du variogramme

Une fois le variogramme expérimental calculé, l'étape suivante consiste à modéliser ce variogramme à l'aide de modèles connus afin de pouvoir lui appliquer le krigeage. Pour cela, on a besoin de deux fonctions du logiciel R :

```
> vgm(psill, model, ...)
```

Cette fonction permet de générer un modèle de variogramme. En tapant la commande `vgm()`, R nous renvoie tous les modèles possibles.

```
> fit.variogram(object, model...)
```

Cette fonction permet d'ajuster le variogramme expérimental par un modèle variographique.

Exemple :

```
#Modélisation du variogramme
pop.fit <- gstat::fit.variogram(pop.vgm, vgm("Sph"))
> pop.fit
  model psill range
1 Nug 2267.514 0.000
2 Sph 4593.973 3231.333
plot(pop.vgm, pop.fit)
(cf. figure 16)
```

2. Le Krigeage avec R

Afin de kriger la variable régionalisée avec R, il faut tout d'abord créer une grille sur laquelle on pourra effectuer le krigeage.

Exemple :

```
#Importation de la carte ou domaine de krigeage
aus <- readShapePoly("E:\\base\\commune_bobo_utm.shp")
# Coordonnees de letendue spatiale de bobo
grx <- seq(bbox(aus)[1,1], bbox(aus)[1,2], by = 100)
gry <- seq(bbox(aus)[2,1], bbox(aus)[2,2], by = 100)
x <- rep(grx, length(gry))
y <- rep(gry, length(grx))
y <- sort(y, decreasing = F)
pt <- data.frame(x = x, y = y)
pts <- SpatialPoints(coordinates(pt))
# Definition du systeme de coordonnees
proj4string(pts) <- proj4string(aus)
#Definition d'une table d'attributs
mat <- data.frame(ID = paste("Random", 1:37758, sep = ""))
#Rajout à la table d'attributs
pts <- SpatialPointsDataFrame(pts, data = mat)
# Jointure spatiale
join <- over(pts, aus)
#Rajout des identifiants
```

```

join <- data.frame(ID = pts@data[, "ID"], join)
# Identification des lignes avec NA
pos <- which(is.na(join[, "AREA"]))
# Suppression de ces lignes
join2 <- join[-pos, ]
# Reinitialisation des noms de lignes
rownames(join2) <- NULL
# convertir ce nouveau data.frame
join2 <- SpatialPointsDataFrame(coords = pts@coords [-pos, ], data = join2)
# conversion de la grille en élément spatiale
join2 <- as(join2, "SpatialPoints")
gridded(join2) <- TRUE
par(mar = c(0, 0, 0, 0))
plot(join2)
(cf. figure 17 et 18)

```

Une fois la grille créée, on peut effectuer le krigeage sur nos données à l'aide de la commande suivante:

```

## Krigeage interpolation des la densité sur tout le domaine d'étude
g <- gstat(id = "pop", formula = pop ~ 1, data = base, nmax = 1, set = list(idp = .5))
pop.krige <- predict(g, model = pop.fit, newdata = join2)
pop.krige
# plotage : On représente la carte krigée
par(mar = c(0, 0, 0, 0))
spplot(pop.krige, "pop.pred", col.regions = heat.colors(50))
(cf. figure 19)

```

3. La validation croisée avec R

Le package *gstat* permet de faire une validation croisée automatique pour la méthode du krigeage. La fonction utilisée est la suivante :

```
> krige.cv(formula, data ...)
```

Exemple :

```

# histogramme des scores (les erreurs standardisées)
hist(v$zscore, col = "yellow", main = "Histogramme des erreurs standardisées", xlab = "Erreur standardisée", ylab = "Frequencies")
# spatialisation des erreur
Ville <- readShapePoly("E:\\BNDT\\CORRECT_BOBO\\Commune 2011\\bobo_commune_limite.shp")
out <- bubble(v, "residual", do.sqrt = T, col = c(3, 2), main = "validation")
plot(Ville)
symbols(out$panel.args[[1]], circles = out$panel.args.common$cex * 100,
bg = out$panel.args.common$col, inches = F, add = TRUE)
legend(347504.9, 1248999, legend = out$legend$right$args$key$text[[1]],
pt.cex = out$legend$right$args$key$points$cex, pch = 20,
col = out$legend$right$args$key$points$col, bty = "n", y.intersp = 1) (cf. figure 20)

```

#Critère d'adéquation

$$q < -\text{sum} \left(\frac{v\$residual}{v\$var1.var} \right)$$

Annexe B : Cartographie avec R

Les packages de R utilisés sont : *gstat*, *sp*, *cartography*, *stats*, *geoR*, *geostat*

Exemple : cartographie de la zone d'étude

#departement

```
Burkina <- readShapePoly("E:\\BNDT\\Adm\\bf_reg.shp")
```

```
Bobo <- readShapePoly("E:\\BNDT\\Adm\\Bf_prov.shp")
```

```
comune <- readShapePoly("E:\\BNDT\\Adm\\Bf_dept.shp")
```

```
plot(Burkina)
```

```
plot(Burkina[Burkina$REGION == "HAUTS - BASSINS",], col = "Gray", add = TRUE)
```

```
plot(Bobo[Bobo$PROV == "HOUET",], col = "red", add = TRUE)
```

```
plot(comune[comune$NOMDEP == "BOBO - DIOULASSO",], col = "white", add = TRUE)
```

#legende

```
locator(n = 1) #permet de cliquer endroit à placer la legende et recuperer les coordonnées
legend(729085.5,1167706, legend
```

```
  = c("Region des Hauts Bassins", "Province du Houet", "Commune"), col
```

```
  = c("Gray", "red", "white"), cex = 0.9, bty = "n", pch = 15)
```

#La commune et la ville

```
comune <- readShapePoly("E:\\BNDT\\CORRECT_BOBO\\Commune 2011\\
  \departement de bobo.shp")
```

```
plot(comune, border = "black")
```

```
Ville <- readShapePoly("E:\\BNDT\\CORRECT_BOBO\\Commune 2011\\
  \bobo_commune_limite.shp")
```

```
proj4string(Ville) <- proj4string(comune)
```

```
plot(Ville, col = "gray", add = TRUE)
```

```
border1 <- readShapeLines("E:\\BNDT\\CORRECT_BOBO\\Commune 2011\\
  \amorce_houet.shp")
```

```
plot(border1, add = TRUE)
```

```
border2 <- readShapeLines("E:\\BNDT\\CORRECT_BOBO\\Commune 2011\\axe.shp")
```

```
plot(border2, add = TRUE)
```

```
layoutLayer(title = NA, author = "Armel Soubeiga",
```

```
sources = NULL, frame = TRUE, col = NA, scale = NULL, coltitle = black, north
  = TRUE)
```

#La ville

```
Ville <- carte1 <- readShapePoly("E:\\base\\commune_bobo.shp")
```

```
plot(Ville)
```

Ajout des titres, légende, sources, etc.

```
layoutLayer(title = "Centre urbain/Ville de Bobo – Dioulasso", author  
            = "Armel Soubeiga", sources = " aout 2016", frame = TRUE, col  
            = NA, scale = NULL, coltitle = "black", north = TRUE)
```

#Affichage du numero de chaque arrondissement au milieu de celui – ci

```
point < –as.data.frame(coordinates(Ville))
```

```
colnames(point) < – c("X", "Y") #Extraction des coordonné des arrondissements
```

```
point < – merge(point, Ville, by.x = "row.names", by.y = "row.names", all.x = TRUE)
```

```
text(point[, c("X", "Y")], c("2ér Arrondissement", "7ème Arrondissement", "3ème Arrondissement",  
"1ème Arrondissement", "6ème Arrondissement", "5ème Arrondissement", "4ème Arrondissement"),  
     = .5)
```