



**UNIVERSITÉ
D'ORLÉANS**



**ÉCOLE DOCTORALE SCIENCES DE L'HOMME ET DE LA SOCIÉTÉ
LABORATOIRE CEDETE**

THÈSE présentée par :

Yélézouomin Stéphane Corentin SOME

soutenue le **13 juillet 2010**

pour obtenir le grade de : **Docteur de l'université d'Orléans**

Discipline/ Spécialité : Géographie – Aménagement - Environnement

**MODELISATION DE LA DISTRIBUTION SPATIALE
DES FORMES MOLECULAIRES M ET S D'*Anopheles gambiae sensu stricto*
AU BURKINA FASO AVEC LES SIG ET L'ANALYSE SPATIALE**

THÈSE dirigée par :

Monsieur Jean-Marie FOTSING, Professeur, Université d'Orléans

RAPPORTEURS :

Monsieur Jean-Pierre LACAUX, Physicien des Observatoires, CNES-MEDIAS

Monsieur François de Charles OUEDRAOGO, Maître de Conférences (CAMES), Université de Ouagadougou

JURY :

Monsieur Jean-Marie FOTSING, Professeur, Université d'Orléans

Monsieur Jean-Pierre LACAUX, Physicien des Observatoires, CNES-MEDIAS

Monsieur François de Charles OUEDRAOGO, Maître de conférences CAMES, Rapporteur

Monsieur Carlo COSTANTINI, Directeur de recherche, IRD

Monsieur Jean-Marc ZANINETTI, Professeur, Université d'Orléans

SOMMAIRE

LISTE DES ABREVIATIONS	3
DEDICACE	5
REMERCIEMENTS	6
INTRODUCTION GENERALE	9
PREMIERE PARTIE :	15
CADRAGE THEMATIQUE, DEMARCHE SCIENTIFIQUE ET METHODE	15
CHAPITRE I : ETAT DE L'ART, DEFINITION DU CHAMP CONCEPTUEL ET POSITIONNEMENT SCIENTIFIQUE	17
CHAPITRE II : BIO-ECOLOGIE D' <i>An. gambiae</i> AU BURKINA FASO : UNE RELATION A FORMALISER	40
CHAPITRE III : APPROCHE METHODOLOGIQUE DE L'ETUDE	68
DEUXIEME PARTIE	96
STRUCTURATION DES DONNEES, ELABORATION ET EVALUATION DES MODELES	96
CHAPITRE IV : ORGANISATION ET STRUCTURATION DE LA BASE DE DONNEES	98
CHAPITRE V : DISTRIBUTION SPATIALE DES MEMBRES DU COMPLEXE <i>An. gambiae</i> : STRUCTURES GLOBALES ET CONFIGURATIONS LOCALES	123
CHAPITRE VI : MODELISATION DE LA DISTRIBUTION SPATIALE	156
CONCLUSION GENERALE	181
ANNEXES	191
ANNEXE 1 : METHODES DE COLLECTE DES MOUSTIQUES SUR LE TERRAIN	192
ANNEXE 2 : IDENTIFICATION DES FORMES MOLECULAIRES	194
LISTE DES FIGURES	199
LISTE DES TABLEAUX	202

LISTE DES ABREVIATIONS

2iE	Institut International d'Ingénierie de l'Eau et de l'Environnement
ACP	Analyse en Composantes Principales
ADN	Acide Désoxyribonucléique
ANCOVA	Analyse de Covariance
ARN	Acide Ribonucléique
AVHRR	Advanced Very High Resolution Radiometer
BDOT	Base de Données sur l'Occupation des terres
BNDT	Banque National de Données Topographique
CEDETE	Centre d'Etudes sur le Développement des Territoires et Environnement
CIT	Convergence Intertropicale
CNRFP	Centre National de Recherche et de Formation sur le Paludisme
CRDI	Centre International de Recherche sur le Développement
CTA	Combinaisons Thérapeutiques à base d'Artémisinine
DDT	Dichlorodiphényltrichloroéthane
ESRI	Environmental Systems Research Institute Inc
ETM+,	Enhanced Thematic Mapper Plus
ETP	Evapotranspiration
FAO	Food Alimentary Organization
FIT	Front intertropical
GIS	Geographic Information System
GPS	Global Positioning System
IBC	Interventions à Base Communautaire
IDW	Inverse Distance Weigthing
INSD	Institut National de Statistique et de la Démographie
IRD	Institut de Recherche pour le Développement
ISO	International Organization for Standardization
LISA	Local Indices for Spatial Association
LIV	Lutte Intégrée contre les Vecteurs
LocClim	Local Climate
LPI	Local polynomial interpolation
MCD	Modèle Conceptuel de Données
MCT	Modèle Conceptuel de Traitement

MTI	Moustiquaires Traitées aux Insecticides
NDVI	Normalized Vegetation Index
NIH	National Institute of Health
NOAA	National Oceanic and Atmospheric Administration
OCEAC	Organisation de Coordination pour la lutte contre les Endémies en Afrique Centrale
OMD	Objectifs du Millénaire pour le Développement
OMS	Organisation Mondiale de la Santé
PCIME	Prise en Charge Intégrée des Maladies de l'Enfant
PCR	Polymerized Chain Reaction
PNLP	Programme National de Lutte contre le Paludisme
RBF	Radial Basic Fonction
SIG	Système d'Information Géographique
SNIS	Système National d'Information Sanitaire
SPOT	Satellite pour l'Observation de la Terre
TM	Thematic Mapper
TPI	Traitement Préventif Intermittent
USGS	United States Geological Service
UTM	Universal Transverse Mercator
WHO	World Health Organization

DEDICACE

*A Vous qui êtes sensibles à la douleur d'autrui,
A Vous qui peinez devant l'injustice,
A Vous qui prenez des risques pour défendre le faible, l'opprimé,
Je dédis ce travail.*

*A vous qui souffrez de l'injustice, de l'indifférence et qui appelez de tous votre être au secours,
Trouvez en ce document les raisons de garder espoir et de croire qu'après la nuit froide et obscure,
le soleil s'élèvera dans votre existence et en éclairera tous les recoins pour vous donner de
savourer les délices d'une vie belle et fraternelle sur la terre des Hommes.*

*Au Père Tout Puissant pour sa grâce et sa miséricorde infinies,
Il m'a accordé de voir la manifestation de la force de son Bras.
J'ai mis ma confiance en Lui, je serai toujours dans la gloire de son Nom.*

REMERCIEMENTS

Nous arrivons tous dans ce monde dans les mains des Hommes et nous dépendons des Hommes pendant les premières années de notre vie. Quand nous quittons ce monde, c'est encore les Hommes qui s'occupent de la gestion de notre dépouille mortelle. Nul succès n'est donc possible sans la contribution des Hommes et nul mérite ne peut s'envisager en dehors des opportunités créées par les Hommes. Nous sommes tous, puissants comme faibles, redevables des autres dans nos succès et nous devons faire preuve de gratitude.

Mais qui remercier et qui laisser ? Dans tous les cas, il est impossible de citer tout le monde. Et pour ne frustrer personne, j'en étais arrivé à me questionner sur la nécessité d'un mot de remerciement.

Mais une sagesse africaine dit : Les yeux voient mais ils ne peuvent raconter. La bouche ne voit pas mais elle raconte ce qu'elle ne peut voir. Telle est la réalité de notre existence terrestre. Les honneurs ne reviennent pas toujours à celui à qu'ils sont dus.

J'ai alors pensé que le pire, c'est de ne rien faire. Ne rien faire, c'est refuser de reconnaître les contributions diverses qui m'ont permis d'arriver à ce résultat. Ce serait injuste car de nombreuses personnes se sont données tant de peine pour que ce travail voie le jour. Pour que jamais elles ne regrettent d'avoir pris de l'initiative, d'avoir compatié, d'avoir espéré pour moi et avec moi, d'avoir prié et souffert, je me suis décidé à leur adresser quelques mots de remerciement.

Les pensées et les paroles sont créatrices de force, d'engagement. Les pensées positives engendrent la force, le courage, l'idéal d'une humanité plus harmonieuse, plus juste, plus équitable. Ce que je voudrais, c'est que ces mots de remerciement créent en toute personne qui les lira, un dragon puissant de sagesse et de paix ; un esprit si puissant que tout leur être matériel et subtil transpire d'amour et de charité. Alors, à vous tous qui avez tant donné pour moi, l'ami d'un jour, le fils, le collègue, le père, le mari, le nouvel ami et même le parfait infortuné inconnu, j'adresse ma profonde gratitude. Je pris la Toute Puissance Divine, la Création elle-même, Celle de qui nous provenons tous, mauvaises comme bonnes gens, de vous inspirer toujours cet élan transcendant vers l'autre pour un avenir de paix et de grâce.

Les grandes choses commencent par la petite idée qui jaillit de la pensée pure de tout être. Alors, que nul estime que ce qu'il a fait est peu ou pas assez, du moment où il a eu cette petite idée de faire quelque chose de bien. La montagne est constituée de roches, la roche, de minéraux, le minéral de molécules, la molécule d'atomes et l'atome de cette énergie invisible mais combien puissante qui donne l'essence de toute chose. Vous croyez sans doute que vous vous êtes sentis

impuissant et vous avez souffert de votre impotence en vous disant : voilà ce que je ferai pour l'aider si j'avais été décideur. Hé bien ! L'univers n'est pas resté sourd. Il a entendu les plaintes qui montent de votre cœur meurtri et à décider que votre compassion porte le fruit dont voilà la matérialisation. Grâce à vous le miracle s'est réalisé !

Ce mémoire ne peut donc se réduire à son contenu scientifique. Il est porteur de force, de persévérance, de joie, d'espérance. Il est également un intégral d'énergie, de combat, de foi en l'homme et en l'amour. Il est donc une propriété de l'universel de la persévérance, de l'amour infini, de l'existence créatrice d'une humanité grandissante et en quête d'harmonie.

Amour et sagesse, lumière et vérité en vous et autour de vous, pour l'avènement d'un monde meilleur sur la terre des hommes.

Je voudrais citer des noms pour qu'ils incarnent les interactions multiples que j'ai eues avec les êtres de tout bord. Chacun ressentira ma gratitude selon ce qu'il aura fait dans sa conscience. Un nom dans cette liste ne fait pas de vous un homme de bien ou de mal, car l'ultime jugement vous sera fait par votre conscience qui est expression de la présence de l'humanité en vous. Si vous n'avez pas trouvé votre nom dans cette liste alors que vous pensez avoir fait quelque chose de bien, ne m'en voulez pas car ce document vous appartient. Vous avez vu le nom d'une personne indigne à vos yeux, ne vous en faites pas, c'est pour que les sentiments de joie que vous dégagez le purifie afin qu'il devienne lui aussi un jour, pour un autre commun des mortels, un artisan de la paix universelle. Sans préséance, merci à vous :

- Moussa Guelbéogo, Sagnon N'Falé, N Ismael Bassolé, du CNRFP,
- Dapola Evariste Constant Da, Tanga Pierre Zoungrana, P. Honoré Somé et Stanislas Ouaro, de l'Université de Ouagadougou,
- Gérald Guillaumet, Youssoufi Touré, Anne Lavigne, Gabriel Bergounioux, Guillaume Géroir, Jean-Marc Zaninetti, Pascal Porée, Catherine Alhéonor et Sylvie Bargain de l'Université d'Orléans,
- Jean-Marie Fotsing, mon directeur de thèse
- Jean-Pierre Lacaux du CNES-Media et François de Charles Ouédraogo de l'Université de Ouagadougou, les rapporteurs, Carlo Costantini et Jean-Marc Zaninetti les examinateurs
- Hélène Gurgel Da Costa, Maureen Douka, Joachim Etouna, Gervais Wafo, Raphaël Pin-Diop tous doctorants ou ancien doctorant en accueil au Centre IRD d'Orléans

- Carlo Costantini, Didier Fontenille, Frédéric Simard, Frédéric Huynh, Evelyne Poncet, Jocelyne Petit, Christine Chauviat, Christophe Mahuzier, Jean-Marie Fotsing, Kenji Ose de l'IRD
- Paul Gnies, Joseph Wéthé, Amadou Hama Maïga, Sina Thiam, Salimata Diakité/Coulibaly, Malick Zoromé, Lydie Yiougo, Yacouba Konaté, Yonkeu Samuel, Salifou Déné, Martial Amoussou, Diane Somé, Feu Lamine Traoré, mes étudiants de M2 MGIEC et GIRE, M1, L3 du 2iE, le corps des enseignants et chercheurs du 2iE
- Kadidiatou Somé/Ouattara, mon épouse, Biauviel Somé notre fils, Félicien et Bernadette Somé mon père et ma mère,
- Feu Jean-François Somé, Feu Jean-François Dabiré, Pascal Somé, Urbain Dabiré, Thomson et Alice Somda, Serges et Zenabou Somé, Ghislain Dabiré, Abdon et Gisèle Hien, Zan Traoré, Mamou Bayala, Désiré Palé, Bamassi Fayama, Tchani, François et Marie-Rose Maurin, Soiré Konaté, Mathieu Kientga, Arlette Dabiré, Lydie Wafo, Issiaka Sombié, Dramane et Elsa Sanon, Fabrice et véronique Béogo, Evariste Ouédraogo, Estelle et Douada Barro, Boubacar Diaby, Alkassim Bah, Labasse Koumaré, Camara Papou et Maude et Gustave Yaméogo, Arsène et Audrey Dabiré et tous les autres parents et amis en France et au Burkina dont les noms n'ont pu être cités.

INTRODUCTION GENERALE

Contexte et justification

« L'environnement agit quotidiennement sur notre santé par le biais du milieu qui nous entoure : eau, air ou sol. En cas de détérioration brutale de l'environnement, comme dans les camps de réfugiés par exemple, les impacts sanitaires sont d'ailleurs considérables »¹, affirme la Fondation Nicolas Hulot pour la Nature et l'Homme. Pour le Centre International de Recherche sur le Développement (CRDI), « Santé des écosystèmes égale santé humaine ».

Cette mise en miroir de la santé et de l'environnement n'est pas nouvelle. Le fondateur de la médecine moderne, Hippocrate, (v. 460 - 377 av. J.-C.) ne disait-il pas que « pour approfondir la médecine, il faut considérer d'abord les saisons, connaître la qualité des eaux, des vents, étudier les divers états du sol et le genre de vie des habitants² ». La découverte par Louis Pasteur en 1865, du microbe et le développement de la microbiologie ont conduit à la stérilisation, à l'hygiène du milieu et à l'assainissement dont l'impact sur l'espérance de vie de l'homme est indéniable. Dans les années 1980, l'émergence de la médecine environnementale et de l'éco-toxicologie s'inscrit dans cette dynamique de prise en compte de l'environnement dans les questions de santé et du bien-être des hommes.

La communauté internationale a reconnu ce lien inextricable entre environnement et santé et l'a déjà exprimé de façon très formelle à plusieurs reprises. C'est le cas aux points 3 et 4 du préambule et aux principes 1 et 7 de la déclaration de la Conférence des Nations Unies sur l'Environnement, tenue à Stockholm du 5 au 16 juin 1972. Cette reconnaissance est réaffirmée à Rio de Janeiro dans la déclaration de Conférence des Nations Unies sur l'Environnement et le Développement du 3 au 14 juin 1992.

D'après le rapport O.M.S., (2006), « plus de 13 millions de décès sont dus annuellement à des causes environnementales évitables. Près du tiers des décès et des maladies qui surviennent dans les régions les moins développées sont provoqués par l'environnement. Plus de 40% des décès des suites du paludisme et environ 94% des décès provoqués par des maladies diarrhéiques, deux des principaux tueurs d'enfants au niveau mondial, pourraient être évités par une meilleure gestion de

¹ http://www.fnh.org/francais/doc/en_ligne/sante/dossier1_intro.htm

² D'après "Airs, eaux, lieux", in Hippocrate - De l'art médical, Le livre de poche classique
http://www.fnh.org/francais/doc/en_ligne/sante/dossier1_art1.htm

l'environnement [...] D'une façon ou d'une autre, l'environnement affecte de manière significative plus de 80% des principales maladies. »³

D'après ce rapport, « les quatre principales maladies influencées par un environnement de mauvaise qualité sont la diarrhée, les infections des voies respiratoires inférieures, les différentes formes de lésions involontaires et le paludisme »⁴.

Le paludisme constitue un problème majeur de santé publique dans les régions tropicales. Selon l'Organisation Mondiale de la Santé (OMS), 350 à 500 millions de cas de paludisme sont enregistrés chaque année dans le monde, avec plus d'un million de décès. C'est une affection d'origine parasitaire due à un sporozoaire hématozoaire du genre *Plasmodium*. Elle est transmise à l'homme lors du repas de sang d'un anophèle femelle infecté. C'est donc une maladie à transmission vectorielle. Les couches de population les plus touchées sont les enfants de moins de cinq ans. Le paludisme est la cause de 18% des décès d'enfants de moins de cinq ans. L'Afrique est de loin le continent le plus concerné avec plus de 80% de décès. Environ 25 à 35% des consultations cliniques, 20 à 45% des hospitalisations et 15 à 35% des décès en milieu hospitalier sont dus à cette affection en Afrique subsaharienne WHO, (2005). Au Burkina Faso, près de 41,1% des décès survenus en milieu hospitalier en 2001 étaient imputables à cette affection Sirima et *al.*, (2003).

En 1996, une attention particulière est accordée aux maladies liées à l'environnement. Des initiatives s'intéressant au paludisme apparaissent, mais à travers des programmes larges concernant les maladies en rapport avec l'environnement. Le programme Ecosystèmes et Santé humaine (Écosanté) du Centre International de Recherche sur le Développement (CRDI) en est un exemple. En 1997, la lutte contre le paludisme fait l'objet d'un regain d'intérêt en raison de la recrudescence de la maladie dans les pays en voie de développement et de son lien direct avec l'environnement. Sous l'incitation du Fogarty International Center (FIC), la Multilateral Initiative on Malaria (MIM) est créée en 1997. En 1998, Roll Back Malaria naît. Le paludisme et le SIDA sont les seules maladies citées parmi les Objectifs du Millénaire pour le Développement (OMD).

Les programmes de lutte antipaludique reposent sur la lutte anti-plasmodiale combinée à celle antivectorielle. La lutte anti-vectorielle a d'ailleurs constitué la base des premiers succès des programmes d'éradication du paludisme dans certaines contrées, particulièrement en Europe. Elle vise la réduction du taux de survie de la population vectorielle et du contact homme-vecteur. Les

³ <http://www.who.int/mediacentre/news/releases/2006/pr32/fr/index.html>

⁴ <http://www.aidh.org/sante/sante-environn.htm>

premiers succès de cette bataille dans les années 1950, avec l'utilisation des insecticides de synthèse comme le Dichlorodiphényltrichloroéthane (DDT) et la dieldrine, ont suscité un espoir d'éradication du paludisme.

Cependant, l'objectif d'éradication n'a pas été atteint, parce que les vecteurs ont vite développé une résistance aux insecticides. Ce qui a contribué à la résurgence de la maladie partout où elle avait connu un recul Greenwood et Mutabingwa, (2002), Trape *et al.*, (1996). Rien d'étonnant que de nouvelles initiatives soient prises pour limiter le contact homme-vecteur. Actuellement la lutte antivectorielle s'attache aussi à réduire l'accès des vecteurs à l'homme par l'utilisation de matériaux imprégnés d'insecticides. Toutefois, cette option pose la question des conséquences de la réduction de la transmission du *Plasmodium* sur l'immunité des enfants en bas âge dans les zones hyperendémiques D'Alessandro *et al.*, (1995) ; Snow et Marsh, 1995 ; Binka *et al.*, (1996) ; Nebie, (2002). Aussi, la lutte génétique s'inscrit-elle parmi les initiatives développées pour le contrôle de la population vectorielle. Elle vise le remplacement de la population de vecteurs sauvages par des moustiques réfractaires au *Plasmodium* Collin et Besansky, (1994).

Il est largement admis aujourd'hui qu'une lutte antivectorielle ne sera efficace à long terme qu'à la condition d'une connaissance approfondie des cycles de transmission, de la biologie et de l'écologie des vecteurs. La non prise en compte de l'hétérogénéité comportementale des vecteurs impliqués dans le système de transmission, c'est-à-dire, la méconnaissance de la biologie et de l'écologie des vecteurs, a été sans doute à la base des échecs enregistrés dans l'éradication du paludisme en Afrique Chauvet *et al.*, (1969). L'état actuel des connaissances sur l'écologie des anophèles reste insuffisant pour permettre la mise en place de mesures efficaces de contrôle vectoriel. Par exemple, les causes de l'hétérogénéité spatiale observées dans l'abondance et la distribution des membres du complexe *Anopheles gambiae* sont encore méconnues. Il en est de même pour les mécanismes qui interviennent dans la régulation des populations.

Il est devenu ainsi nécessaire, dans la perspective du contrôle de la population des vecteurs, de prendre en compte cette diversité spécifique et de répondre efficacement à un certain nombre de questions telles que :

- quelle est la compétence des nouvelles espèces en matière de transmission du paludisme ?
- cette spéciation en cours est-elle doublée d'une diversification au niveau éthologique, écologique ou biologique ?
- un processus d'endémisme est-il en cours ?

Compte tenu des effets dévastateurs de cette maladie sur la santé de la population mondiale, de nombreux projets ont été initiés. Ils visent une meilleure connaissance de la biologie, de l'écologie et de l'éthologie de cette espèce, voire de ce complexe d'espèces, pour plus d'efficacité dans le contrôle de la population vectorielle. Ainsi, un projet de NIH⁵ « *Ecological genomic of An. gambiae* », qui porte à la fois sur le Burkina Faso et le Cameroun, est initié conjointement par l'Université Notre Dame aux Etats-Unis, le CNRFP⁶, l'OCEAC⁷ et l'IRD⁸. Ses objectifs sont :

- étudier les mécanismes génétiques impliqués dans l'adaptation du moustique à son environnement ;
- décrire le biotope des deux « formes moléculaires » M et S d'*Anopheles gambiae* au Cameroun et au Burkina Faso ;
- étudier et identifier les paramètres du milieu favorisant la présence de l'une ou l'autre de ces deux formes.

Pour atteindre ces objectifs, le projet « *Ecological genomic of An. gambiae* » a mis en place un protocole de recherche à plusieurs niveaux dont l'aire d'application la plus détaillée (le niveau micro géographique) doit coïncider avec une zone de sympatrie, c'est-à-dire une zone où cohabitent, à proportion presque égale, *An. gambiae* forme moléculaire M et *An. gambiae* forme moléculaire S.

Sujet

Le projet « *Ecological genomic of An. gambiae* » dont dérive ce travail d'étude et de recherche, vise donc à résoudre un problème de santé dans lequel intervient l'environnement en tant que cadre de vie du vecteur. Cette problématique soulève un aspect important, celui de la distribution spatiale des formes moléculaires M et S d'*Anopheles gambiae*, d'où le titre que porte cette étude : **« modélisation de la distribution spatiale des formes moléculaires M et S d'*Anopheles gambiae* sensu stricto au Burkina Faso par les SIG et l'analyse spatiale ».**

Aussi, ce travail de recherche s'inscrit-il dans la problématique « environnement et santé », dans le volet lutte contre les vecteurs. Il ne s'agit pas d'une recherche épidémiologique, encore moins de médecine, mais plutôt une étude dont le but est de contribuer à la lutte contre le paludisme par une meilleure connaissance de la géographie des anophèles. La présente étude propose une réflexion sur la distribution spatiale des vecteurs. Son ambition est de répondre à des questions essentielles dans la définition de politique et de stratégie de lutte anti-vectorielle. Ces questions sont les suivantes : comment sont réparties, sur le territoire burkinabé, les formes moléculaires M et S *An. gambiae* ?

⁵ NIH : National Institute for Health

⁶ CNRFP : Centre National de Recherche et de Formation sur le Paludisme

⁷ OCEAC : Organisation de Coordination pour la lutte contre les Endémies en Afrique Centrale

⁸ IRD : Institut de Recherche pour le Développement

Dans l'environnement, qu'est-ce qui détermine cette répartition spatiale ?

Objectifs et hypothèses

Ce travail a pour objectifs de :

- décrire la distribution spatiale des formes moléculaires M et S d'*An. gambiae* s.s. au Burkina Faso
- identifier les facteurs qui déterminent cette distribution
- modéliser la distribution spatiale

Il s'appuie sur trois hypothèses : la distribution non aléatoire des formes moléculaires M et S d'*An. gambiae* au Burkina Faso, la détermination de l'organisation spatiale par les facteurs écologiques et la contribution différentielle des différents facteurs environnementaux. Les deux dernières hypothèses ouvrent la perspective d'une formalisation mathématique à partir de données empiriques.

Ces objectifs et hypothèses invitent la géographie à une réflexion plus globale, celle de la chorologie des espèces, particulièrement celles impliquées dans la transmission de certaines maladies. La biogéographie est la branche de la géographie dont l'objet d'investigation est la répartition spatiale des espèces et les causes de cette répartition. Elle dispose d'un ensemble de concepts, de méthodes et d'outils qui pourraient être renouvelés sur le plan opérationnel avec l'apparition de techniques nouvelles de traitement de l'information spatiale. Il s'agit en l'occurrence des systèmes d'information géographique, de la télédétection et de l'analyse spatiale. Si la biogéographie peut fournir un cadre théorique pour l'articulation de l'ensemble du travail, il ne peut suffire ni du point de vue technologique ni du point de vue thématique.

Approche méthodologique

Aussi ce travail s'inscrit-il dans une approche pluridisciplinaire qui intègre à la fois : la biogéographie, comme cadre théorique de structuration de l'étude, l'entomologie, comme discipline qui fournit les connaissances sur les formes moléculaires M et S d'*An. gambiae*. L'analyse spatiale, le système d'information géographique et l'analyse de données statistiques peuvent être appréhendés comme des méthodes et techniques de structuration et de traitement de données.

Ce travail s'inscrit dans une démarche de construction d'une structure formalisée qui rend compte des relations spatiales entre les formes moléculaires M et S d'*An. gambiae* s.s. et le milieu. Il doit permettre une représentation formelle simplifiée de la distribution spatiale de chaque forme moléculaire d'*An. gambiae* s.s. à travers une fonction mathématique et une carte de distribution. Le

processus de construction de ces modèles offre une approche de problèmes anciens (la chorologie des espèces) avec des outils nouveaux et des méthodes réactualisées par les nouvelles technologies de traitements de l'information spatiale.

Plan du mémoire

Ce mémoire est structuré en deux parties. La première, intitulée « Cadrage thématique, démarche scientifique et méthodes », pose les bases de la réflexion. Elle est composée de trois chapitres. Le premier chapitre, « Etat de l'art, définition d'un champ conceptuel et positionnement scientifique », définit, à l'issue d'un état de l'art, le champ d'investigation. Il permet une analyse des concepts et insère l'étude dans une discipline scientifique. Ce qui est indispensable à l'orientation de la réflexion. Le second chapitre « Bio-écologie d'*An. gambiae* au Burkina Faso : une relation à formaliser » est une présentation du paludisme, des anophèles et du Burkina Faso. Ces trois éléments constituent les objets de l'étude, les porteurs de la problématique et des informations traitées dans ce mémoire. Leur connaissance est indispensable à la compréhension du problème étudié, au traitement des données, à l'élaboration de la méthodologie et à l'interprétation des résultats. Dans le troisième chapitre « Approche méthodologique » sont présentées la démarche méthodologique, les méthodes et les techniques de collecte et de traitement des données.

La deuxième partie de ce travail, « structuration des données, élaboration et évaluation des modèles », est tournée vers la vérification des hypothèses et l'évaluation des objectifs. Elle est composée de trois chapitres. Elle présente les résultats du travail à travers la structuration de la base de données et des processus au chapitre IV. Le chapitre V contient la présentation de l'organisation spatiale des formes moléculaires M et S d'*An. gambiae* s.s. au Burkina Faso et les facteurs déterminants de cette organisation. Le chapitre VI propose les modèles de distribution et leur évaluation.

PREMIERE PARTIE : CADRAGE THEMATIQUE, DEMARCHE SCIENTIFIQUE ET METHODE

Cette première partie s'appuie sur la revue de la littérature scientifique pour situer ce travail de recherche dans une discipline scientifique : la géographie. Elle contient aussi une présentation du paludisme, de l'anophèle et du Burkina Faso. Elle présente enfin la méthodologie de l'étude. Cette partie a pour but de cadrer ce travail de recherche tant sur le plan thématique, théorique, méthodologique que géographique. Elle répond aux questions suivantes : dans quelle perspective est conduite cette étude ? Où se situe le site de l'étude et quelles sont ses propriétés ? Comment envisager la réalisation de cette étude ? Quels sont les techniques et les outils qui sont utilisés ?

Cette partie est composée de trois chapitres :

- Le premier vise à fixer les limites de nos ambitions à partir de la littérature scientifique ;
- Le second est une présentation des anophèles et du Burkina Faso (la zone d'étude).
- Le troisième chapitre est consacré à la présentation de la démarche, des méthodes et des techniques.

CHAPITRE I : ETAT DE L'ART, DEFINITION DU CHAMP CONCEPTUEL ET POSITIONNEMENT SCIENTIFIQUE

Cette étude est loin d'être la première qui s'intéresse à la distribution spatiale des espèces vivantes. Elle n'est ni la première contribution de la géographie dans le domaine de la santé, ni la première dans le domaine de la modélisation spatiale. Les travaux antérieurs sur la modélisation en géographie, la chorologie des espèces, l'analyse spatiale, les maladies à transmission vectorielle ou l'épidémiologie, ont donné une richesse aux concepts qui s'exprime par la diversité des interprétations. Ils ont apporté beaucoup à l'évolution de la perception de termes et des techniques qui sont en usage dans ce mémoire. Dans cet ensemble de résultats scientifiques et techniques, il faut trouver sa voie en faisant un usage original des acquis pour répondre à une problématique donnée. Pour cela, il faut partir de ce qui est fait et définir son champ d'investigation en reprécisant les concepts structurants du thème et en indiquant un positionnement scientifique. C'est l'objet de ce chapitre.

Ce chapitre est structuré en 2 sections. La première s'ancre dans la littérature scientifique pour faire une synthèse sur les différents travaux qui abordent la question de la distribution spatiale des vecteurs de maladies. Elle doit permettre aussi de montrer le rôle de la géographie dans l'étude des problématiques liées à la santé et positionner ce travail sur le plan scientifique. La seconde section situe ces investigations dans le champ de la géographie. Elle s'appuie sur le paradigme biogéographique qui lui fournit des hypothèses et des indicateurs pour construire un schéma conceptuel.

1.1. GEOGRAPHIE ET ETUDE DES RELATIONS ENTRE ENVIRONNEMENT ET SANTE

Que vient chercher la géographie dans un questionnement sur la santé humaine ? La maladie est-elle un phénomène qui entre dans le domaine de compétence des géographes et pourquoi ? Ce sont des questions récurrentes auxquelles il faut apporter des réponses qui imposent de définir non seulement la géographie mais également sa démarcation avec les autres sciences dans un esprit de conciliation. Il s'agit de s'offrir une place dans ce processus de lutte contre les maladies sans tomber dans ce que Bavoux, (2007) qualifie de « boulimie et débordement pangéographique ».

Les réponses à ces questions s'appuient sur une définition de l'objet d'étude de la géographie et sa place dans l'étude des problématiques liées à la santé. Il est question, tout en délimitant le champ de

compétence de la géographie, de montrer comment elle interagit avec les autres disciplines pour contribuer à une réflexion sur les questions de santé.

Cette analyse du rôle de la géographie dans l'étude des questions de santé se fonde sur trois axes :

- l'axe conceptuel qui repose sur une évolution de la perception de la lutte contre les maladies en général et les maladies à transmission vectorielle en particulier ; et offre l'opportunité d'une nouvelle contribution de la géographie à la résolution des problèmes de santé,
- l'axe méthodologique fondé sur l'approche géographique dans le contexte de la lutte contre les maladies,
- l'axe technologique orienté sur l'utilisation des outils propres au géographe dans l'étude des problèmes de santé.

I.1.1 GÉOGRAPHIE ET LUTTE CONTRE LES MALADIES

Comment la géographie peut-elle intervenir dans la réflexion sur les problématiques liées à la santé ? A travers son objet d'étude : l'espace. Ce paragraphe développe cette réponse et se subdivise en deux sous paragraphes. Le premier est consacré à la définition de l'objet de la géographie. Le second se focalise beaucoup plus sur la relation entre cet objet et les questions de santé.

I.1.1.1 L'objet de la géographie : l'espace

Durant son évolution sur plus de 2500 ans, la géographie a connu de nombreuses remises en cause et de nombreuses reconstructions de son objet. L'unanimité au sein même de la discipline n'est pas définitivement acquise. Toutefois, la tendance actuelle à laquelle nous souscrivons et qui semble fédérer une grande partie de la communauté des géographes, consiste à dire que l'objet de la géographie, c'est l'espace et son organisation. « Comme beaucoup d'autres, nous pensons que l'objet de la géographie contemporaine porte sur l'espace et son organisation. [...] L'espace est formé, soit par des lieux définis par leurs coordonnées géographiques, soit par des portions plus ou moins étendues de la surface terrestre dont les éléments sont des lieux unis par des relations fonctionnelles que leur confèrent ou leur ont conféré les sociétés humaines ». Chamussy et *al.*, (1977) cité par Ciattoni et *al.*, (2005). Cette définition de l'objet de la géographie basée sur l'espace est sans doute la plus répandue aujourd'hui dans l'enseignement supérieur, même si l'acceptation de la notion d'espace est aussi sujette à discussion.

« L'espace est [...] en réalité, l'ensemble des lieux répertoriés et qualifiés par les géographes, la totalité des signes qu'ils identifient et interprètent pour différencier les objets, les phénomènes à la surface de la terre. Les géographes l'inventorient, le découpent et le ramènent à des éléments isolables et combinables, parlant par exemple d'espaces naturels, ruraux, urbains, perçus, vécus, etc. Ils le cartographient, en extraient les structures directrices, en décryptent les formes, les fonctions et les dynamiques [...]. La géographie permet de répondre aux questions suivantes qui sont essentielles dans tout processus humain : à quoi ressemble la terre ? Où sommes-nous sur la terre et qu'est-ce qu'il y a sur la terre ? » La géographie donne aux hommes la connaissance de l'espace dans lequel ils se situent et les clés pour déchiffrer ses fonctionnements » Bavoux, (2007). La notion d'espace se conçoit donc comme milieu géographique, c'est-à-dire un espace doté de propriétés définies en fonction des problématiques. Dans le cadre de ce travail, l'espace milieu de vie, habitat des anophèles, influence leur répartition. Sa dimension géométrique (forme, taille) comme ces caractéristiques biophysiques, participe à la distribution des fréquences de ces vecteurs du paludisme dans l'espace.

I.1.1.2 Relation entre espace et lutte contre les maladies

L'espace étant l'objet d'étude de la géographie, la lutte contre la maladie s'inscrit en partie dans les compétences du géographe. En effet, la maladie affecte les sociétés et les hommes qui sont situés quelque part sur la terre. Le géographe peut y observer la diffusion, les régularités, les singularités et les discontinuités. Cette relation de la géographie à la réflexion sur les problèmes de santé est particulièrement éloquente dans le cas du paludisme qui est une maladie à transmission vectorielle. Sa propagation au sein des humains n'est possible que par l'intermédiaire d'un vecteur dont la survie dépend énormément du milieu géographique, de l'environnement. La lutte contre les maladies, surtout celles à transmission vectorielle, tire d'énormes intérêts de la géographie qui permet de caractériser l'espace. L'intérêt de la géographie pour la résolution des questions médicales est bien antérieur à celui des géographes pour la question de santé. « La géographie de la santé s'appuie sur une très longue tradition, initiée par Hippocrate lui-même qui avançait le rôle du climat et de l'environnement, mais aussi de la localisation géographique, sur l'apparition des maladies »⁹. La géographie médicale est l'initiative d'un médecin, l'Allemand Leonhard Ludwig Finke (1747-1837).¹⁰ C'est ce qu'exprime Barrett Frank, (2000) en titrant leur article, "Finke's 1792 map of human diseases: the first world disease map?" Le fondateur de l'épidémiologie, le Britannique John

⁹ http://www.scienceshumaines.com/-Oales-sciences-humaines-et-la-sante-0a_fr_13795.html

¹⁰ Barrett A., Frank (2000) "Finke's 1792 map of human diseases: the first world disease map?" *Social Science & Medicine* 50 915-921. <http://www.csiss.org/classics/content/106>

Snow¹¹ (1813-1858) n'était pas non plus un géographe. Son fameux essai sur la propagation de l'épidémie de choléra en 1854 et l'hypothèse d'une dissémination de celle-ci par l'intermédiaire de la distribution d'eau en utilisant les cartes le consacre comme utilisateur de la géographie pour la résolution des questions de santé. Ce qui caractérise ces précurseurs de l'usage de la géographie dans le domaine de la santé, c'est la prise en compte de l'espace. La place de la géographie dans la lutte contre la maladie se situe dans la spatialisation des phénomènes morbides et dans l'intelligibilité de l'espace pour les questions de santé.

L'espace, porte d'entrée du géographe dans l'étude des phénomènes, donne une autre dimension et un autre angle d'attaque dans la lutte contre la propagation des maladies. La géographie peut alors contribuer avec efficacité à la planification de lutte. Cette contribution de la question spatiale dans la lutte contre les maladies est de plus en plus exigée par les spécialistes de la santé et les études intégrant la dimension spatiale ou mésologique de la maladie sont légions. L'émergence du concept « écosanté » procède de l'intégration de la lutte contre les maladies dans un système global qui prend en compte l'environnement (espace géographique) à travers ses différentes déclinaisons. Ainsi, l'usage de plus en plus courant de méthodes et outils traditionnellement connus comme géographiques est perceptible dans les sciences de la santé. Le concept de complexe pathogène¹² développé par M. Sorre en 1933 a imposé aux plus sceptiques la nécessité d'intégrer les phénomènes morbides dans un système plus large. ¹³Le-Bras et Malvy, (2004) confirment cet avis en soutenant le fait que « le concept de complexe pathogène défini par le géographe Maximilien Sorre en 1933 mérite d'être « revisité » à la lumière des nouveaux outils permettant de mieux identifier les déterminants biologiques (génétique moléculaire), d'environnement (système d'information géographique), des comportements, dépendants des activités des services de soins. Son utilisation invite à porter sur toute menace de maladie un regard à l'échelle du monde et à envisager les conséquences des bouleversements de l'époque tels que la mondialisation des personnes, des denrées et des biens, la dynamique démographique, l'urbanisation rapide, l'industrialisation, la déforestation, l'évolution rapide des comportements, la capacité d'efficacité mais aussi de nuisance des systèmes de santé ». En exemple, nous citons des publications d'articles dans des revues spécialisées comme Bio-

¹¹ http://fr.wikipedia.org/wiki/John_Snow

¹² « L'interdépendance des organismes mis en jeu dans la production d'une même maladie infectieuse permet de dégager une unité biologique d'un ordre supérieur : le complexe pathogène. Il comprend, avec l'homme et l'agent causal de la maladie, ses vecteurs et tous les êtres ou facteurs qui conditionnent ou compromettent leur existence » Sorre, 1961, cité par Raphaël Pin-Diop 2006.

¹³ Le Bras M et Malvy J-M. D., Le complexe pathogène tropical : Regard nouveau sur un concept ancien, Médecine tropicale 2004, vol. 64, n° 6, pp. 613-618 (<http://cat.inist.fr/?aModele=afficheN&cpsidt=16594466>)

Med : Costantini et *al.*, (2009)¹⁴, Thomas et *al.*, (2003)¹⁵, de Mushinzimana et *al.*, (2006),¹⁶ de Ratana et *al.*, (2005)¹⁷, et de Vallée et *al.*, (2007)¹⁸, Sabatier et *al.*, (2004)¹⁹. Si donc on retrouve la géographicit  dans de nombreuses questions scientifiques, c'est parce que « [la g ographie] est belle, utile et ais e tout ensemble. On peut dire qu'elle est n cessaire   tout le monde²⁰ ». C'est parce que son objet est associ    des questions existentielles, pratiques, que l'usage des produits de la g ographie sert   tout le monde   un certain moment.

En somme, de par son objet, ses domaines de comp tences, la g ographie doit apporter sa contribution aux questions de sant  et ce travail de recherche participe de cette dynamique. Elle est g ographique par ses objectifs et par son approche m thodologique. Si l'espace fait partie int grante de l'objet de la g ographie, tout usage de l'espace n'est pas forcément g ographique. La g ographie a son originalit  dans l'usage de l'espace comme moyen de lecture et compr hension des ph nom nes.

I.1.2 CONTRIBUTION DE LA GEOGRAPHIE A LA LUTTE CONTRE LES MALADIES

L'objet de la g ographie justifie son implication dans la recherche sur les questions de sant . L'approche des ph nom nes spatiaux en g ographie a cependant des sp cificit s. En effet, l'approche m thodologique des g ographes, en tant que sp cialistes du spatial, diff re sans doute des autres disciplines scientifiques et constitue une deuxi me raison de l'implication de la g ographie dans le domaine de la lutte contre les maladies. Comment l'espace intervient-il au plan

¹⁴ Costantini, C., Ayala, D., Guelbeogo, W.M., Pombi, M., **Som , C.Y.**, Bassole, I.H., Ose, K., Fotsing, J.-M., Sagnon, N.F., Fontenille, D., Besansky, N.J., et Simard, F., 2009, Living at the edge: biogeographic patterns of habitat segregation conform to speciation by niche expansion in *Anopheles gambiae*: BioMed Center Ecology, v. 9:16.

¹⁵ Thomas P Eisele, Joseph Keating, Chris Swalm, Charles M Mbogo, Andrew K Githeko, James L Regens, John I Githure, Linda Andrews et John C Beier, 2003, Linking field-based ecological data with remotely sensed data using a geographic information system in two malaria endemic urban areas of Kenya, *Malaria Journal* 2003, 2:44, BioMed Central

¹⁶ Emmanuel Mushinzimana¹, Stephen Munga¹, Noboru Minakawa, Li Li, 2006Chen-chiang Feng, Ling Bian, Uriel Kitron, Cindy Schmidt, Louisa Beck, Guofa Zhou, Andrew K Githeko¹ et Guiyun Yan, 2006, Landscape determinants and remote sensing of anopheline mosquito larval habitats in the western Kenya highlands , BioMed, *Malaria Journal*2006, 5:13

¹⁷ Ratana Sithiprasasna, Won Ja Lee, Donald M Ugsang et Kenneth J Linthicum, Identification and characterization of larval and adult anopheline mosquito habitats in the Republic of Korea: potential use of remotely sensed data to estimate mosquito distributions, BioMed Central, *International Journal of Health Geographics* 2005, 4:17

¹⁸ Julie Vall e, Marc Souris, Florence Fournet, Audrey Bochaton, Virginie Mobillion, Karine Peyronnie and G rard Salem, 2007, Sampling in health geography: reconciling geographical objectives and probabilistic methods. An example of a health survey in Vientiane (Lao PDR) BioMed Central, *Emerging Themes in Epidemiology* 4:6 doi:10.1186/1742-7622-4-6 <http://www.ete-online.com/content/4/1/6>

¹⁹ P. Sabatier, Babouchkine, J.-M., Morvan, J., Polidori, L., Lacaux, J.-P., Braak, L., Dubois, M.A., et Chaude, H., 2004, S2.Dengue. Surveillance spatiale et de la dengue. Conception d'outils d'acquisition, de cheminement, d'analyse et de pr vision  pid miologique: ITBM-RBM v. 25 (2004) p. 292-296.

²⁰ D. Martineau, 1970, Nouvelle G ographie, Paris, in Jean-Jacques Bavoux (2007)

méthodologique en géographie ? En quoi cette approche de l'espace constitue-t-elle un apport dans l'analyse des problématiques de santé ?

I.1.2.1 L'espace dans l'analyse géographique

Les géographes, dans leur diversité, ont une conception assez partagée de l'espace qui est leur dénominateur commun. Pour eux, l'espace est une porte d'entrée dans la connaissance des phénomènes. Il est le ferment d'un ensemble de questionnements permettant de connaître l'état, le fonctionnement et la dynamique des phénomènes. Ces questionnements fondent une approche particulière en géographie : l'analyse spatiale. Cette dernière, en tant qu'approche géographique, localise les phénomènes, révèle leurs structures et leurs configurations. Elle s'appuie sur l'organisation spatiale des phénomènes pour les comprendre et les expliquer. Considérer l'espace comme source de compréhension des phénomènes est en soit une approche qui confère à la géographie, science de l'espace, une dimension originale. Ainsi, pour le géographe, le caractère d'un lieu dépend, au moins partiellement, de la situation dans l'espace. La récurrence de certaines structures et configurations spatiales doit donc être étudiée pour comprendre et expliquer le fonctionnement et la dynamique des phénomènes dans l'espace. La statistique classique, non spatiale par tradition, est largement bâtie sur le postulat d'un espace neutre, simple support des phénomènes étudiés. Dans cette perspective, la localisation d'observations dans l'espace (absolu) et leur position les unes par rapport aux autres (espace relatif) n'exercent aucune influence sur la nature même de ces observations. « Cette négation de la première loi de la géographie de Tobler, selon laquelle chaque phénomène est relié à tous les autres, mais des phénomènes proches dans l'espace auront tendance à être davantage liés que des phénomènes éloignés, ne peut que faire frémir le géographe » Banos, (2001).²¹ Cette position centrale de l'espace comme moyen de compréhension des phénomènes est si importante qu'on ne saurait faire de la géographie en dehors de l'espace. C'est du moins ce qu'affirme Abler et *al.*, (1971) in Banos, (2001) en disant : « la géographie n'aurait pas lieu d'être si l'ubiquité caractérisait toute chose ».

L'analyse spatiale, approche spécifique de l'espace dans l'analyse des phénomènes, est née dans les années 1960 dans le monde anglo-saxon et s'est particulièrement appuyée sur des disciplines comme l'économie spatiale et la sociologie de l'Ecole de Chicago. Ses précurseurs sont Von Thüm, Christaller et, beaucoup plus tard Paul Claval et Philippe Pinchemel. Elle prend corps dans les différentes branches de la géographie qui lui fournissent les paradigmes spécifiques. L'étude de

²¹ Arnaud Banos, « A propos de l'analyse spatiale exploratoire des données », *Cybergeo : European Journal of Geography* [En ligne], Systèmes, Modélisation, Géostatistiques Systèmes, Modélisation, Géostatistiques,]-->,document 197, mis en ligne le 18 octobre 2001. URL : <http://www.cybergeo.eu/index4056.html>

l'armature urbaine d'un pays n'utilise pas les mêmes concepts que celle relative à l'organisation agraire, même si ces deux études peuvent interférer dans certains cas. Ainsi l'espace est territoire, biotope ou écosystème et autre chose en fonction des circonstances. Il est perçu, vécu ou construit. Il est physique ou conceptuel. Cette diversité d'appréhension de l'espace en géographie s'intègre dans différents schèmes dont les branches de la discipline sont porteuses. L'espace est donc un élément de fécondité dans l'entreprise scientifique du géographe. Il est à la fois diversement utilisé selon les circonstances et porte toujours en lui une source d'explication et de compréhension des phénomènes à étudier.

Dans le cadre de ce travail, l'espace, lieu de vie du complexe *Anopheles gambiae*, vecteur du paludisme, présente une variabilité d'état qui le rend plus ou moins propice à la vie de ce culicidé. L'espace contribue plus ou moins directement à l'amplification ou à la réduction de la transmission du paludisme. Nul ne peut donc comprendre la propagation du paludisme en excluant l'espace qui détermine, au moins en partie, le cycle de transmission de la maladie. Cette vision de l'espace, milieu de vie, provient de la biogéographie et a été déterminante dans la conceptualisation de la démarche de cette étude. Ainsi, l'analyse spatiale en tant qu'approche géographique a pris corps dans le paradigme biogéographique et a permis de proposer une perception objective de la distribution spatiale de membre du complexe *An. gambiae*.

I.1.2.2 Analyse spatiale et analyse des problèmes de santé

La lutte contre les maladies nécessite de comprendre leur étiologie et leur modèle de diffusion. Dans le processus d'éclosion des maladies, l'espace peut prendre une importance particulière, car la maladie survient dans certains milieux et pas dans d'autres, parce que l'espace y offre les conditions favorables. Les travaux du médecin James Lind 1768, celui du père de l'épidémiologie John Snow, sur l'épidémie de choléra à Londres en 1854, de l'épidémiologiste Pavlovsky en 1930 avec sa théorie des foyers naturels des maladies, ceux de M. Sorre (1933) et son concept de complexe pathogène, en sont des illustrations éloquentes. Cela est particulièrement vrai pour les maladies infectieuses et les maladies à transmission vectorielle. La géographie peut permettre de révéler l'organisation spatiale de la maladie en mettant en exergue les structures et les configurations spatiales. Elle connaît un renouveau dans les années 1960 avec des ambitions de modélisation et de quantification. C'est le triomphe des quatre grandes familles, celle de Von Thünen, celle Christaller révisée par Lösch, celle de Reilly et Thiessen, et moins congrûment celle de Zipf, Brunet (2000b)²². C'est la période

²²Roger Brunet; conférence prononcée le 24 novembre 1999 et publiée dans le Bulletin de la Société de Géographie de Liège, 2000, n°2, p. 21-30.

d'apparition des modèles de diffusion associés au calcul des probabilités. C'est aussi l'époque des modèles de simulation qui visaient à reproduire des systèmes dynamiques. Ces modèles s'intéressaient à anticiper les chances ou les effets d'un changement. C'est encore l'avènement des modèles de configurations spatiales, notamment de réseaux et graphes sagittaux, visant à modéliser des systèmes. La géographie, à travers la statistique spatiale exploratoire et l'analyse des configurations spatiales, a prouvé son importance dans la compréhension des problèmes de santé et dans la lutte contre les maladies.

Lorsqu'il s'agit de maladies à transmission vectorielle, comme c'est le cas dans ce travail de recherche, l'analyse de la répartition spatiale des espèces vectrices de maladies peut apporter une contribution importante dans la compréhension de la survenue et de la dynamique des phénomènes morbides. Cette compréhension soutient la conception des plans de riposte efficace. La contribution de ce travail, à la compréhension de la distribution des formes moléculaires M et S d'*An. gambiae* s.s., s'intègre dans une perspective de recherche de solution efficace pour lutter contre la transmission vecteur/homme du paludisme. En mettant en relief les facteurs déterminants de la distribution des différentes formes moléculaires au Burkina Faso, cette étude ouvre le champ à d'autres investigations comme l'utilisation de méthode biologique dans le contrôle des populations d'anophèles, la mise en place de pesticides plus adaptés à la population culicidienne, etc.

L'approche spatiale de la géographie apporte certes une contribution intéressante à la connaissance et à la lutte contre les maladies. La mise en œuvre de cette approche est possible grâce à l'utilisation des technologies dont il convient également d'analyser la contribution à la résolution des problématiques liées à la santé.

I.1.2.3 Techniques géographiques et étude des problèmes de santé

Les géographes utilisent certaines techniques en vue d'opérationnaliser leur approche. Il s'agit essentiellement de la cartographie, des systèmes d'information géographique et de la télédétection. Selon le cas, les outils spatiaux servent à formuler des hypothèses, à vérifier des hypothèses posées au préalable, à orienter les activités de terrain et à faire des prédictions (Pin-Diop, 2006). Ces techniques offrent de ce fait des possibilités de production de données et de traitement fort utile dans de nombreuses disciplines et participent également au positionnement de la géographie comme une science importante dans l'analyse des questions de santé. L'engouement qu'elles suscitent auprès de bien d'autres disciplines est sans doute lié aux opportunités particulières qu'elles offrent à la

communauté scientifique. C'est ce que l'on présente très succinctement dans les sous paragraphes suivants.

L'outil cartographique

Le cartographe par défaut est devenu, dans certaines contrées, à certaines périodes, le géographe. Ainsi, « du XVe au XVIIe siècle, dans le contexte des grandes découvertes maritimes, la figure du géographe et celle du cartographe se confondent en Europe » Zaninetti, (2005). « La cartographie désigne la réalisation et l'étude des cartes. Elle mobilise un ensemble de techniques servant à la production des cartes. La cartographie constitue un des moyens privilégiés pour l'analyse et la communication en géographie. Elle sert à mieux comprendre l'espace, les territoires et les paysages. Elle est aussi utilisée dans des sciences connexes, comme la démographie, l'économie, dans le but de proposer une lecture spatialisée des phénomènes » (Encyclopédie scientifique en ligne)²³.

Outils de spatialisation par excellence, la cartographie est à la fois pour le géographe, un outil d'analyse et un puissant moyen de communication. C'est outil de communication graphique qui s'appuie sur un langage visuel et qui permet une vision simultanée et synthétique des phénomènes spatiaux. Les travaux comme ceux de Jacques Bertin²⁴, Serge Bonin²⁵ Roger Brunet²⁶ sur la sémiologie graphique²⁷ pour une utilisation de la carte en tant que média, en tant que moyen de communication mais également en tant que méthode d'analyse en sont une preuve. « Les cartes servent à représenter des phénomènes géographiques, c'est-à-dire des phénomènes dont la configuration spatiale produit du sens » (Encyclopédie scientifique en ligne). La cartographie reconnue plus ou moins unanimement comme un outil privilégié du géographe est l'une des techniques qui a introduit la discipline géographique dans de nombreux champs de recherche et particulièrement celui de la santé. L'évolution technologique, à travers l'avènement de la télédétection et du SIG, a facilité la réalisation des cartes.

²³ <http://www.techno-science.net/?onglet=glossaire&definition=2569>

²⁴ Jacques Bertin, 1967, sémiologie graphique : les diagrammes, les réseaux, les cartes, EHESS, édition originale, 2005 4ème édition, 452 pages

²⁵ Serge Bonin, 1975, initiation à la graphique, EPI, 172 p.

²⁶ Brunet Roger: la carte mode d'emploi. FAYARD / RECLUS, 1987, 270 p.

²⁷ « On appelle sémiologie graphique l'ensemble des règles permettent l'utilisation d'un système graphique de signes pour la transmission d'une information » Béguin, Michèle et Pumain, Denise (1994), La représentation des données géographiques. Statistique et cartographie. Paris, Armand Colin.

L'outil système d'information géographique (SIG)

« Aujourd'hui la géomatique est considérée comme un puissant outil pour l'élaboration des travaux qui exigent une connaissance plus approfondie de la dynamique spatiale et temporelle des phénomènes et qui requiert la manipulation, le stockage et la visualisation des gros volumes de données » Fotsing, (2003). L'apparition des premiers logiciels de SIG « bureautiques » sur PC dans les années 1990 (Arcview, Mapinfo, etc.), la réduction considérable des coûts des logiciels et des ordinateurs les rendent accessibles aux petites entreprises individuelles, aux laboratoires de géographie des facultés de lettres désargentés. Cette accessibilité aux technologies des SIG, va galvaniser l'usage de la cartographie et de l'analyse spatiale Zaninetti, (2005). Désormais, l'usage de la carte et de l'analyse spatiale est techniquement accessible. « Jusqu'à cette époque récente, la réalisation d'une carte relevait d'un parcours du combattant. Le spécialiste de la cartographie devait avoir d'énormes connaissances en géodésie, en mathématique etc. et avoir une dextérité extraordinaire dans le maniement de la plume et autres outils de dessin » Somé, (2006). La prise en charge de nombreux calculs et l'amélioration de la technique de dessin déjà avec la cartomatique puis avec les SIG ont entraîné un intérêt pour l'espace dans de nombreuses disciplines. Cet intérêt pour l'espace dans de nouvelles disciplines est également provoqué par les SIG. Ces derniers ont fourni des modules puissants de calcul de certains paramètres et d'évaluation de certains indicateurs qui ont nécessité de nombreuses recherches sur la structure de l'information géographique. Nous citons comme exemple de réflexion Souris, (2006). Les modèles d'interpolation et d'estimation, les modèles de transformation géométrique et de projection, les fonctions d'extrapolation et de mise en page, sont des composantes de traitement des SIG qui ont été mis à contribution dans le cadre de cette étude. Toutes ces routines de calcul ne sont pas ignorées et ne peuvent pas être ignorées dans le cadre de la réflexion sur les problèmes de santé. Les travaux en santé utilisant les SIG comme outils d'analyse ou de cartographie se multiplient. A titre illustratif, citons la thèse en épidémiologie de Tran, (2004) soutenue en 2004 à l'Université de Strasbourg 1, celle de Biteau, (2006) soutenue à l'Université Montpellier II, celle de Bouyer, (2006) en entomologie médicale soutenue également à l'Université Montpellier II et celle de celle de Balenghien, (2006) à l'Université des sciences des technologies et de la santé de Grenoble soutenue en 2006. A titre illustratif pour les publications scientifiques, rappelons également, les références déjà utilisées dans la section précédente telles que Costantini et *al.*, (2009)²⁸ Thomas et *al.*, (2003), de Mushinzimana et *al.*, (2006) et de Ratana et *al.*, (2005).

²⁸ Nous sommes coauteur avec les autres membres du projet de cet article.

Les SIG, dans leur fonction de stockage et d'intégration de données, utilisent bien souvent des informations provenant de la télédétection, la plus grande source actuelle de données spatiales.

La télédétection

En tant qu'ensemble de méthodes et de techniques permettant d'acquérir des informations sur les surfaces sans contact direct avec celles-ci Bonn, (1991), la télédétection a permis au cours des trois dernières décennies de produire une grande quantité de données spatiales Bouzou et *al.*, (2001); Somé, (2000). « L'application de la télédétection à la surveillance et au contrôle des vecteurs de maladies remonte à 1949, lorsque Audy utilisa des photographies aériennes en noir et blanc pour déterminer les foyers endémiques de typhus en Asie du sud-ouest, à partir de la distribution d'un certain type de végétation » Pin-Diop, (2006). Depuis le lancement des premiers satellites civils d'observation de la terre dans les années 1970, les progrès scientifiques dans le domaine de la télédétection sont énormes. Les résolutions spatiales, spectrales et temporelles sont en progrès. Les méthodes de traitement des données issues de la télédétection se sont également beaucoup améliorées. Les logiciels sont de plus en plus performants et intègrent une grande diversité d'indices (indice de végétation, indice minéralogique, indice ferrugineux, indice hydrologique, etc.). Ceux-ci sont rattachés à des surfaces spécifiques qui peuvent avoir un intérêt dans le profil épidémiologique de certaines maladies. Par son caractère répétitif, synoptique et diversifié du point de vue des thématiques, la télédétection est une source d'information indispensable dans la prévention et le suivi de certains phénomènes qui affectent la santé des populations. La télédétection, en permettant le suivi de certains paramètres environnementaux, constitue une source d'information importante pour l'étude de certaines maladies. Ainsi, et de plus en plus, les capteurs de télédétection sont à même de fournir aux épidémiologistes l'information pertinente sur le milieu pour l'étude et la surveillance des maladies liées à des conditions environnementales particulièrement les maladies à transmission vectorielle Tran, (2004). Son impact sur les intervenants dans le domaine de la santé a conduit à l'émergence du concept de « Télésanté » qui s'appuie à la fois sur la télédétection, les SIG, l'analyse cartographique et d'autres technologies de l'information et de la communication (Internet, télécommunication) dont Sabatier et *al.*, (2004), Vallée et *al.*, (2007) en sont des illustrations.

1.2. DEFINITION DU CHAMP CONCEPTUEL ET POSITIONNEMENT SCIENTIFIQUE

Dans les paragraphes précédents, l'accent a été mis sur le rôle que peut jouer la géographie dans le domaine de la santé. Cette fécondité n'est pas sans inconvénient car, comme la géographie, un travail

de réappropriation de concept et des outils est effectué dans ces disciplines. Il est évident, au regard de ce qui précède, que les termes qui constituent le thème de ce travail de recherche peuvent faire l'objet d'une interprétation plurielle. « Une même définition dans une même discipline peut cacher des paradigmes, méthodologies et enjeux différents, voire divergents ou contradictoires, et une même posture théorique dans deux disciplines différentes peut recourir à des mots différents » Langlois, (2005). Il convient, en raison des interprétations auxquelles ils sont sujets, de les redéfinir dans le contexte de ce travail. De prime abord, trois concepts sont essentiels : la modélisation, l'analyse spatiale et le SIG.

I.2.1 MODELISATION : DEFINITION ET TYPE

« Le terme modélisation désigne à la fois l'activité de produire un modèle et le résultat de cette activité » Langlois et Reguer, (2005). « La modélisation est l'art ou la science de produire des modèles au sens fort et de hisser au rang de modèles des observations rassemblées, c'est-à-dire d'identifier les éléments majeurs et les relations les plus assurées qui expriment ou bien la structure d'un objet particulier, ou bien la structure commune à des objets d'une même classe » Brunet et *al.*, (1993).

Le modèle, quant à lui, est « une structure formalisée pour rendre compte d'un ensemble de phénomènes qui possèdent entre eux certaines relations » Sanders, (2001) Cette première définition fait allusion aux relations entre des objets au sein d'une structure. Elle présuppose une totalité et une interdépendance entre les éléments d'un système. Cette totalité n'est pas le fruit du hasard. Cette première acceptation du modèle renvoie à la notion de système et donc à la perspective holistique. C'est le modèle - structure. « Nous ignorons si ce que nous disons est vrai, mais nous savons que cela a un sens » Langlois et Reguer, (2005). La modélisation se situe ici dans une démarche de construction de la structure logique, du fonctionnement d'un phénomène conçu comme système.

Un modèle est une « représentation schématique d'un processus, d'une démarche raisonnée » Guermond, (2005). C'est une « représentation idéalisée du monde réel construite pour démontrer certaines de ses propriétés » Haggett, (1965) in Brunet, (2000b). Le modèle est également une « représentation formelle et épurée du réel ou d'un système de relations [...] à des fins d'interprétation » Brunet, (2000b). Cette seconde acceptation du modèle renvoie à la notion de modèle - schéma, une représentation graphique. Elle consiste donc à « associer à une formalisation verbale (comme en mathématique, en physique, en chimie, en géographie, etc.) une formalisation graphique (comme un graphique associé à un graphe, un diagramme de phase associé à un système différentiel, un schéma

moléculaire associé à une formule chimique, une carte associée aux valeurs d'un tableau statistique). Le modèle - schéma est une représentation du modèle - structure » Langlois et Reguer, (2005). Elle peut permettre la mise en forme empirique d'observations non encore expliquées. Dans cette perspective, « nous ignorons si ce que nous disons a un sens, mais c'est vrai » Langlois et Reguer, (2005). La démarche, dans ce cas, est de partir d'un jeu de données observées ou d'un système déjà formalisé pour construire une représentation d'un phénomène de façon à en simplifier soit la description, soit l'explication, soit l'interprétation.

Quelle que soit la perspective, la modélisation procède de :

- la représentation qui est un exercice d'abstraction de la réalité. Elle aboutit à un schéma fidèle de la réalité mais qui lui est différent. La carte topographique en est une illustration presque parfaite ; elle est une représentation fidèle du territoire mais qui lui est différent ;
- la formalisation qui est synonyme de rigueur méthodologique et confère au modèle sa capacité à fournir les moyens de sa répliquabilité sur des phénomènes pouvant être différents. C'est dire que le modèle fournit un langage, un corpus de normes et un canevas pour la représentation du phénomène ;
- l'interprétation : le modèle en tant que résultat d'un processus cognitif, débouche sur une interprétation c'est-à-dire qu'il doit donner un sens Brunet, (2000a).

Toute modélisation doit donc avoir pour objectifs de :

- comprendre : saisir les motivations qui ont un sens pour chaque homme ;
- expliquer : trouver une loi générale extérieure à l'objet ;
- interpréter : donner une signification par la mise en relation d'un champ de représentation (signifié) avec un autre (signifiant).

Le sens du concept de modélisation, dans ce travail de recherche, ne s'écarte pas du contenu qui lui est donnée ci-dessus. Il faut donc entendre par modélisation une démarche de construction d'un modèle. Le modèle est à la fois une structure formalisée pour rendre compte d'un ensemble de phénomènes qui possèdent entre eux des relations et une représentation formelle et épurée du réel ou d'un système de relations à des fins d'interprétation.

Le processus de la modélisation, tel que défini, s'appuie sur une démarche scientifique qui comprend à la fois le choix des données, le choix des échelles, de la métrique pour l'observation et la collecte des données, leur structuration, leur organisation. Pour cela, la modélisation recourt, d'une certaine

façon, à d'autres modèles existants dans le cadre de disciplines scientifiques et techniques. La modélisation est donc difficile à envisager en dehors des cadres théoriques scientifiques qui ont fondé les hypothèses et les modes de raisonnement. Elle ne peut se réaliser non plus sans les outils techniques et méthodes d'opérationnalisation de la conception. Dans le cadre de cette étude, l'analyse spatiale constitue un des cadres méthodologiques de la modélisation.

I.2.2 L'ANALYSE SPATIALE

L'analyse spatiale est l'étude formalisée de la configuration et des propriétés de l'espace produit et vécu par les sociétés humaines Pumain et Saint-julien Thérèse, (1997). Elle est aussi un « ensemble de méthodes mathématiques et statistiques visant à préciser la nature, la quantité et la qualité attachées aux lieux et aux relations qu'ils entretiennent – l'ensemble constituant l'espace – en étudiant simultanément attributs et localisation » Brunet, (2000a). Elle « recouvre un ensemble de théories géographiques et de méthodes provenant de la physique, des mathématiques et de l'informatique » Mathian et Piron, (2001). L'analyse spatiale est donc à la fois cadre de pensées et ensemble de méthodes. En tant que cadre de pensées, elle considère l'espace à la fois comme support d'existence et comme produit de l'action. A ce titre, l'espace peut servir de justification et d'explication scientifique autonome. En tant qu'ensemble de méthodes, l'analyse spatiale regorge de moyens (procédures, modèles, techniques, indices, etc.) qui permettent à la fois la transformation des données, leur analyse et leur interprétation d'un point de vue spatial.

L'analyse spatiale est utilisée dans le cadre de cette étude à la fois comme méthode et technique de traitement de données (analyse des séries géostatistiques, choix des techniques d'estimation, d'extrapolation, etc.) et comme méthode d'analyse géographique. Elle doit permettre de montrer s'il existe une organisation / une structuration dans l'espace des populations d'*An. gambiae* s.s. au Burkina Faso. Elle devra, si cette organisation existe, fournir les outils pour mettre en exergue les facteurs d'organisation permettant d'expliquer la configuration observée.

L'analyse spatiale, dans le cadre de cette thèse, ne peut donc se concevoir en dehors de la thématique qui est la distribution spatiale d'*An. gambiae* s.s.. En liaison avec le concept de modélisation, il s'agit de trouver un modèle qui représente la façon dont les configurations spatiales des populations du complexe *An. gambiae* s'expriment. Ce modèle, représentation simplifiée de cette configuration, doit permettre de la comprendre et de l'expliquer. L'espace est une composante importante de cette modélisation, c'est pourquoi, l'analyse spatiale est retenue comme méthode. Encore faut-il des outils pour l'application des techniques !

I.2.3 LE SYSTEME D'INFORMATION GEOGRAPHIQUE (SIG).

Parmi les nombreuses définitions des SIG, nous retenons celle de Thériault, (1995) pour qui, le SIG est défini comme un ensemble de principes, de méthodes, d'instruments et de données à référence spatiale, utilisé pour saisir, conserver, transformer, analyser, modéliser, simuler et cartographier les phénomènes et les processus distribués dans l'espace géographique.

Les SIG ont deux composantes structurelles :

- la composante utilisateur qui comprend les hommes et les projets ;
- la composante informatique qui regroupe les ordinateurs, les logiciels SIG et les périphériques spécifiques tels que les systèmes de radio positionnement, les scanners, imprimantes, tables à numériser, etc.

Par sa composante utilisateur, le SIG sert de cadre d'organisation de la collecte de données et de leur structuration pour la formalisation de la distribution spatiale d'*An. gambiae* s.s.. Par sa composante informatique, il offre à l'analyse spatiale des outils de structuration et de stockage des données en vue d'en faciliter l'accès. Il concentre des outils d'application des techniques (statistique, calcul d'indice, interpolation et estimation, représentation cartographique, etc.).

En synthèse de cette analyse conceptuelle, ce travail de recherche s'inscrit dans une démarche de construction d'une structure formalisée qui rend compte des relations spatiales entre *An. gambiae* s.s. et le milieu. Elle vise à une représentation formelle simplifiée pour permettre d'expliquer, de faire comprendre les configurations spatiales observées. Cette synthèse permet non seulement une meilleure définition des objectifs, mais aussi d'inscrire cette étude dans un champ scientifique qui fournira les outils à la réalisation des objectifs.

I.2.4 PARADIGME BIOGEOGRAPHIQUE : UNE MATRICE CONCEPTUELLE POUR LA CONSTRUCTION DU PROJET DE MODELISATION

Après la définition des concepts qui ont permis de donner une vue d'ensemble sur le thème, cette section précise les objectifs visés par ce travail de recherche et l'inscrit dans un champ disciplinaire. Ce champ disciplinaire lui fournit les moyens théoriques nécessaires pour la formulation des bonnes hypothèses et des indicateurs qui en permettront la vérification. Cette section est donc consacrée à la présentation des objectifs, du positionnement scientifique, des hypothèses ainsi que des indicateurs.

I.2.4.1 Objectifs de recherche

L'objectif global de ce travail est d'étudier la distribution spatiale des formes moléculaire M et S *An. gambiae* s.s. et les causes de cette répartition au Burkina Faso en utilisant les SIG et l'analyse spatiale. C'est, en quelque sorte, l'autécologie d'*An. gambiae* s.s. au Burkina Faso avec l'utilisation des SIG et de l'analyse spatiale. Il s'inscrit dans la perspective de lutte contre les maladies à transmission vectorielle qui se décline par le biais du contrôle de la population de vecteurs. Elle se décompose en trois objectifs spécifiques :

1. décrire la distribution spatiale d'*An. gambiae* s.s. au Burkina Faso. Ce premier objectif doit permettre de répondre aux questions suivantes : comment se répartissent, sur le territoire burkinabé, les formes moléculaires M et S d'*An. gambiae* s.s. ? Y a-t-il une structure ou une direction particulière, une tendance à des regroupements ou toute autre forme d'organisation dans la répartition spatiale des membres du complexe *An. gambiae* ?
2. identifier les facteurs environnementaux qui déterminent la distribution spatiale des formes moléculaires M et S d'*An. gambiae* s.s. au Burkina Faso. Ce qui intéresse, dans ce second objectif, ce sont les causes de la distribution observée et décrite. Nous cherchons donc une explication ou plutôt une justification de la distribution spatiale décrite.
3. modéliser la distribution spatiale des membres du complexe *An. gambiae* au Burkina Faso. Il s'agit de formaliser la description obtenue en mettant à contribution l'explication qui en découle. Ce modèle doit permettre de reproduire la distribution sans être contraint à l'observation directe sur le terrain tout en minimisant le plus possible les écarts avec la réalité. Il doit permettre d'expliquer la distribution des espèces concernées.

L'atteinte des objectifs ci-dessus cités n'est envisageable que si l'on s'inscrit dans une perspective théorique en positionnant le thème dans une démarche scientifique éprouvée. Pour cela, il convient de choisir un champ disciplinaire. C'est de cela qu'il sera question dans le paragraphe « Positionnement scientifique ».

En effet, celui qui ambitionne de décrire, de comprendre et d'expliquer un phénomène sur le plan scientifique, doit s'adosser à une science existante, à moins d'en créer une autre. Modélisation, analyse spatiale et SIG n'ont donc de pertinence que s'ils s'inscrivent dans un champ disciplinaire approprié qui leur confère un encrage thématique. Ce dernier permet de disposer d'indicateurs qui ont un sens pour la mesure des relations, d'avoir des méthodes d'élaboration des mesures, etc. Le champ disciplinaire qui apparaît le plus convenable dans le cadre de cette étude est la biogéographie

parce qu'elle est la branche de la géographie qui s'intéresse à la répartition des êtres vivants à la surface de la terre ainsi qu'aux causes de cette répartition.

I.2.4.2 Paradigme biogéographique et conceptualisation de la modélisation de la distribution spatiale

Il est aujourd'hui établi que la distribution des espèces animales et végétales à la surface de la terre n'est ni uniforme, ni homogène, ni isotrope. Chacune y occupe une aire qui lui est propre, c'est-à-dire une surface d'étendue variable, continue ou discontinue, qui circonscrit l'ensemble des localités où les populations d'espèces se rencontrent à l'état spontané Lacoste et Salanon, (2006). C'est ce qui permet la délimitation des aires de distribution géographique des espèces ainsi que celles des autres unités taxonomiques des organismes vivants.

La biogéographie qui s'intéresse à la répartition des espèces à la surface de la terre en a fait son objet d'étude. De ce fait, elle a développé un paradigme, c'est-à-dire « un ensemble de concepts fondamentaux, d'hypothèses théoriques générales, de démarches méthodologiques et de techniques qui font consensus et qui guident la recherche scientifique » en son sein Bavoux, (2007). Ainsi, le paradigme biogéographique nous offre un cadre de pensée, une perspective théorique et une matrice conceptuelle dans laquelle nous trouvons les concepts de base nécessaires à la théorisation de notre objet d'étude. La plupart des modèles écologiques développés tirent leur origine dans ce paradigme biogéographique ainsi que le confirme Guisan et Zimmermann, (2000) en affirmant : « The core theory of predictive modeling of biotic entities originates from major trends published in the field of biogeography. »

Dans cette perspective biogéographique, la distribution spatiale des organismes vivants est le résultat de l'influence de facteurs tant internes qu'externes à ceux-ci.

Les facteurs internes aux organismes qui déterminent leurs aires de distribution sont :

- la capacité de propagation,
- l'amplitude écologique,
- le potentiel d'évolution.

La capacité de propagation d'un organisme vivant est sa capacité d'expansion. Elle dépend à la fois de son potentiel reproductif et de sa capacité de dissémination. Son potentiel reproductif est son aptitude à produire de nouveaux individus en un temps donné dans des conditions optimales. Ce potentiel reproductif n'exprime pas la capacité de reproduction totale car cette dernière dépend d'autres paramètres comme la longévité, la durée de vie féconde, etc. La capacité de dissémination

exprime l'aptitude de l'organisme à se déplacer, à se disperser dans l'espace, soit de façon passive, soit de façon active.

L'amplitude écologique est la capacité d'un organisme à occuper une diversité d'habitats. A la capacité de propagation se combine donc la valence écologique pour déterminer l'aire potentielle de répartition d'un organisme vivant, d'une espèce. La valence écologique exprime la tolérance d'une espèce par rapport à un facteur de l'environnement déterminé, étant donné que tout processus biologique participant au développement d'un organisme ne peut s'effectuer que dans des limites de tolérance déterminées pour un facteur donné. Aussi, existe-t-il une limite minimale et une limite maximale entre lesquelles tout processus biologique est possible dans un milieu donné par rapport à un facteur environnemental. Entre ces limites se situe le préférendum ou l'optimum écologique. Pour une espèce donnée, il est possible de définir, en s'appuyant sur les préférences de l'ensemble des paramètres, l'environnement optimal de développement de l'espèce : la niche écologique théorique ou potentielle. Dans la nature, il est toutefois difficile de faire coïncider cette niche potentielle avec la niche réalisée.

Le potentiel évolutif recouvre la plasticité intraspécifique, c'est-à-dire la variabilité des caractères plus ou moins réversibles à l'intérieur des individus d'une même espèce et liés à des conditions du milieu. Il induit tous les phénomènes d'adaptation, de sélection, de mutation, d'hybridation observables à l'intérieur d'une espèce et qui permettent de s'adapter efficacement à la pression exercée par les facteurs environnementaux. Le potentiel évolutif tient donc compte des phénomènes de sélection et des stratégies adaptatives.

Le milieu constitue un facteur externe dans la distribution des espèces à la surface de la terre à travers les conditions qu'il offre. Pour l'essentiel, les principaux facteurs externes sont constitués, des facteurs écologiques que l'on définit comme « tout élément du milieu, biotique ou abiotique, susceptible d'agir sur les êtres vivants, au moins durant une phase de leur cycle de développement » (Dajoz, 2006). Les facteurs écologiques peuvent être regroupés en trois types : les facteurs abiotiques, les facteurs biotiques et les facteurs historiques.

Les facteurs abiotiques sont l'ensemble des facteurs écologiques non liés à un organisme vivant, comme le climat, le relief, les sols, etc. A l'opposé, les facteurs biotiques relèvent du vivant. C'est par exemple la compétition pour les ressources (eau, air, lumière, etc.), les phénomènes de prédation, etc. La distribution actuelle ne peut toutefois s'expliquer entièrement que par des facteurs actuels. Le

recours à des faits historiques peut concourir à l'explication des distributions actuelles des êtres vivants à la surface de la terre.

De la synthèse sur les facteurs internes et externes de la distribution des espèces à la surface de la terre émergent deux lois : la loi du minimum de Liebig (1840)²⁹ et la loi de tolérance de Shelford (1911)³⁰. Celles-ci supportent la conception théorique en autécologie, l'étude des rapports entre une espèce et l'environnement.

Par généralisation de la loi du minimum de Liebig, un facteur écologique est une contrainte pour le développement de l'espèce lorsqu'il est absent ou réduit au-dessous d'un seuil critique ou bien s'il excède une limite maximale.

La loi de Shelford dit que chaque être vivant présente vis-à-vis des divers facteurs écologiques des limites de tolérance entre lesquelles se situe un optimum. Cette loi a trois pour corollaires :

1. une espèce a des fréquences réduites ou est complètement absente dans un milieu lorsque les facteurs écologiques de celui-ci ont des intensités proches des limites de tolérance ou lorsque les limites de tolérance sont dépassées ;
2. en réaction aux facteurs écologiques défavorables, les espèces peuvent modifier leur cycle de développement ;
3. à l'intérieur d'une même espèce, émergent des stratégies adaptatives qui s'expriment en termes d'accommodation, de mutation, d'écotypage ou de spéciation.

Ces lois contribuent, toutes deux, à donner un contenu opératoire au concept de facteur limitant qui désigne l'ensemble des facteurs écologiques qui constituent vis-à-vis d'une espèce ou d'un taxon donné, un frein, une contrainte au développement de l'espèce ou du taxon et jouent donc à la fois sur les fréquences de population et sur les stratégies adaptatives.

Nous retenons *An. gambiae* forme moléculaire M et *An. gambiae* forme moléculaire S comme des espèces ayant chacune leurs exigences écologiques propres et le territoire du Burkina Faso comme milieu ayant une variété de conditions environnementales. L'objectif étant de modéliser la distribution spatiale d'*An. gambiae* forme moléculaire M et d'*An. gambiae* forme moléculaire S, nous nous appuyons sur l'étude des relations qui existent entre ces espèces. Chaque forme est étudiée

²⁹ « La croissance d'un végétal n'est possible que dans la mesure où tous les éléments indispensables pour l'assurer sont présents en quantités suffisantes dans le sol. Ce sont les éléments déficitaires, proches du minimum limitant, qui conditionnent la production des cultures » http://agora.qc.ca/mot.nsf/Dossiers/Justus_von_Liebig ; http://fr.wikipedia.org/wiki/Justus_von_Liebig

³⁰ « Pour tout facteur de l'environnement existe un domaine de valeurs (ou intervalle de tolérance) dans lequel tout processus écologique sous la dépendance de ce facteur pourra s'effectuer normalement »

dans ses relations avec différents paramètres du milieu. La biogéographie, sous ses aspects autoécologiques, est en phase avec la problématique et sert de cadre théorique et thématique.

En adéquation avec ce paradigme de la biogéographie et l'objectif de cette thèse, plusieurs hypothèses peuvent être formulées.

I.2.4.3 Les hypothèses de la recherche

La première hypothèse, qui émane du paradigme biogéographique et qui est pertinente, est qu'il existe une structure dans la distribution spatiale des membres du complexe *An. gambiae* au Burkina Faso. Il est possible d'observer, dans ce pays, une variation de la densité des différents membres du complexe en fonction des conditions environnementales. Les populations des différents membres du complexe sont d'autant plus élevées dans un milieu que celui-ci présente des conditions proches de leur niche écologique théorique. Elles satisfont alors aux exigences écologiques des formes moléculaires, conformément aux lois de Schelford et de Liebig.

La deuxième hypothèse est que tous les facteurs écologiques ne contribuent pas de la même façon à la détermination de cette structure spatiale observée. Certains facteurs environnementaux peuvent devenir plus facilement que d'autres, une contrainte importante, des facteurs limitant pour le développement et la propagation de l'espèce.

De la troisième hypothèse, nous retiendrons qu'il est possible d'exprimer les fréquences de population en fonction des paramètres les plus déterminants dans la distribution spatiale des membres du complexe *An. gambiae*.

Soient $P_{(m)}$, $P_{(s)}$ les fréquences de population respectivement d'*An. gambiae* de forme moléculaire M et de forme moléculaire S.

Soient E_m et E_s , les facteurs environnementaux déterminants de la distribution respectivement des formes moléculaires M et de S d'*An. gambiae* s.s. ;

■ $P_{(m)} = f(E_m)$

■ $P_{(s)} = f(E_s)$

■ $E_s = - E_m$

Cette hypothèse signifie qu'il est possible, connaissant certains paramètres d'un milieu, d'anticiper sur les fréquences de population de forme moléculaire M et S dans ce milieu. Elle suppose également

une opposition en terme de déterminants environnementaux entre les formes moléculaires M et S. Pour tester ces hypothèses, un certain nombre d'indicateurs ont été retenus.

I.2.4.4 Les indicateurs retenus pour la vérification des hypothèses de recherche

La vérification de la première hypothèse s'appuie sur deux indicateurs géostatistiques utilisés en analyse spatiale :

1. l'ellipse de déviation standard. C'est l'expression sous la forme d'une ellipse de la dispersion des points autour du centre moyen correspondant soit à un centroïde, soit à un barycentre. Il permet de mesurer si un ensemble d'objets géographiques ont une tendance directionnelle dans l'espace. *« The ellipse allows you to see if the distribution of features is elongated, and hence has a particular orientation. You can calculate the standard deviational ellipse using either the locations of the features or using the locations influenced by an attribute value associated with the features. The latter is termed a weighted standard deviational ellipse ».*³¹
2. Autocorrélation spatiale et les indices locaux d'association spatiale (LISA) :
 - a. L'autocorrélation spatiale permet de mesurer l'influence que la proximité géographique peut avoir sur les valeurs attributaires observées sur des entités spatiales, le degré de dépendance spatiale d'un processus ponctuel Zaninetti, (2005). Elle vérifie la loi de la proximité de Tobler d'après laquelle, toutes choses égales par ailleurs, les points voisins ont une propension plus grande à avoir des valeurs plus proches que les points éloignés. Ce test est un préalable à l'utilisation d'un certain nombre d'indices. Une distribution autocorrélée présentera un biais dans l'estimation et l'interprétation d'un certain nombre d'indices et de tests statistiques. C'est pourquoi, il est indispensable de tester l'autocorrélation Aubry, (2000). Ce test est fait dans cette thèse à partir de la méthode de concentration des valeurs « faible/forte (Getis – Ord Générale G) » et l'indice de Moran au seuil de signification de 95%.
 - b. Les indices locaux d'association spatiale (LISA) servent à identifier les agrégats ou associations locales et les singularités, ainsi que les configurations spatiales locales d'un semi de points marqués d'un caractère quantitatif. Ils mettent en exergue l'existence d'une configuration spatiale dans la distribution des fréquences d'observation de M et de S au Burkina Faso. Les indices retenus dans le cadre de cette thèse sont : l'indice Getis/Ord Gi* et l'indice I d'Anselin Local Moran.

³¹ ESRI, ArcGIS, Help for spatial statistics toolbox

L'existence d'une structure spatiale dans la distribution des formes moléculaires M et S d'*An. gambiae* s.s., en tant qu'hypothèse, repose sur la mise en rapport de la diversité des milieux et des exigences écologiques. Aussi, une confirmation de cette hypothèse renvoie-t-elle à une interrogation sur la caractérisation du milieu qui met en rapport facteurs écologiques et exigences des populations d'*An. gambiae* forme moléculaire M et forme S, essence même de la deuxième hypothèse. La question qui se pose est : quels sont les paramètres qui déterminent la distribution spatiale de la population d'*An. gambiae* de forme moléculaire M et S ? Ces paramètres s'entendent alors comme facteurs écologiques de distribution des espèces cibles.

Il s'agit moins d'identifier la structure des données relatives à la fois aux populations d'*An. gambiae* forme moléculaire M et S que de mesurer l'impact des facteurs environnementaux sur leur distribution spatiale. Pour cela, la statistique exploratoire multidimensionnelle est mise à contribution sous sa composante analyse des axes principaux. Compte tenu des objectifs, de la structure et du type des données, l'analyse en composantes principales est retenue comme méthode d'analyse des données. En sortie, cette analyse met en exergue les corrélations entre les variables, les axes de ressemblance et de dissemblance. Elle servira de première base de tri de variables pertinentes entrant dans la définition de la niche écologique réalisée pour chaque forme moléculaire M et S *An. gambiae* s.s.. Elle ouvre vers la modélisation de la distribution spatiale des membres du complexe.

L'analyse en composantes principales a permis de révéler la structure des données. L'hypothèse relative à la mise en relation des facteurs environnementaux et la répartition spatiale des populations des formes moléculaires *An. gambiae* s.s. est plus envisageable. La régression ou l'analyse de la covariance (ANCOVA) sont des techniques statistiques appropriées pour cela. La régression met en équation les fréquences transformées des formes moléculaires prises individuellement et les facteurs environnementaux. Elle permet d'évaluer la possibilité d'une prédiction des fréquences en fonction d'un ou de plusieurs paramètres environnementaux. Le seuil de confiance retenu pour la régression est de 95%. L'application d'une analyse de la covariance est indispensable si une ou plusieurs variables nominales se révèlent déterminantes dans la distribution de ces vecteurs du paludisme.

Cette étude vise donc trois objectifs : la description de la distribution spatiale des formes moléculaires M et S d'*An. gambiae* s.s., l'explication de cette distribution et sa formalisation. Elle s'appuie sur la biogéographie qui fournit les hypothèses. L'analyse spatiale et l'analyse de données sont retenues comme méthodes de traitement de l'information pour la vérification des hypothèses.

CONCLUSION DU CHAPTIRE I

De ce chapitre, nous retenons que la géographie, de par son objet, son approche et ses outils apporte une contribution indispensable à la connaissance des problèmes de santé. Ce travail s'insère dans cette dynamique d'ouverture de la géographie sur les problématiques de santé en abordant la distribution spatiale des vecteurs de maladie en l'occurrence les formes moléculaires M et S, vecteurs principaux du paludisme au Burkina Faso. Cela va consister à formaliser, à travers des modèles, les relations entre ces deux formes moléculaires et l'environnement / l'espace. Le paradigme biogéographique s'avère efficace à fournir un cadre théorique et des hypothèses éprouvées afin de fédérer les contributions des autres disciplines scientifiques et techniques pour atteindre les objectifs de description, d'identification des déterminants environnementaux et de modélisation de la distribution spatiale des formes moléculaires M et S d'*An. gambiae* au Burkina Faso.

Ce chapitre a ainsi permis de cadrer ce travail de recherche tant au plan conceptuel qu'au plan théorique. Toutefois, la délimitation du thème resterait incomplète tant que le champ d'application n'aurait pas été défini. C'est l'objet du chapitre suivant consacré à la présentation de l'anophèle et de son environnement, le Burkina Faso.

CHAPITRE II : BIO-ÉCOLOGIE D'*An. gambiae* AU BURKINA FASO : UNE RELATION A FORMALISER

« La modélisation en géographie commence par deux questions inévitables et associées : où sommes-nous ? Qu'est-ce qui a des chances d'être en jeu ? » Brunet, (2000a). Ce chapitre vise à répondre à ces deux questions. Son objectif est de présenter le champ d'application de ce travail de recherche. La connaissance du champ d'application (l'anophèle et le Burkina) est indispensable non seulement à la conception d'une méthodologie opérationnelle mais aussi à l'interprétation des données. Parce que le paludisme confère à cette étude un caractère utilitaire par son engagement dans une problématique d'actualité sur les questions de santé, nous y donnons également une présentation de cette maladie. Ce chapitre est composé de trois sections : la première est un exposé sur le paludisme dont les deux formes moléculaire M et S d'*An. gambiae* sont vecteurs, la seconde est consacrée à la bioécologie des anophèles, enfin, la troisième fait une présentation du Burkina Faso, la zone d'étude.

II.1 LE PALUDISME UNE MALADIE DE SANTE PUBLIQUE DONT L'ANOPHELE EST LE VECTEUR

Le Paludisme (ou *malaria*) est une maladie endémique qui sévit de façon continue dans les régions tropicales et subtropicales de l'Afrique, de l'Asie du sud et du sud-est et de l'Amérique du sud. Elle doit son nom au mot Latin *Palus* (= marais) ou au mot Italien *malaria* (= mauvais air). C'est une maladie infectieuse due à des parasites protozoaires du genre *Plasmodium*. Ces parasites sont transmis d'un individu à l'autre par les femelles des moustiques du genre *Anopheles*. Il existe quatre espèces de *Plasmodium* capables d'infecter l'Homme : *Plasmodium falciparum*, Welch, (1897), *Plasmodium vivax*, Grassi et Felletti, (1890), *Plasmodium malariae* (Laveran, 1881) et *Plasmodium ovale* Stephens, (1922).

II.1.1 LE PALUDISME DANS LE MONDE ET AU BURKINA FASO

Le Paludisme est la maladie la plus répandue dans le monde. Selon l'Organisation Mondiale de la Santé (O.M.S), 102 pays y sont exposés. Plus de 40% de la population mondiale, soit plus de deux milliards de personnes, vivent dans des régions à risque. Quelques 300 millions de cas cliniques de paludisme sont enregistrés chaque année dans le monde et entraînent près de 1 à 1,5 millions de décès par an. Près de 90% de ces décès sont enregistrés en Afrique subsaharienne où les plus touchés sont les enfants de 0 à 5 ans et les femmes enceintes OMS, (1999).

Au Burkina Faso, le paludisme est la première cause de consultation, d'hospitalisation et de décès dans les formations sanitaires. Il représente plus de trois millions de cas cliniques avec

15 000 décès déclarés chaque année surtout parmi les enfants de 0 à 5 ans CNLP (1994). D'après la Banque Mondiale³² citant le Programme national de lutte contre le paludisme du Burkina (PNLP), environ 5 450 000 cas de paludisme ont été enregistrés, en 2007, dans l'ensemble des formations sanitaires du pays. 12 000 cas de décès sont recensés sur ce nombre soit un taux de létalité de 2 %. Chez les enfants de 0 à 5 ans qui ne représentent que 19% de la population, 2 615 000 cas de paludisme ont été enregistrés. En moyenne, chaque enfant de cette tranche d'âge a consulté un médecin ou un agent de santé au moins une fois pour le paludisme. Le total des cas de paludisme enregistrés chez les femmes enceintes est de 154 000 et 94 en sont décédées. Plus de la moitié des cas et des décès par paludisme surviennent pendant la saison pluvieuse, entre août et novembre (correspondant à la saison de forte transmission de la maladie).

Selon le Programme National de lutte contre le Paludisme (PNLP, 2007)³³ au Burkina Faso, cette maladie reste une endémie stable dans tout le pays, avec une recrudescence saisonnière (mai à octobre). Selon les données statistiques du système national d'information sanitaire (SNIS), le paludisme est la première cause de consultation (35%), d'hospitalisation (41%) et de décès (37%) en 2005. Les enfants de moins de 5 ans restent les plus touchés durant l'année 2005. Dans cette tranche d'âge, le paludisme représente 45% des motifs de consultation, 55% des hospitalisations et 57% des décès.

II.1.2 LA LUTTE CONTRE LE PALUDISME

Au Burkina Faso, le plan stratégique 2006-2010 de lutte contre le paludisme, estimé près de 45 milliards de francs CFA, se résume en 13 points :

1. rendre disponible et accessible le traitement aux Combinaisons Thérapeutiques à base d'Artémisinine (CTA) au niveau des formations sanitaires et de la communauté ;
2. promouvoir les Interventions à Base Communautaire (IBC) par la clarification du statut et rôle des relais communautaires ;
3. rendre accessible le traitement du paludisme grave dans les centres de référence ;
4. accélérer le passage à l'échelle de la Prise en Charge Intégrée des Maladies de l'Enfant (PCIME) ;
5. promouvoir la lutte intégrée contre les vecteurs (LIV) ;

³² <http://web.worldbank.org/WBSITE/EXTERNAL/ACCUEILEXTN/PAYSEXTN/AFRICAINFRENCHEXT/BURKINAFASOINFRENCHEXTN/0,,contentMDK:21744044~pagePK:1497618~piPK:217854~theSitePK:463455,00.html>

³³ PNL, 2007, Plan stratégique de lutte contre le paludisme 2006-2010, <http://209.85.129.132/search?q=cache:T1SAeMEeo0YJ:www.rollbackmalaria.org/countryaction/nsp/burkina.pdf+PNLP+Burkina+p%C3%A9riode+de+transmission&cd=7&hl=fr&ct=clnk&gl=fr&client=firefox-a>

6. rendre disponibles et accessibles les moustiquaires traitées aux insecticides (MTI) à travers différents canaux (Secteur public, privé) ;
7. mettre à l'échelle le traitement préventif intermittent (TPI) chez la femme enceinte ;
8. renforcer le plaidoyer, la mobilisation sociale et la communication pour le changement de comportement CNLP, (2007) ;
9. renforcer les capacités institutionnelles du PNLP ;
10. renforcer le partenariat et la collaboration multisectorielle dans la lutte contre le paludisme;
11. orienter les domaines de recherche adaptés aux besoins de gestion du programme ;
12. contribuer à la réalisation des travaux de recherche sur le paludisme ;
13. renforcer le système de suivi évaluation.

La lutte antipaludique a été définie comme l'ensemble des mesures destinées à supprimer, ou tout au moins à réduire, la mortalité et la morbidité dues au paludisme O.M.S, (1974). Elle comporte des actions curatives fondées sur la chimiothérapie des malades, et des actions préventives basées sur la chimioprophylaxie et/ou la réduction du contact homme/vecteur.

II.1.2.1 La lutte contre le Parasite

La lutte contre le parasite repose sur la chimiothérapie des malades et la chimioprophylaxie pour les groupes à risques (femme enceinte, enfant de moins de 5 ans et personnes provenant de zones non impaludées). Cette chimiothérapie connaît des limites aujourd'hui à cause de l'apparition de souches de *Plasmodium* résistantes aux antipaludiques Charmot, (1987); Wernsdorfer, (1991).

II.1.2.2 La lutte contre les vecteurs

La lutte anti-vectorielle repose actuellement sur deux types de méthodes à savoir : les méthodes de lutte anti-larvaire et celles contre les adultes. La lutte anti-larvaire consiste à réduire les populations d'anophèles à leur source de production. Trois méthodes sont couramment utilisées :

- les méthodes physiques, par suppression des gîtes larvaires Mouchet et Carnevale, (1991) ;
- les méthodes chimiques, par utilisation de substances chimiques (huile, vert de Paris ou acetoarsenite de cuivre) qui détruisent les larves Service, (1993) ;
- les méthodes biologiques, par élevage de poissons larvivores (*Gambusia affinis*, *Poecillia reticulata*) ou par épandage dans les gîtes, de bactéries (*Bacillus thuringiensis israeliensis*, *Bacillus sphaericus*) empêchant le développement des larves Lacey et al., (1984).

Les méthodes de lutte contre les anophèles adultes reposent sur l'utilisation de barrières physiques, de répulsifs ou d'insecticides Diabaté, (1999); Ilboudo-Sanogo, (2001). L'usage des matériaux imprégnés d'insecticides pose la question de la conséquence de la réduction de la transmission du *Plasmodium* sur l'immunité des enfants en bas âge dans les zones hyper-endémiques Binka et al., (1996); Nébié, (2002); Snow et Marsh, (1995). Tous ces problèmes font qu'il existe aujourd'hui peu d'outils efficaces pour lutter contre le paludisme, d'où la nécessité d'envisager d'autres méthodes de lutte en attendant la mise en place d'un vaccin.

II.1.2.3 Les perspectives de la lutte contre le paludisme

Face aux doutes de la communauté scientifique vis à vis de l'effet des matériaux imprégnés d'insecticides Binka et al., (1996); Nébié, (2002); Quinones et al., (1997); Snow et Marsh, (1995); Trape et Roger, (1996) et dans l'attente d'un vaccin efficace et financièrement accessible, il apparaît nécessaire de développer d'autres méthodes de contrôle vectoriel. Parmi ces méthodes, la lutte génétique apparaît comme une alternative intéressante.

La lutte génétique envisagée est le remplacement éventuel des populations vectrices sauvages par d'autres trans-géniques incapables de transmettre le parasite par l'introduction de gènes de résistance au *Plasmodium* au sein de la population vectorielle Curtis, (1996); Carlson et al., (1995). Avec le séquençage du génome d'*Anopheles gambiae* publié en 2002 Holt et al., (2002), il devient possible de développer de nouvelles cibles pour le contrôle de la transmission.

Quelle que soit la stratégie de lutte anti-vectorielle adoptée, la caractérisation exacte des espèces, la connaissance de leur bio-écologie, du flux génétique entre elles et leur répartition spatio-temporelle constituent un préalable si l'on veut rendre la lutte plus sélective et plus efficace.

II.1.3 LE CYCLE DE TRANSMISSION DU PALUDISME

Le cycle de développement des *Plasmodium* comporte deux phases : une chez le moustique et une autre chez l'Homme (figure 1).

Chez l'homme, par sa piqure, la femelle de l'anophèle infestée inocule des centaines de parasites (les sporozoïtes) qui gagnent rapidement le foie où ils se multiplient. Au bout de 9 à 18 jours selon l'espèce de *Plasmodium*, les parasites qui se trouvent sous une forme appelée mérozoïte, passent dans le sang, envahissent les hématies et se multiplient à nouveau. Après un certain nombre de multiplications, certains mérozoïtes se transforment en gamétocytes mâles et femelles.

Le moustique s'infeste en piquant une personne dont le sang contient des gamétocytes. Le développement des parasites dans l'organisme de l'insecte aboutit à une nouvelle génération de parasites, les sporozoïtes qui migrent vers les glandes salivaires où ils demeurent jusqu'à ce que l'insecte pique de nouveau un individu. Et un nouveau cycle reprend ainsi.

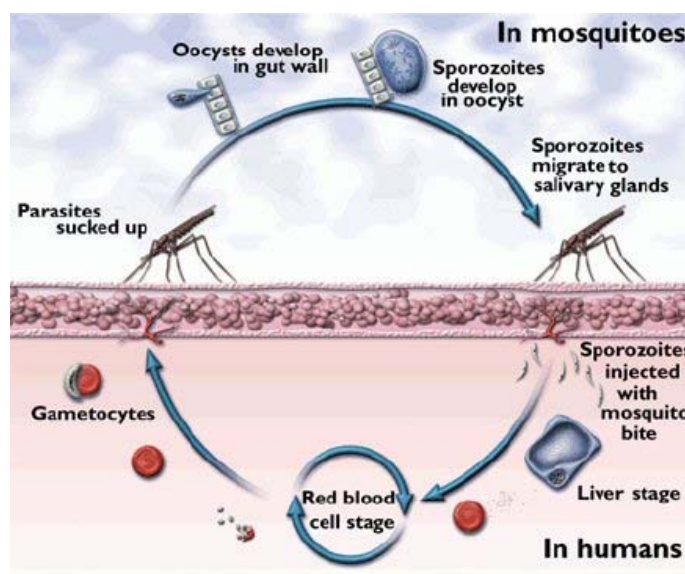


Figure 1 : cycle du plasmodium, <http://www.traveldocotor.co.uk/malaria.htm>

II.2 BIO-ÉCOLOGIE D'*Anopheles gambiae*

Connaître la bio-écologie d'une espèce est une condition *sine qua non* à l'étude de sa répartition dans un espace donné. Il est difficile de saisir les rapports entre un milieu et un organisme sans connaître l'offre écologique de ce milieu et les exigences vitales de l'espèce. Ce paragraphe est un exposé sur la bio-écologie d'*Anopheles gambiae*. Même si la connaissance de ce complexe d'espèces par rapports aux différents facteurs écologiques pris individuellement n'est pas toujours très précise, une synthèse de celle-ci constitue une avancée significative dans la conduite de ce projet.

Ce point, consacré à la bio-écologie des anophèles, s'articule autour de la position zoologique de celles-ci, de leur cycle de développement et de leur écologie à différents stades de développement.

II.2.1 LA POSITION ZOOLOGIQUE

Les anophèles sont des Arthropodes appartenant à la classe des Insectes, à l'ordre des Diptères, du sous-ordre des Nématocères et à la grande famille des *Culicidae*. Cette famille compte plus de 3 450 espèces et sous-espèces regroupées en trois sous-familles : la sous-famille des *Toxorhynchitinae*, la sous-famille des *Anophelinae* et celle des *Culicinae*. Parmi ces sous-familles, celle des *Anophelinae*

est constituée de 3 genres (*Bironella* ; *Chagasia* ; *Anopheles*). Le genre *Anopheles* est le plus important sur le plan médical. Il regroupe 400 espèces dont 70 sont des vecteurs de maladies humaines Bruce-Chwatt et Zulueta, (1985).

En Afrique subsaharienne, ce sont les membres du complexe³⁴ *Anopheles gambiae* et *Anopheles funestus* qui sont les vecteurs majeurs du paludisme. Certaines observations faites sur l'écologie des vecteurs ont mis en exergue une hétérogénéité au sein de cette population *Anopheles gambiae*. Les investigations ont révélé plus tard que *Anopheles gambiae* est un complexe d'espèces.

II.2.2 PRESENTATION DU COMPLEXE *Anopheles gambiae*

La plupart des espèces d'anophèle ont été décrites et nommées suivant le « concept morphologique de l'espèce ». Avec le développement de la génétique, le concept moderne de l'espèce est essentiellement basé sur la mise en évidence de barrière de reproduction entre spécimens. Un complexe d'espèces est un ensemble d'espèces non identifiables morphologiquement et caractérisées par un isolement reproductif. Le complexe *An. gambiae* est le plus étudié au monde Mouchet et al., (2004).

Les premières indications d'une hétérogénéité au sein des populations d'*An. gambiae* furent en effet basées sur des considérations purement écologiques avec la description de «variants» capables de coloniser des gîtes larvaires saumâtres Muirhead-Thomson, (1951). Sur la base de croisements effectués en laboratoire, ces variants se sont avérés partiellement isolés des souches inféodées aux gîtes d'eau douce, elles-mêmes formant trois groupes de croisements distincts Davidson, (1964); Davidson et Jakson, (1962); Paterson, (1964). Les travaux de Coluzzi en cytogénétique Coluzzi et Sabatini, (1967), (1968a, b), ont permis une avancée considérable des moyens d'identification des différentes espèces constitutives du complexe *An. gambiae*, basée sur l'existence d'inversions chromosomiques polymorphes mises en évidence par coloration des chromosomes géants isolés des glandes salivaires des larves au stade IV ou des ovaires des femelles semi-gravides Green, (1972). Les progrès de la biologie moléculaire ont alors permis le développement de sondes ADN spécifiques Collins et al., (1987); Crampton et Hill, (1993) qui ont débouché sur la mise au point d'un test diagnostique basé sur l'amplification par PCR des gènes codant pour les ARN ribosomaux Scott et al., (1993). Ce test permet l'identification rapide des spécimens collectés sur le terrain, quels que soient le sexe et le stade de leur développement. Il nécessite très peu de matériel de départ. Son

³⁴ Un complexe d'espèces est un ensemble d'espèces jumelles parce que morphologiquement très voisines, voire indifférentiables, dont l'identification demande l'utilisation de critères mixiologiques, cytogénétique, biochimique ou moléculaire. Chaque espèce d'un complexe présente des comportements, une écologie et des compétences vectorielles spécifiques. (J. Mouchet et al. 2004)

efficacité a été grandement démontrée Fontenille et *al.*, (1993) et il est maintenant pratiqué en routine dans de nombreux laboratoires.

Le complexe *An. gambiae* comprend aujourd'hui sept espèces bien différenciées :

- *An. gambiae* s.s. (sens strict) Giles, (1902);
- *An. arabiensis* Patton, (1904);
- *An. bwambae* White, (1985);
- *An. melas* Theobald, (1903);
- *An. merus* Doenitz, (1902);
- *An. quadriannulatus* Theobald, (1911);
- *An. quadriannulatus* B, dénomination provisoire de la septième espèce du complexe identifiée très récemment en Ethiopie Hunt et *al.*, (1998).

An. gambiae s.s., *An. arabiensis*, sont les membres du complexe présents au Burkina Faso Petrarca et *al.*, (1986). *An. gambiae* s.s. présente deux formes chromosomiques appelées Mopti et Savane, caractérisées par une limitation de flux génique. Ces deux formes correspondent respectivement à deux profils moléculaires dénommés : forme moléculaire M et forme moléculaire S Favia et *al.*, (1997); Favia et *al.*, (2001). Le cycle de développement de ces formes moléculaires n'a pas été particulièrement décrit. Aussi, retient-on pour leur cycle de développement, celui des anophèles.

II.2.3 LE CYCLE DE DEVELOPPEMENT DES ANOPHELES

Les anophèles sont des holométaboles³⁵ qui présentent quatre stades au cours de leur vie. Ces stades constituent le cycle de développement. Ce cycle de développement des anophèles est marqué par deux phases : une phase aquatique et une phase aérienne. La phase aquatique regroupe les trois premiers stades. Elle est appelée aussi phase pré-imaginale. Elle va de l'œuf à la nymphe via les stades larvaires. La phase aérienne concerne l'adulte ailé ou imago (Figure 2).

³⁵ Holométaboles : A métamorphose complète.

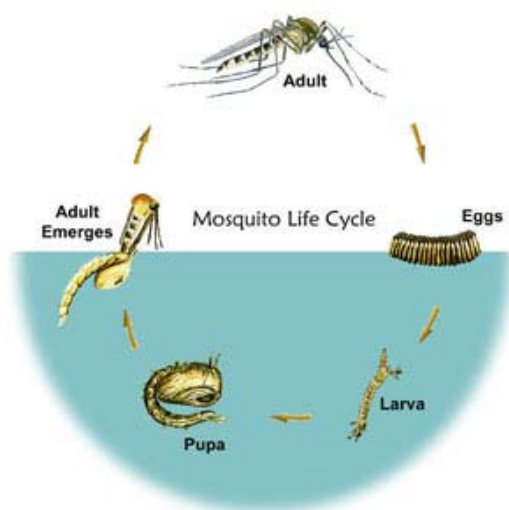


Figure 2 : Cycle de développement de l'Anophèle In Yé (2005)³⁶

II.2.3.1 La phase de vie aquatique

Cette phase débute avec la ponte des œufs à la surface de l'eau. Les œufs au nombre de 30 à 300 par ponte sont déposés un à un à la surface de l'eau et flottent jusqu'à l'éclosion. Sous les tropiques, les œufs éclosent habituellement 2 à 3 jours après la ponte Rodhain et Perez, (1985). La larve émerge en utilisant une « dent d'œuf » placée en position postéro-dorsale sur sa tête. Elle possède les trois tagmes caractéristiques des insectes, c'est à dire la tête, le thorax et l'abdomen. La larve est dépourvue de pattes mais possède de nombreuses soies qui recouvrent tout le corps. Elle est capable de nager dans l'eau grâce à des mouvements ondulants de son corps. Elle respire l'air extérieur par des stigmates dorsaux situés à l'extrémité de l'abdomen, ce qui lui impose une position de repos parallèle à la surface de l'eau.

Les larves ne se développent pas en continu, elles passent par quatre stades différents. La petite larve qui émerge de l'œuf est appelée larve de premier stade ou larve de stade I. Elle mesure environ 1,5 mm de long et subit trois mues successives qui la transforment en larve de stade II, III et enfin de stade IV. La larve de 4^e stade mesure environ 8 à 10 mm de long. La vie larvaire dure de 8 à 12 jours dans des conditions favorables Rodhain et Perez, (1985).

La larve de quatrième stade, au terme de son développement, se transforme en une nymphe, dont la forme générale est celle d'une virgule. La nymphe ne se nourrit pas, elle représente un stade intermédiaire. Elle subit de très profondes transformations morphologiques et physiologiques pour

³⁶ Yazoumé Yé, 2005, Incorporating environmental factors in modelling malaria transmission in under five children in rural Burkina Faso, doctor scientiarum humanarum, ruprecht-karls-university Heidelberg.

donner un moustique adulte aérien après une période de 1 à 4 jours sous les tropiques Rodhain et Perez, (1985).

II.2.3.2 La phase de vie aérienne

L'adulte ou imago qui émerge de la nymphe est un insecte au corps long et grêle, doté de longues pattes et de longues pièces buccales en forme d'aiguille. Peu après leur émergence, les adultes (mâles et femelles) s'envolent et les femelles s'accouplent aussitôt avec des mâles plus âgés qui essaient au-dessus de l'eau. En général, les femelles ne s'accouplent qu'une seule fois dans leur vie, mais pondent périodiquement pendant toute leur existence. Les spermatozoïdes sont stockés dans une spermathèque où ils conservent leur pouvoir fécondant jusqu'à la mort de la femelle Bruce-Chwatt et Zulueta, (1985).

Après 24 à 48 heures de vie adulte, la femelle va à la recherche de son premier repas sanguin. Elle a besoin de sang comme source de protéines pour le développement et la maturation de ses œufs. Dès la fin du repas, commence la digestion et simultanément les ovaires se développent. Les femelles gravides se mettent en quête d'un lieu favorable à la ponte. La fécondation de l'ovule a lieu lors de la ponte. Une fois qu'elles ont déposé leurs œufs, elles retournent prendre un autre repas de sang pour aller pondre à nouveau et le cycle reprend. La longévité de l'adulte est estimée à une vingtaine de jours, guère plus d'un mois dans les conditions habituelles Robert et *al.*, (1989). La période allant du repas sanguin à la ponte est appelée cycle gonotrophique. Elle dure environ 48 heures dans les conditions de températures comprises entre 23°C et 25°C Bruce-Chwatt et Zulueta, (1985).

Le cycle gonotrophique comprend trois phases :

- la recherche de l'hôte et la prise du repas de sang,
- la digestion du sang et la maturation ovarienne,
- la recherche du lieu de ponte par la femelle gravide.

La ponte a lieu généralement au crépuscule. Les caractéristiques physico-chimiques de l'eau des gîtes sont analysées par des chémorécepteurs situés sur les pattes.

II.2.4 L'ÉCOLOGIE DES ADULTES

Les anophèles sont actifs du coucher au lever du soleil. Toutefois, les heures d'agressivité maximale varient selon les espèces. Il y a également des variations dans les comportements trophiques. Certains moustiques piquent préférentiellement à l'intérieur des habitations, ce sont des endophages, d'autres préfèrent piquer à l'extérieur des habitations, ce sont des exophages.

Les anophèles présentent également des choix pour leur repas sanguin. De nombreuses espèces d'anophèles se nourrissent à la fois sur l'homme et sur l'animal. Certaines espèces prennent leur repas sanguin principalement sur les animaux. Ce sont des zoophiles. D'autres ne se gorgent pratiquement que sur l'homme. Ce sont des anthropophiles. Les espèces anthropophiles ont un taux d'infection plus élevé que les zoophiles White, (1974).

Les vecteurs de paludisme sont inégalement repartis sur les continents. En Afrique subsaharienne, *Anopheles gambiae sensu lato* et *Anopheles funestus* sont les vecteurs majeurs du paludisme Mouchet et Carnevale, (1991). *Anopheles gambiae s.s.* et *An. arabiensis* sont les deux espèces du complexe les plus répandues Touré, (1988). Leurs aires de répartition se chevauchent partiellement : *An.gambiae s.s.* est classiquement adapté aux zones de forêts et de savanes humides alors qu'*An. arabiensis* peuple les environnements plus secs jusqu'en bordure du Sahara Chauvet et *al.*, (1969); Coluzzi et *al.*, 1979; Coz, (1973).

En zone de savane, ils sont souvent en sympatrie et très bien adaptés à l'homme et à son environnement. Ils se comportent comme des espèces endophiles et anthropophiles. C'est auprès de l'homme qu'elles trouvent les gîtes larvaires et ceux de repos qui leur conviennent le mieux Coluzzi, (2002); Gillies et Meillon, (1968). Cependant *An. arabiensis* est plus zoophile et exophile qu'*An.gambiae s.s.* Costantini et *al.*, (1998); Robert et Carnevale, (1984); White, (1974). D'après Rajagopalan et *al.*, (1991), la plupart des anophèles se dispersent dans un rayon d'environ un kilomètre autour de leur gîte. Pour *An. gambiae s.s.*, ce rayon de dispersion autour des gîtes varie en moyenne entre 1 et 1,6 km. Ce rayon est de 0,8 km pour *An. funestus*.

II.2.5 L'ÉCOLOGIE DES LARVES (STADES PRE-IMAGINAUX) : CARACTERISTIQUES DES HABITATS LARVAIRES

Les larves des deux vecteurs majeurs du complexe *An. gambiae* tendent à occuper des gîtes de diverses natures, mais tous sont généralement des gîtes temporaires ou semi permanents ensoleillés comme les flaques d'eau Gillies et Coetzee, (1987); Gillies et Meillon, (1968); Hamon et *al.*, (1956).

Les gîtes de reproduction d'*An. gambiae s.s.* et d'*An. arabiensis* coïncident généralement. Les deux espèces ont été rencontrées ensembles dans les mêmes gîtes au Nigeria Service, (1970b). Cependant, d'autres études ont montré une répartition hétérogène significative entre ces deux espèces Minakawa et *al.*, (1999); Sagnon, (1999); Sagnon et *al.*, (2000); Service et *al.*, (1978); White et Rosen, (1973).

Une hétérogénéité temporelle dans la composition des gîtes a été observée au Mali Sangaré, (2000); Touré et *al.*, (1998). Pour Le-Sueur et Sharp, (1988), les paramètres tels que le pH, l'oxygène dissous, la turbidité n'expliquent pas la distribution des membres du complexe *An. gambiae* d'eau douce au Natal.

D'autres études n'ont pu mettre en évidence une relation entre les facteurs environnementaux et la répartition des membres du complexe *An. gambiae* Minakawa et *al.*, (1999); Robert et *al.*, (1998); Ye-Ebiyo et *al.*, (2003). Pour Minakawa et *al.*, (1999), l'abondance d'*An. gambiae* s.s. serait associée à la proximité des concessions. Koenraadt et *al.*, (2004), ont mis en évidence les effets de la précipitation et de l'évapotranspiration sur la dynamique temporelle des membres du complexe *An. gambiae* au Kenya.

Au Burkina Faso, les larves *An. arabiensis* et *An. gambiae* tendent à occuper des gîtes spécifiques. Selon Sagnon, (1999), il est probable, au regard des statistiques sur la structure de la population larvaire des trois taxa, que *An. gambiae* forme moléculaire M se reproduit préférentiellement dans des gîtes de nature plus permanente. La reproduction dans les gîtes temporaires représente pour cette espèce seulement un deuxième choix, tandis que *An. arabiensis* et *An. gambiae* forme moléculaire S sont plus adaptés à se reproduire dans les gîtes temporaires.

Cette différence dans l'occupation des gîtes par ces trois taxa du complexe *An. gambiae* pourrait s'expliquer par des propriétés de répulsion inter-spécifique des derniers stades larvaires Guelbeogo, (1999); MC-Crae, (1984). Dans tous les cas, l'influence des facteurs environnementaux dans cette répartition n'est pas encore bien connue Charlwood et Edoh, (1996); Sagnon, (1999). La connaissance de la bio-écologie des larves s'avère indispensable si l'on veut élaborer des stratégies efficaces de lutte antivectorielle.

Fort de cette présentation de la bio-écologie du complexe *An. gambiae*, nous pouvons définir son environnement.

II.2.6 L'ENVIRONNEMENT DU COMPLEXE *AN. GAMBIAE*

L'environnement est une notion relative dont la perception effective est fonction d'une entité biologique de référence. C'est à partir des composantes initiales, modifiées par les organismes à mesure de leur implantation, que s'est progressivement constitué « un environnement », propre à l'ensemble biologique localement considéré Lacoste et Salanon, (2006). L'environnement ne peut se définir que par rapport à un organisme vivant. L'environnement d'*An. gambiae* s.s. peut être défini à

partir d'une synthèse de son cycle de vie. Cette synthèse sur la bio-écologie du complexe nous permet de présumer des facteurs environnementaux qui peuvent influencer sa prolifération.

Très souvent, les facteurs climatiques comme les précipitations, les températures, et l'insolation sont considérés. La disponibilité et l'accessibilité en eau sont deux autres paramètres importants parce qu'elles interviennent dans la biologie des larves. La qualité de l'eau (taille, Ph, stabilité, écoulement, etc.) dépend du contexte physique. Ce sont autant d'éléments qui obligent à intégrer des paramètres tels que la topographie (altitude) et la disponibilité de la matière organique. La présence humaine (nécessité d'un repas sanguin pour la maturation des œufs) et la végétation (alimentation en jus sucré, lutte contre les prédateurs) interviennent aussi dans la définition de l'environnement du vecteur.

L'environnement d'*An. gambiae* s.s., simple à première vue, est plus complexe qu'il ne le paraît. Cette complexité est liée à la nécessité d'une variation des exigences en fonction du stade de développement doublée d'une variation en fonction du membre du complexe concerné. Prendre en compte la forte variabilité saisonnière de certains paramètres environnementaux tels que la disponibilité en eau, le climat, conduit à un niveau très élevé de complexité. De plus, la mesure ou l'évaluation objective de certains paramètres ne sont pas évidentes compte tenu de l'extension spatiale de la zone d'étude.

De cette présentation sur les anophèles, nous retenons qu'il n'y a pas un vecteur du paludisme mais plutôt des vecteurs du paludisme au regard de la diversité des espèces qui composent le complexe *An. gambiae*.

Les études sur la biologie et l'écologie des vecteurs du paludisme sont très riches mais non spécifiques aux espèces ciblées dans cette étude. Ainsi, l'état de la connaissance sur la bio-écologie des anophèles est-il ouvert et la définition d'un biotope précis pour les anophèles devient très complexe. A cette diversité spécifique des anophèles, s'ajoutent, pour une même espèce, des exigences écologiques variables en fonction du stade de développement. Le niveau de complexité devient alors très élevé et impose de réduire le champ d'investigation, d'où la limitation aux formes moléculaires S et M. Le contrôle de la population vectorielle ne pouvant se concevoir en dehors de l'environnement de vie de celle-ci, il est indispensable de comprendre les déterminants mésologiques de distribution des formes moléculaires S et M d'*An. gambiae*. C'est l'objet de la section trois de ce chapitre.

II.3 PRESENTATION DE LA ZONE D'ETUDE³⁷

Le Burkina Faso est un pays situé au cœur de l'Afrique occidentale, entre 5°30' de longitude ouest et 2°40' de longitude, est et entre 9°30' et 15°15' de latitude nord. Il est limité au nord et au nord-ouest par le Mali, à l'est par le Niger, au sud-est par le Bénin, au sud par le Togo, le Ghana, et au sud-ouest par la Côte d'Ivoire (Figure 3). Sa superficie³⁸ est de 274 200 km². Quelles sont les caractéristiques de ce pays, tant au plan physique qu'humain, qui pourraient contribuer à mesurer et à expliquer la distribution spatiale des populations des deux formes moléculaires *An. gambiae* s.s. ?

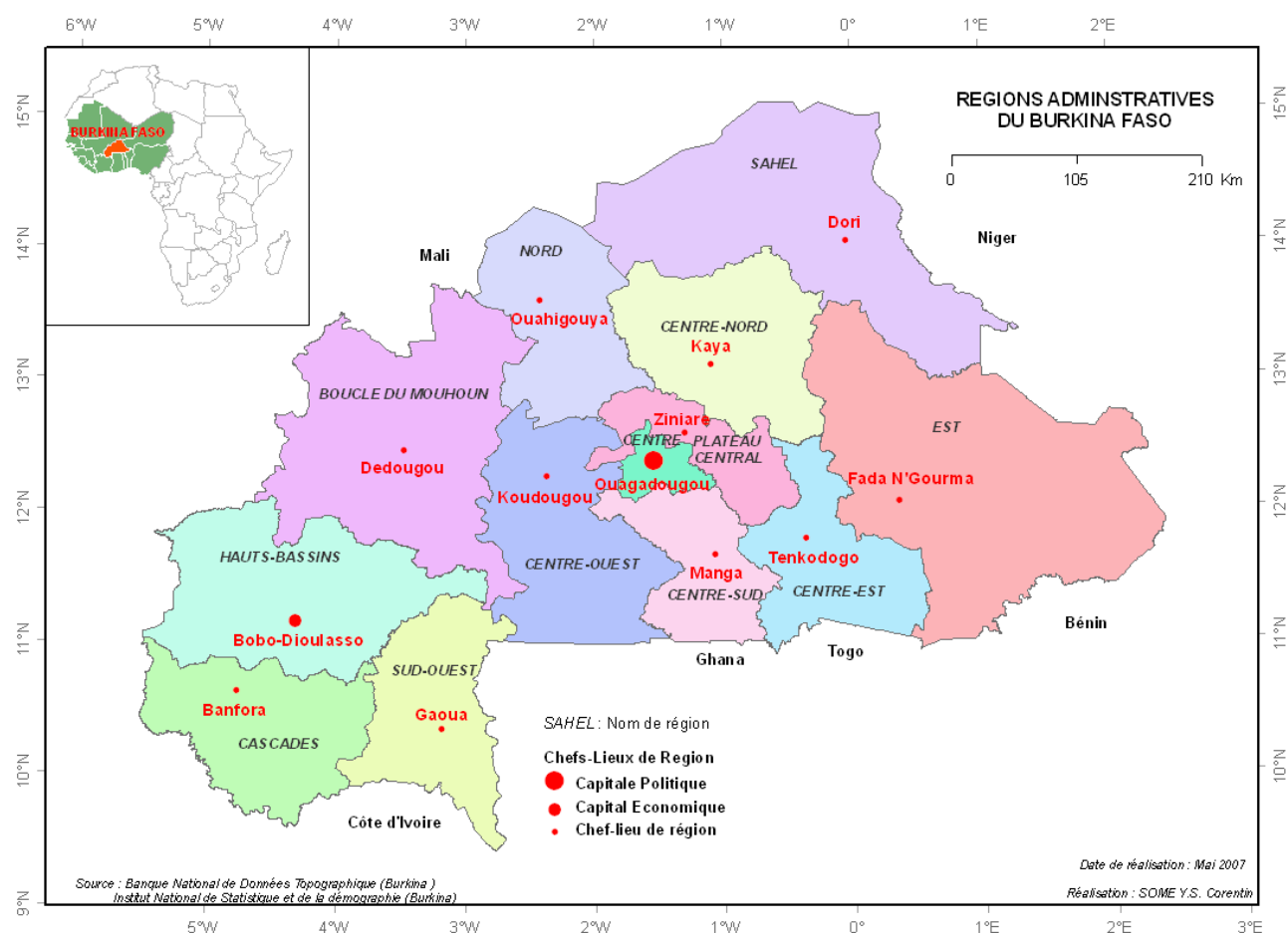


Figure 3 : Situation du Burkina Faso

³⁷ Les éléments utilisés pour la présentation du Burkina Faso, le site de l'étude, ont été retenus sur la base de leur contribution à la compréhension de la répartition des formes moléculaires M et S d'*An. gambiae* (section précédente). Nous avons choisi de faire une présentation très visuelle, graphique parce que nous voulons faire un modèle spatial. Cette forme est plus efficace dans la communication. C'est pourquoi, nous avons préféré utiliser des cartes, des graphiques et des tableaux.

³⁸ <http://www.primature.gov.bf/burkina/infopratique/fsites.htm>
http://www.inforoute-communale.gov.bf/col_cadrel.htm

Le Burkina Faso est composé de 8 200 villages³⁹ et secteurs repartis en 351 communes/départements, 45 provinces et 13 régions qui sont des circonscriptions administratives et des collectivités territoriales décentralisées.

II.3.1 LA POPULATION

La population du Burkina Faso est de 13 117 100 habitants⁴⁰. Cette population est composée de plus de 72 ethnies classées en 7 groupes linguistiques : langues Berbère, Dogon, Kru, Ouest Atlantique, Gur ou Voltaïque, Manden, Nilo Saharien (Figure 4)

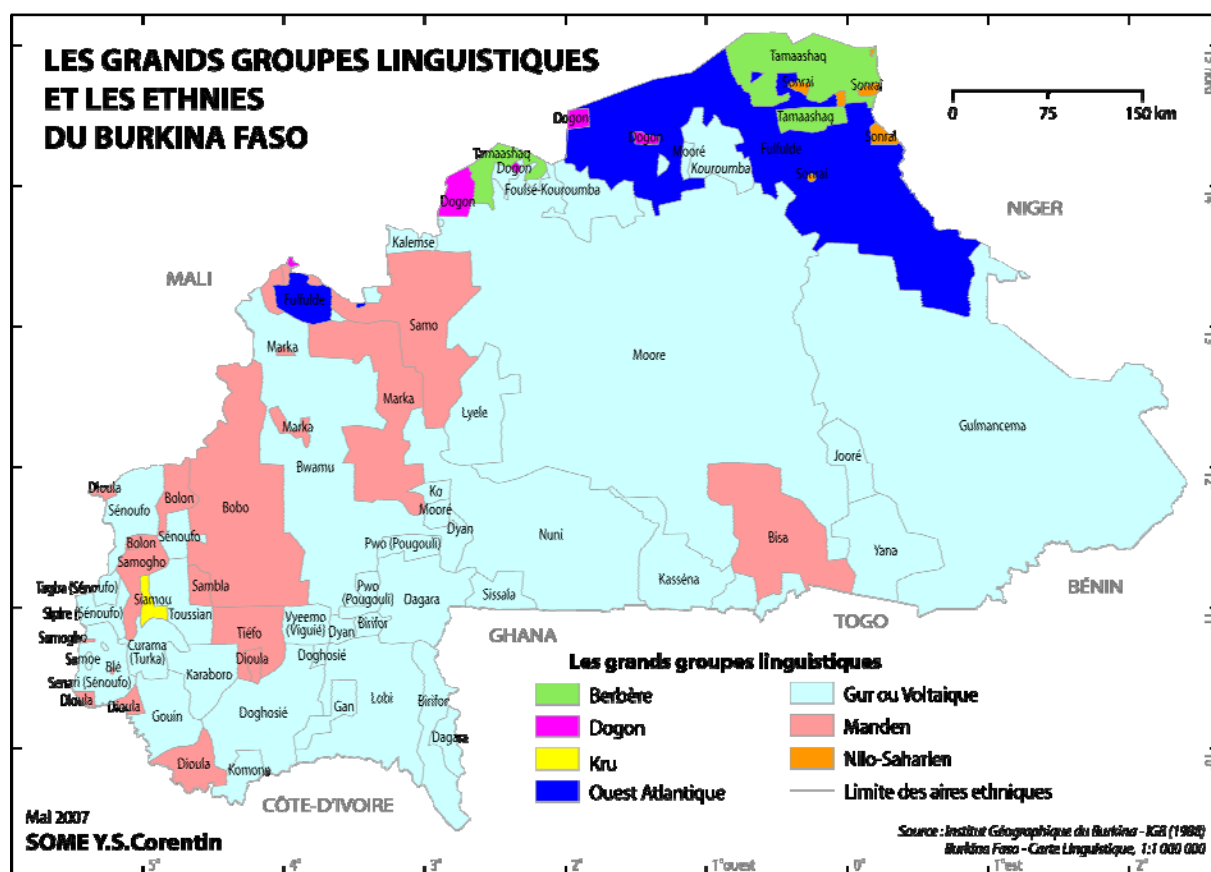


Figure 4 : Répartition des ethnies au Burkina Faso

Ces ethnies ont des organisations sociales et politiques différentes. Certaines sont des sociétés très hiérarchisées et sédentarisées qui formaient, avant la colonisation, de puissants empires. C'est le lieu de citer l'empire des Mossés, le royaume des Gulmacéma, les Emirats du Liptako et du Yagha, etc. Ces entités politiques fortes occupent aujourd'hui les parties centre, nord et est du pays. Les régions sud et ouest du pays sont occupées par des sociétés plutôt segmentaires, avec des systèmes politiques très peu centralisés, et sont connues sous le nom de société acéphale ou anarchico-

³⁹ Les villages sont des localités dotées du statut de village dans l'organisation du territoire au Burkina Faso. Il n'est pas à confondre avec la localité qui est plus générique.

⁴⁰ <http://www.insd.bf/>

classement de nombreuses zones en parcs nationaux justifie sans doute les faibles densités de population. Plus de 60% des aires protégées du Burkina Faso sont situées dans cette région.

L'extrême nord connaît de faibles densités de population. Même si l'Emirat du Liptako s'y trouve, cette partie du pays est plus propice à l'activité d'élevage qu'à l'agriculture. Les populations y partiquent plutôt une activité pastorale transhumante.

Par contre, le sud et l'ouest concentrent, en dépit des densités de populations relativement faibles, de forte d'implantations humaines. Ce qui est révélateur du caractère très segmentaire des structures sociopolitiques. En outre, cette zone était jadis très peu hospitalière en raison de son infestation par de grandes endémies comme la cécité de rivières ou la maladie du sommeil.

Du point de vue de la répartition des populations, les densités diminuent du centre vers les extrémités. L'histoire du peuplement et les politiques d'aménagement du pays sont vraisemblablement les causes de cette organisation spatiale des densités de population. Le milieu biophysique présente une organisation et une structure différentes. C'est l'objet de la suite de cette section consacrée à la présentation du milieu qui s'articule autour de la géomorphologie, du climat, de l'hydrographie et de la végétation.

II.3.2 LE RELIEF

Le Burkina Faso se caractérise par un relief plat, quasi monotone. C'est une vaste pénéplaine parsemée d'élévations de quelques dizaines de mètres de commandement. Le point le plus haut du pays, le Mont Ténakourou, culmine à 749 mètres. L'altitude moyenne est de 400 m. Près de la moitié du pays se situe entre 250 et 350 mètres d'altitude et plus de 98 % entre 160 et 480 m. La dénivelé maximale ne dépasse pas 600 m (Figure 6).

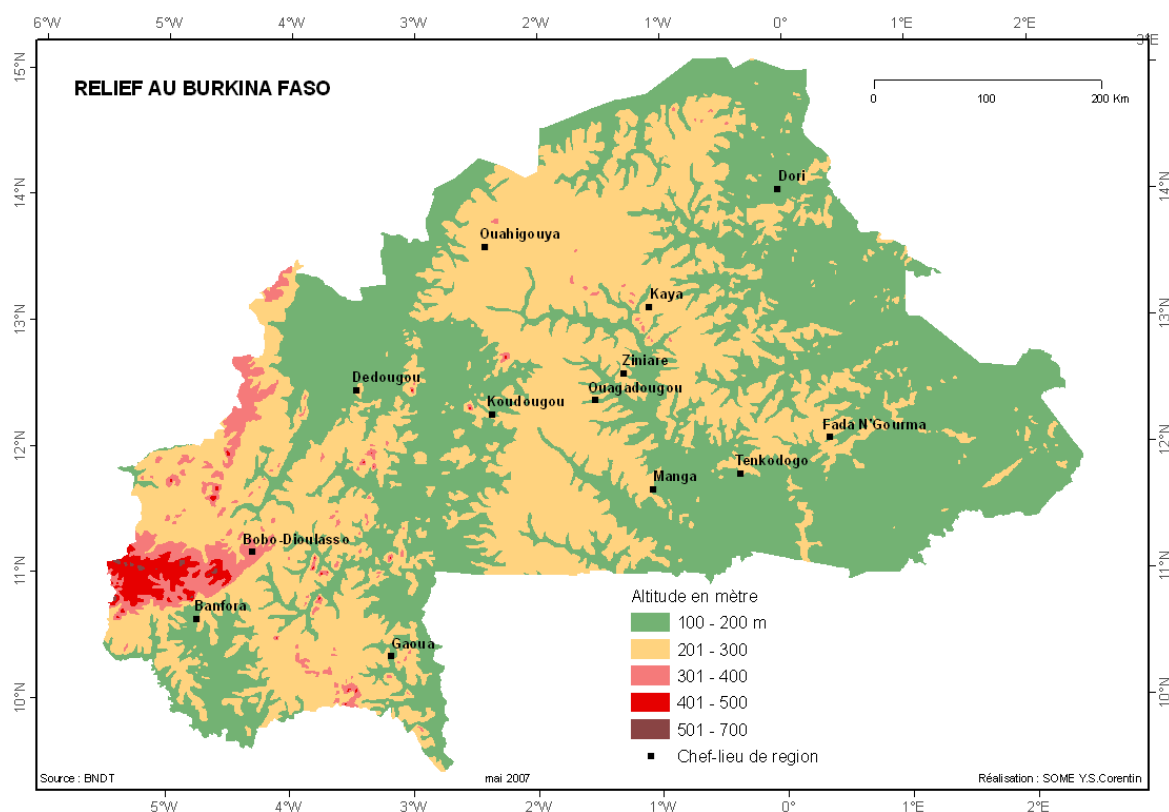


Figure 6 : Les grands ensembles topographiques du Burkina Faso

D'un point de vue géologique, le Burkina Faso est situé sur le craton Ouest Africain. Cinq grandes périodes marquent la mise en place des formations géologiques du Burkina Faso :

- le Précambrien D, avec la mise en place des formations cristallines constituées essentiellement de granito-migmatiques. Ces formations couvrent plus de 136 000 km², soit près de la moitié du territoire ;
- le Précambrien C ou Birimien pendant lequel ont été formées des roches métamorphisées d'origine volcanique ou sédimentaire et des granites syntectoniques. Ces formations couvrent près de 81 500 km², soit environ 1/3 du territoire ;
- le Précambrien A et le Cambro-ordovicien qui connaissent la mise en place de formations sédimentaires plus ou moins métamorphisées, composées surtout de grès, de schistes et de calcaire ;
- le Tertiaire avec la mise en place du continental terminal constitué essentiellement de formations sédimentaires marines (grès et schistes argileux) et des dépôts récents.

Du façonnement de ces roches, nous retenons quatre grands domaines géomorphologiques (Figure 7).

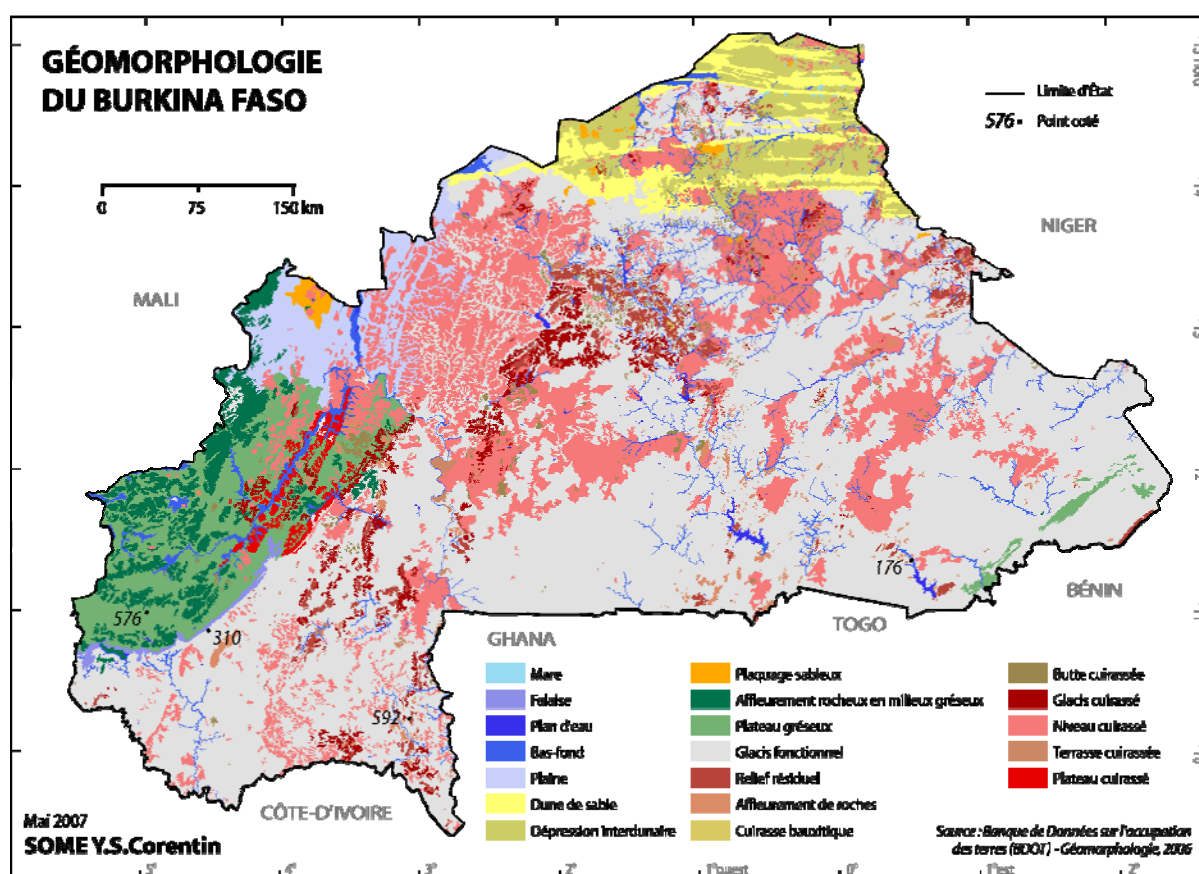


Figure 7 : géomorphologie du Burkina Faso

- 1- le domaine granitique. Il est plat avec, par endroit des reliefs résiduels de type chaos de blocs issus de l'érosion différentielle liée à l'intrusion des venues volcaniques ;
- 2- le domaine volcanique. Il est constitué de chaînons appalachiens tantôt linéaires, tantôt curvilinéaires qui s'étirent sur des dizaines, voire des centaines de kilomètres et qui sont quelquefois coiffés de cuirasses bauxitiques ou ferro-allumineuses ;
- 3- le domaine sédimentaire, à dominance gréseuse, s'étend sur l'ouest et l'extrême est du pays. Il domine les formations du socle avec des altitudes allant de 450 à 500 m. Le contact très marqué avec les formations du socle a permis de dégager des escarpements très abrupts ;
- 4- le domaine éolien qui se situe dans le nord du pays.

Sur ces formations géologiques se sont développés les sols.

II.3.3 LES SOLS

Les sols sont très diversifiés au Burkina Faso. La carte pédologique du Burkina Faso au 1/500 000 compte 11 classes⁴² (Tableau 1).

⁴² La classification considérée est la classification française

Tableau 1 : Les sols au Burkina Faso

Classe des sols	Superficie en km ²	% de la superficie
Sols à mull	14481,25	5,32
Sols à Sesquioxydes	7624,3	2,8
Sols à Sesquioxydes et matière organique rapidement minéralisée	68263,12	25,07
Sols Fersiallitiques	12911,9	4,74
Sols Halomorphes	8854,53	3,25
Sols hydromorphes	48081,52	17,66
Sols isohumiques	4729,6	1,74
Sols minéraux bruts	12684,72	4,66
Sols peu évoluées	84185,05	30,92
Vertisols	2180,01	0,8
Vertisols et paravertisols	8311,52	3,05

Source : O.R.S.T.O.M., (1969-1973)

Parmi les facteurs qui ont contribué et contribuent toujours au façonnement des formes présentées en géomorphologie, le climat occupe une place de choix.

II.3.4 LE CLIMAT

Le climat du Burkina Faso est réglé par trois vents dominants issus de trois zones de haute pression que sont les anticyclones : l'anticyclone de Ste Hélène, l'anticyclone des Açores et l'anticyclone temporaire du Sahara qui renforce le précédent

Le premier de ces vents, communément appelé Mousson, est un alizé provenant de l'anticyclone de Ste Hélène dans l'océan Atlantique. Il souffle du sud-ouest vers le nord-est dans l'hémisphère Nord. C'est un vent humide. Le second, l'Harmattan, est également un alizé. Il est issu de l'anticyclone du Sahara. C'est un vent temporaire sec et desséchant, froid la nuit et chaud le jour. De direction nord-est / sud-ouest dans l'hémisphère Nord, il souffle en sens opposé à celui provenant de Ste Hélène. Il relaie le vent issu de l'anticyclone des Açores, au large des côtes portugaises. La rencontre de ces deux vents crée le front intertropical (FIT) ou convergence intertropicale (CIT). Le balancement du FIT, conjugué avec la position en latitude, détermine un gradient climatique de direction nord-nord-est – sud-sud-ouest. La Figure 8 illustre cela.

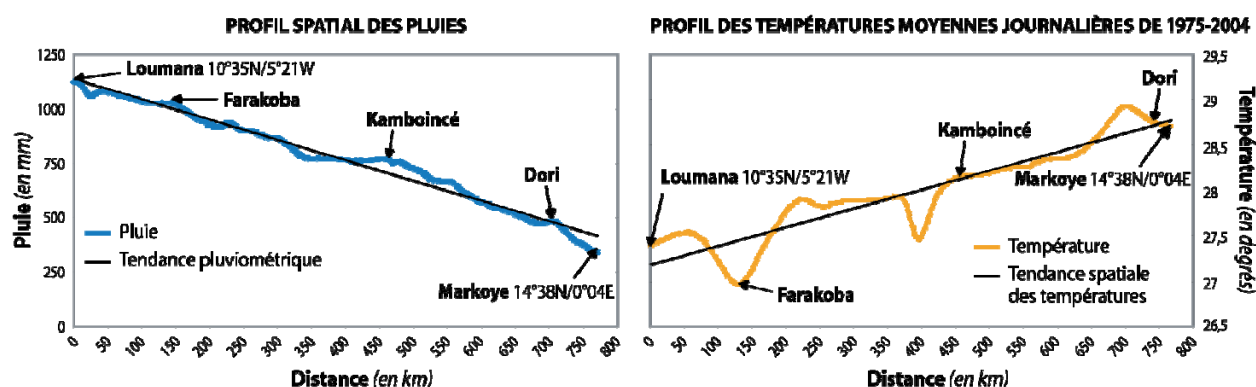


Figure 8 : Profils de distribution des températures et précipitations au Burkina Faso

Ces deux courbes représentent, le long d'un transect de 800 km orienté nord-nord-est – sud-sud-ouest, les valeurs moyennes des précipitations et des températures. Sur ce transect, les précipitations augmentent du nord-nord-est vers le sud-sud-ouest. Les températures diminuent, quant à elles, en sens opposé. Les précipitations passent ainsi de plus de 1000 mm de pluie par an à Loumana, Banfora, Farakoba, Legmoin, Batié, au sud-ouest à moins de 500 mm à Dori, Gorom-Gorom, Markoye, Djibo au nord. La pluviosité d'ensemble du pays est de 777 mm. La droite de tendance spatiale du profil des précipitations a une pente négative, avec une différence d'environ 650 mm de pluie entre l'extrême nord et l'extrême sud-ouest. A l'opposé, les températures moyennes augmentent de 27,6°C à Diébougou, à 29,45°C à Dori, soit une amplitude à peine supérieure à 2,5°C. La droite de tendance du profil des températures est largement positive.

Sur le plan des températures, 4 saisons thermiques se distinguent :

- une froide et sèche qui s'étend de novembre à février,
- une chaude et sèche qui va de mars à mai,
- une froide et humide qui va de juin à août,
- une chaude et humide qui va de septembre à octobre.

La Figure 9 illustre ce propos à travers les diagrammes ombro-thermiques de trois stations météorologiques.

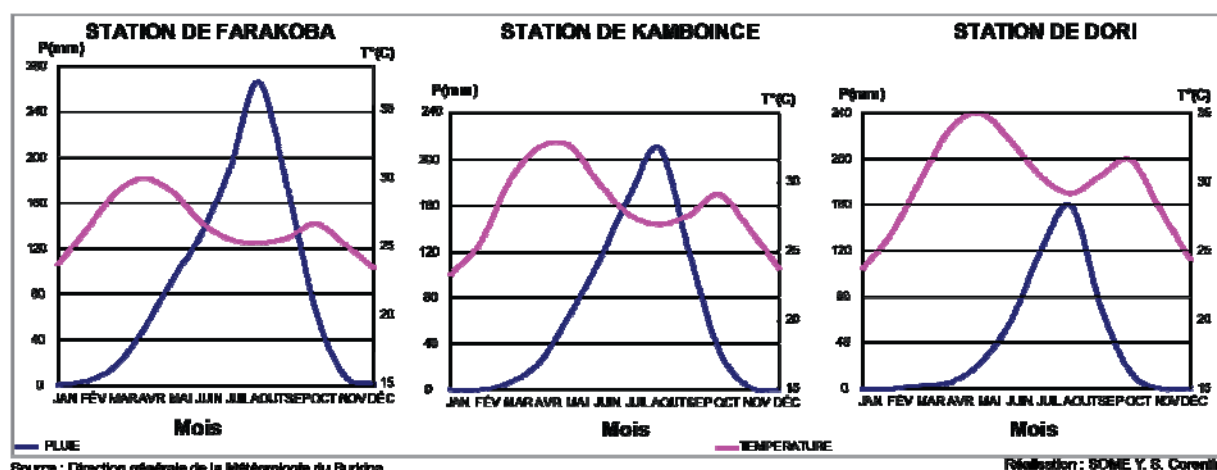


Figure 9 : Diagrammes ombrothermiques de quelques stations de références au Burkina Faso

La station de Dori est située au nord du pays, celle de Kamboince au centre et celle de Farakoba au sud-ouest. Sur ces trois stations, les courbes de températures sont bi-modales. Le premier maximum se situe entre le mois d'avril et le mois de mai, et le second entre le mois de septembre et d'octobre. Les moyennes des températures maximales observées entre 1975 et 2005 sont toujours supérieures à 25°C pour l'ensemble des stations. Elles varient entre 42,2°C à Dori au nord et 29,3°C à Bobo Dioulasso au sud-ouest.

Les minima sont observables au mois d'août puis entre les mois de décembre et de janvier pour toutes les stations et demeurent toujours supérieures à 10°C. Les moyennes des températures minimales, quant à elles, varient de 28,1°C à Dori à 13,3°C dans la vallée du Kou au sud-ouest.

L'amplitude thermique moyenne annuelle est de 13,3°C. Elle varie également du nord au sud-ouest, entre 16,4°C à Dionkélé, au centre-nord et 11,3°C à Niangoloko, à l'extrême sud ouest.

Sur le plan des précipitations, la courbe est uni-modale, avec un maximum au mois d'août pour l'ensemble des stations. Cela témoigne de l'existence, au plan pluviométrique, de deux saisons fortement contrastées par rapport aux saisons thermiques. La première est pluvieuse et intervient entre mars et juin. La seconde est sèche et se manifeste entre septembre - octobre et avril - mai.

La mise en rapport des températures et des précipitations, à travers des climogrammes dans trois stations (Figure 10), permet d'observer des courbes d'allure très allongée.

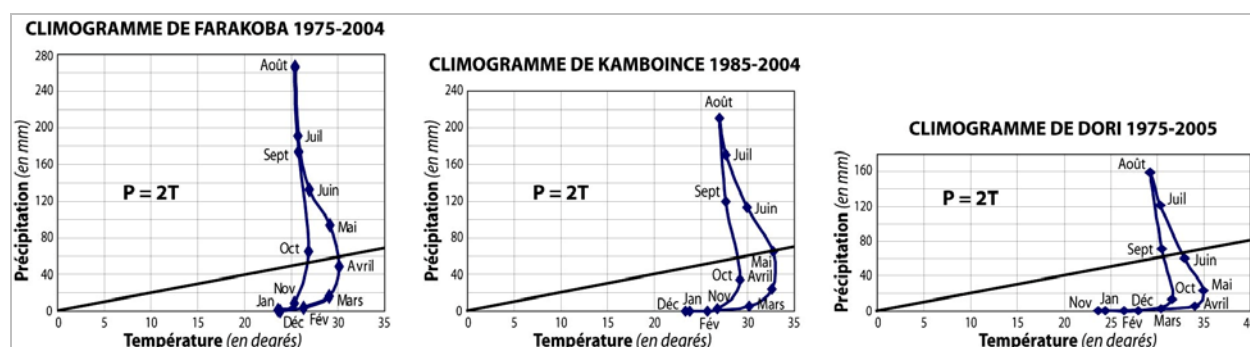


Figure 10 : Climogramme de quelques stations de référence au Burkina Faso

La droite d'équation ($P=2t$) permet de délimiter deux saisons : la première, humide, appelée hivernage, longue de 3 à 5 mois; elle intervient entre la troisième décennie d'avril et la troisième décennie d'octobre au sud-ouest et entre la troisième décennie de juin et la première décennie d'octobre au Nord. La dernière, sèche contraste avec l'hivernage. D'une durée de 7 mois au sud-ouest contre 9 mois au nord, elle occupe le reste de l'année (Figure 11).

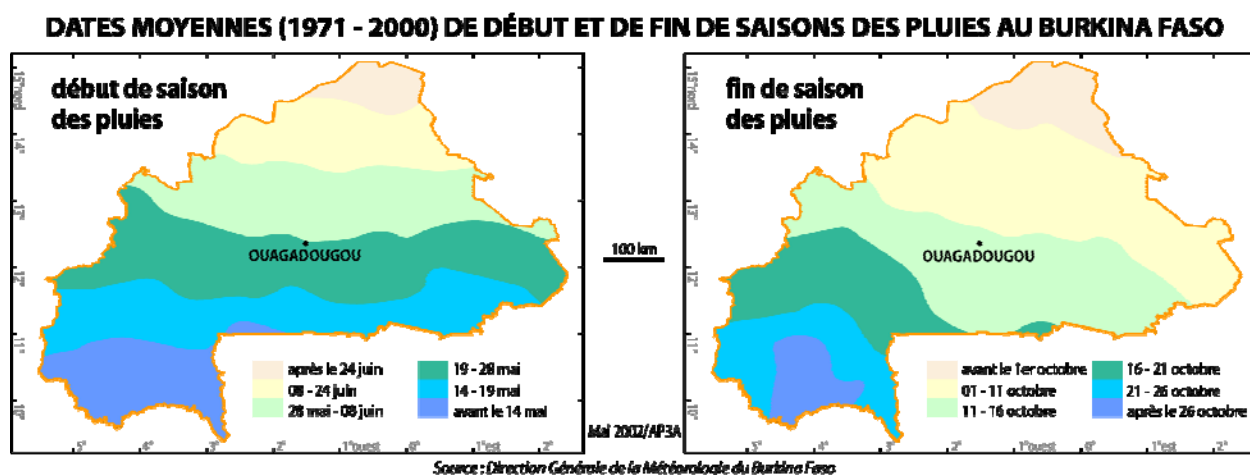


Figure 11 : Début et fins de saison pluvieuse au Burkina Faso

Le Burkina Faso est donc un pays chaud. Il est caractérisé par des températures toujours élevées avec des amplitudes thermiques faibles. Les précipitations marquent fortement les différences entre les saisons et varient entre 1 200 mm à l'extrême sud-ouest et 450 mm au nord. C'est un climat tropical allant du pré-guinéen humide dans l'extrême sud-ouest au sahélien dans le nord. Entre ces deux extrêmes, les variantes observées reposent essentiellement sur les volumes des précipitations et sur la durée de la saison pluvieuse.

Le climat, à travers les précipitations et l'évaporation, influence énormément la dynamique des cours d'eau.

II.3.5 L'HYDROGRAPHIE⁴³

Le réseau hydrographique du Burkina relève de trois bassins internationaux que sont : la Volta, le Niger et la Comoé auxquels s'ajoute un système endoréique. Ces trois bassins correspondent, à l'intérieur du pays, à quatre bassins nationaux que sont : le Mouhoun, le Nakambé, la Comoé et le Niger. Chacun de ces systèmes hydrographiques est composé des principales rivières dont les bassins versants structurent le drainage des eaux.

Dans le Tableau 2, ci-dessous, sont présentés les différents bassins versants internationaux, les bassins nationaux, les sous bassins hydrographiques, ainsi que les bassins endoréiques.

Tableau 2 : Les bassins hydrographique du Burkina Faso

Bassin international	Bassin national	Sous-bassin national	Superficie (km2)
Comoe	Comoe	Comoé-Léraba	17 590
Niger	Niger		83 442
		Beli	15 382
		Gorouol	7 748
		Darjol	1 709
		Faga	24 519
		Sirba-Gouroubi	11 946
		Bonsoaga	7 231
		Dyamangou	3 759
		Tapoa Mekrou	5 707
		Banifing	5 441
VOLTA			172 968
	Nakambé		81 932
		Pendjari-kompienga	21 595
		Nakambé	41 407
		Nazinon	11 370
		Sissili	7 559
	Mouhoun		91 036
		Mouhoun supérieur	20 978
		Mouhoun inférieur	54 802
		Sourou	15 256
Bassin endéroïque	Bassin endéroïque	Bassin endéroïque	300
BURKINA FASO			274 300

Source : (MEE-Burkina, 2002)

Ce réseau compte des lacs naturels et surtout de nombreuses retenues d'eau artificielles. Par exemple, le bassin de la Comoé, qui couvre une superficie de

⁴³ Cette présentation de l'hydrographie s'appuie essentiellement sur la Carte de L'Afrique de l'Ouest, 1 : 500.000, Institut Géographique National; Paris ; le Rapport sur l'Etat de l'Environnement au Burkina de 2002, la Banque nationale de données topographique Burkina et Programme GIRE

17 590 km², comptait en 2000, 27 barrages et 20 lacs naturels. Le bassin du Nakambé qui a une superficie de 81 000 km², abrite plus de 400 barrages dont les plus importants du pays. Il abrite : les barrages hydroélectriques de Bagré et de la Kompienga et le barrage de Ziga. S'y retrouvent aussi de nombreux lacs naturels comme celui de Dem, de Bam. D'après la Direction des Ressources en Eau du Burkina, plus de 800 barrages fonctionnels ont été inventoriés en 2006.

La Figure 12 présente l'hydrographie du Burkina Faso à travers les grandes rivières, leurs affluents et leurs bassins versants.

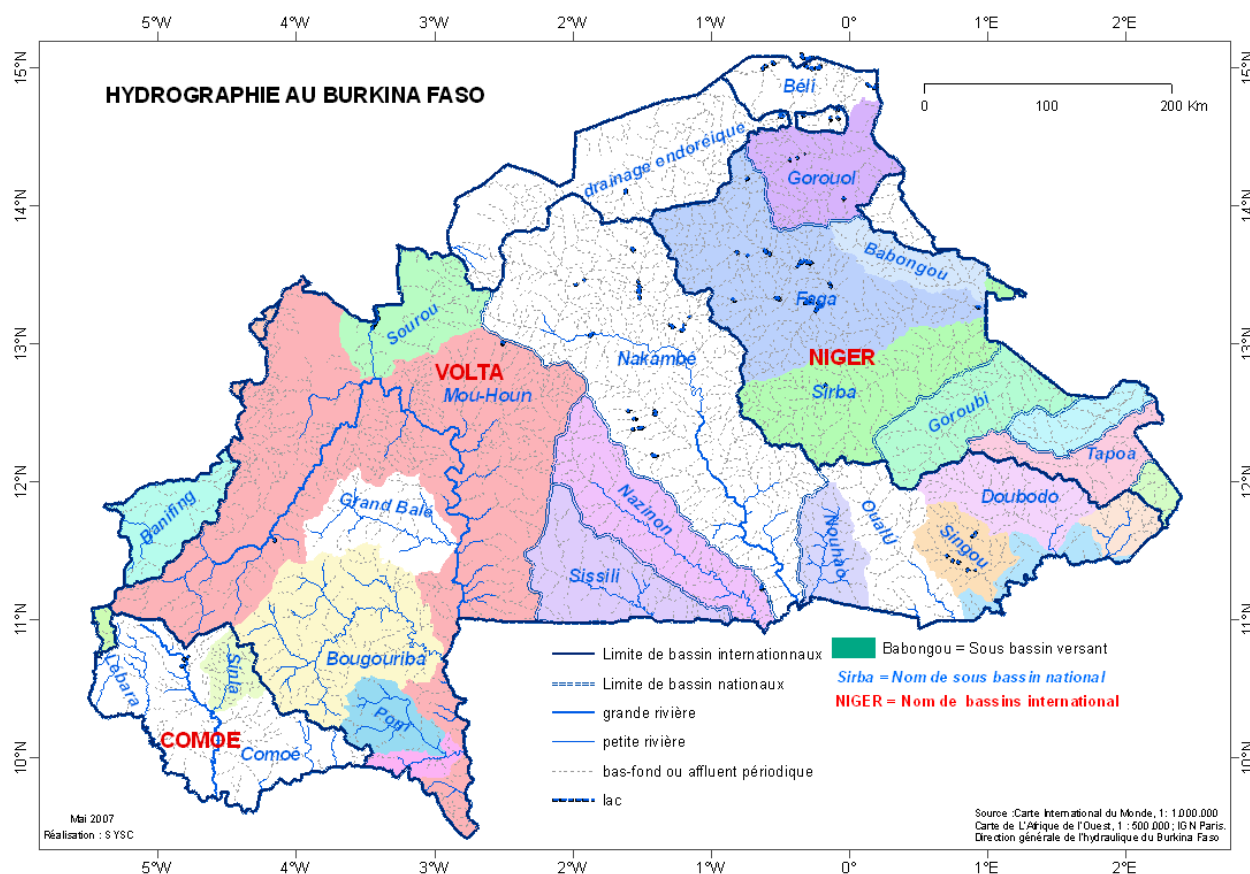


Figure 12 : Les principaux cours d'eau et leur bassins versants au Burkina Faso

Il apparaît sur ce réseau que les cours d'eau les plus importants se situent tous dans la partie ouest et sud du pays. Par ailleurs, la plupart de ces cours d'eau, à l'exception du Nakambé, prennent leur source sur les hautes terres du sud-ouest, dans la région des Hauts Bassins. Ceci renforce le caractère pluvial des régimes hydrologiques.

Ce réseau hydrographique du Burkina Faso est composé essentiellement de cours d'eau à régime pluvial tropical, fortement tributaire des précipitations. Ceux-ci sont, pour l'essentiel, intermittents avec des écoulements qui s'arrêtent quelques jours à quelques semaines seulement après les précipitations Somé, (2000). Moins d'une dizaine d'entre eux sont permanents et coulent toute l'année. La plupart des cours d'eau connaissent leur débit maximum en fin août – début septembre,

pendant la période de maximum pluviométrique. En contrepartie, les étiages sont très sévères. Ils interviennent juste avant le début de la saison de pluies qui se situe entre la fin du mois d'avril et le début du mois de juin selon la position géographique. Il arrive souvent que les écoulements arrêtent complètement. Dans un milieu où la disponibilité en eau constitue une contrainte majeure, l'organisation du couvert végétal est fortement influencée par le climat et l'hydrographie.

II.3.6 LA VEGETATION

En référence aux travaux de Fontes et Guinko, (1995); Guinko, (1984) et à la cartographie de la végétation de Fontes et *al.*, (1994), la végétation du Burkina Faso relève de deux domaines phytogéographiques : le domaine phytogéographique sahélien et le domaine phytogéographique soudanien. Leurs aires de répartition se superposent aux grands ensembles climatiques (Figure 13).

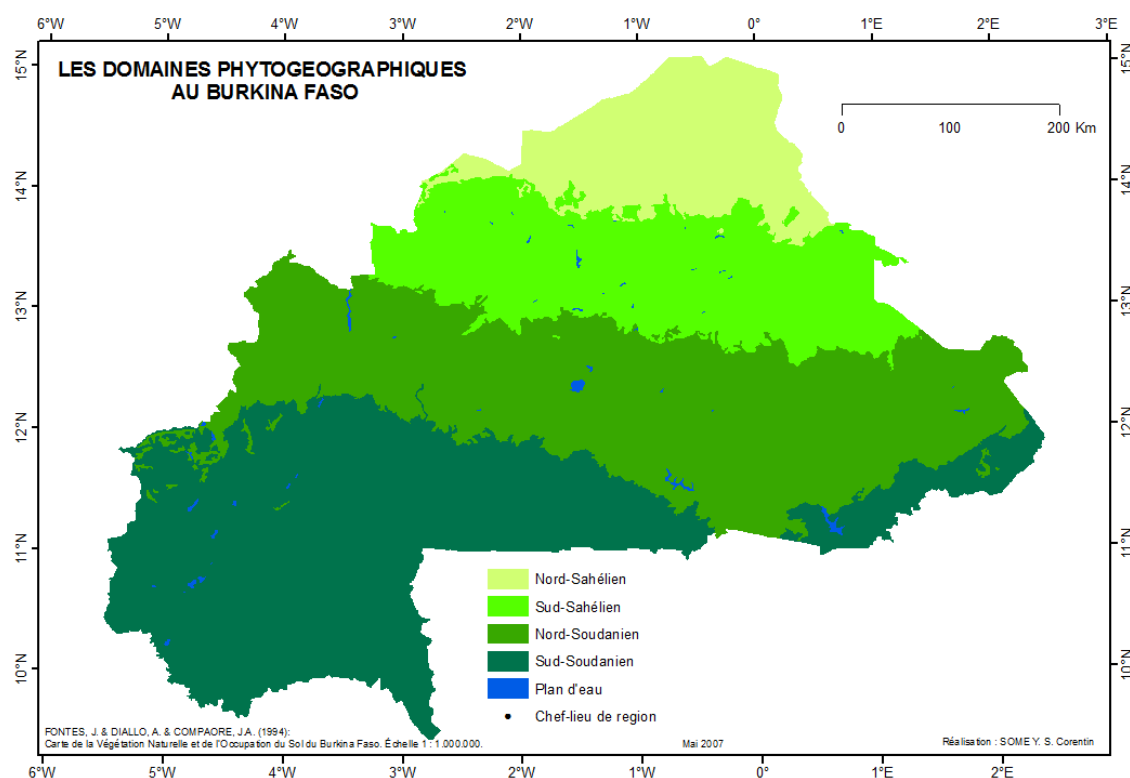


Figure 13 : Domaines phytogéographiques du Burkina Faso

Le premier couvre environ 35% du territoire et se situe au nord du pays, à partir du 13^{ème} parallèle. Sa limite sud, la plus pluvieuse, coïncide plus avec l'isohyète 650 mm. Le second, occupe un peu plus de la moitié sud du pays, au sud du 13^{ème} parallèle. L'isohyète 650 mm correspond à sa limite la plus sèche. Des différences de faciès sont perceptibles dans les différents domaines phytogéographiques (Figure 14).

Le domaine sahélien est celui des steppes. Les formations steppiques présentent des variantes allant d'une steppe presque sans espèces ligneuses à celle arbustive, voire arborée. Elles ont deux faciès majeurs appelés secteurs :

- le secteur nord sahélien caractérisé par des steppes herbeuses à *Aristida spp* et des steppes herbeuses à arbustives à *Cenchrus biflorus*, *Combretum glutinosum*, *Balanites aegyptiaca* ;
- le secteur sud sahélien, plus humide est le domaine de steppes avec une présence relativement plus abondante de ligneux. Ce sont des steppes arbustives à arborées truffées de fourrées et dominées essentiellement par des combretacés et des espèces arbustives du genre *Acacia* (*Acacia tortilis ssp.* *Acacia raddiana*, *Acacia laeta*, *Combretum spp.*). Les vallées de ce secteur abritent quelques savanes arborées à *Butyrospermum parkii*⁴⁴, *Lannea microcarpa*, *Acacia albida*.

Le domaine soudanien est celui de la savane, avec un tapis herbacé haut et continu comparativement au domaine sud sahélien. Cette savane présente de nombreux faciès regroupés en deux secteurs phytogéographiques : le secteur nord soudanien et le secteur sud soudanien.

- Le secteur nord soudanien est constitué pour l'essentiel de savanes fortement anthropisées du fait des densités de population élevées. Il correspond à des savanes arbustives à arborées à *Combretum spp*, *Anogeisus leiocarpus*, *Butyrospermum parkii*, *Laenna spp*. Il renferme aussi des savanes parcs dominées par des espèces arboricoles très utilitaires et protégées. Ce sont en général les parcs à *Butyrospermum parkii*, *Parkia biglobosa*, *Adansonia digitata* et *Tamarindus indica*.
- Le secteur sud soudanien, quant à lui, est constitué pour l'essentiel de savane arborée à boisée avec comme espèces dominantes *Burkea africana*, *Butyrospermum parkii*, *Pterocarpus erinaceus*, *Parkia biglobosa*. Il abrite aussi des forêts claires avec des espèces dominantes telles que *Isobertia doka*, *Burkea africana*, *Terminalia spp*. Les vallées des principales rivières tels comme Mouhoun sont colonisées des forêts galeries très denses, à *Berlinia grandifolia*, *Daniellia oliveri*, *Mitragyna inermis* et quelques reliques de palmiers rafia.

La végétation est donc organisée suivant un gradient nord-nord-est sud-sud-ouest comme le climat.

⁴⁴ *Vitallaria paradoxa* est le nouveau nom de *Butyrospermum parkii*

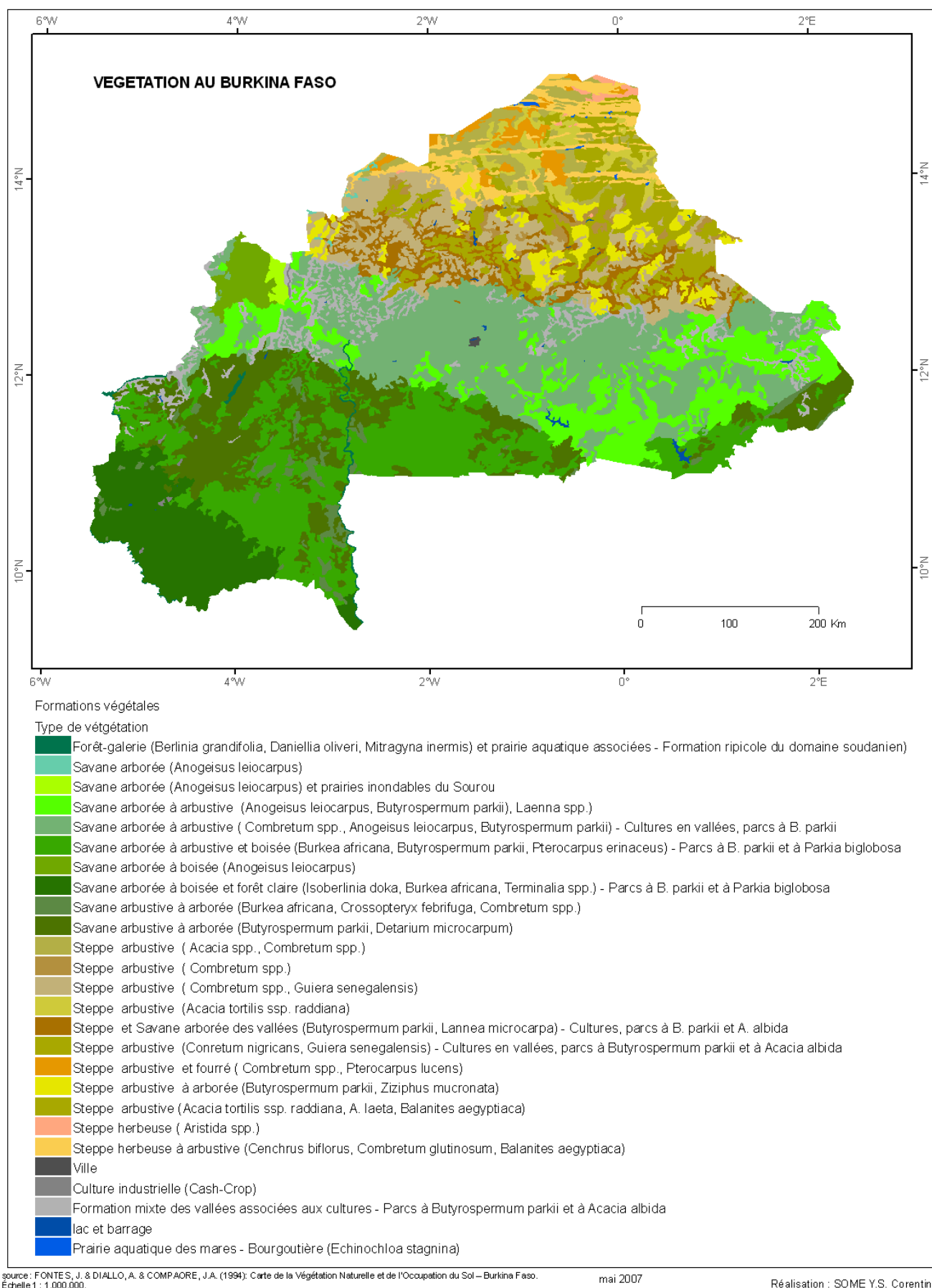


Figure 14 : Formations végétales du Burkina Faso

CONCLUSION DU CHAPITRE II

Au Burkina Faso, le paludisme constitue un problème majeur de santé publique. La stratégie de lutte contre cette maladie y est multiforme et intègre la lutte contre le vecteur, l'anophèle. Les anophèles constituent un complexe d'espèces. A cette diversité spécifique des anophèles, s'ajoutent, pour une même espèce, des exigences écologiques variables en fonction du stade de développement. Compte tenu du niveau de complexité de la question et de la nécessité de réduire le champ d'investigation, nous limitons notre étude aux formes moléculaires M et S.

Le Burkina Faso est un pays de transition entre le désert du Sahara et les zones forestières très humides du Golf de Guinée. Sur moins de 600 km à vol d'oiseau, il s'étend du Sahel, où les précipitations sont de l'ordre de 300 mm par an, à la zone soudanienne pré-guinéenne, avec plus de 1 200 mm/an, soit 900 mm d'écart. Ce pays se caractérise également par son relief plat, une diversité au niveau géomorphologique et un réseau hydrographique dense qui effacent la monotonie sous-jacente. Il présente aussi une diversité culturelle doublée d'une variation de densités régionales de population. Ces densités oscillent entre 10 et près de 500 habitants au km² et sont liées à l'histoire et aux politiques d'aménagement du territoire.

Un complexe d'espèces aux exigences écologiques variant d'un type à un autre et en fonction du stade de développement, dans un environnement offrant une diversité de conditions écologiques, tel est le visage du champ d'application de ce travail de recherche. Un tel schéma, dans une étude portant sur la modélisation, impose la définition d'une démarche méthodologique adaptée, objet du chapitre suivant.

CHAPITRE III : APPROCHE METHODOLOGIQUE DE L'ETUDE

La recherche sur la distribution spatiale des espèces vivantes est très ancienne et de nombreuses méthodologies ont déjà été construites. L'avènement de la télédétection, des SIG, de l'analyse spatiale et de l'analyse de données statistiques, a apporté une nouvelle dynamique dans la recherche sur la répartition des êtres vivants dans l'espace, particulièrement celle des vecteurs de maladies. Ces résultats peuvent aider à concevoir la méthodologie pour modéliser la distribution spatiale des formes moléculaire M et S d'*An. gambiae* au Burkina Faso. Il est donc opportun de faire un aperçu des méthodes de modélisation spatiale des espèces vivantes et d'y situer notre démarche.

L'essence de ce chapitre consiste donc à partir d'un rappel sur les méthodes de modélisation pour présenter la démarche méthodologique utilisée dans le cadre de ce travail. Nous y répondons aux questions portant sur l'articulation d'ensemble de la recherche, les méthodes et les matériels utilisés. Ce chapitre couvre les aspects portant à la fois sur la collecte et le traitement des données, l'évaluation des indicateurs et la vérification des hypothèses. Il est structuré en deux sections :

- la première, « aperçu des méthodes de modélisation de la distribution spatiale des espèces vivantes utilisant l'analyse spatiale et les SIG », fait un état de l'art sur les méthodes de modélisation ;
- la seconde, « démarche, matériel et méthodes », est consacrée à la méthodologie et au matériel utilisé dans cette étude.

III.1 APERÇU DES METHODES DE MODELISATION DE LA DISTRIBUTION SPATIALE DES ESPECES VIVANTES UTILISANT L'ANALYSE SPATIALE ET LES SIG

« Une vieille sagesse enseigne que rares sont les situations vraiment nouvelles et que d'autres ont déjà su résoudre les « nouveaux » problèmes que nous rencontrons » Le-Moigne, (1990). Les modèles de la distribution spatiale des espèces, et particulièrement des espèces qui sont des vecteurs de maladies, sont nombreux et variés en raison de la diversité des sites, de l'écologie et de l'étiologie des espèces. Cette fécondité de la recherche dans le domaine de la modélisation est liée sans doute aux usages dont les modèles font l'objet, au caractère bien souvent utilitaire et pragmatique des modèles. Quels sont ces usages ? Quelles sont les méthodes utilisées pour le processus de modélisation ? Cette section vise à répondre à ces deux questions. Elle a pour but de présenter les usages qui sont faits des modèles de distribution spatiale des espèces et les différents types de modèles. Cette présentation devrait contribuer à faire comprendre la dynamique de modélisation dans laquelle s'inscrit ce mémoire de recherche. Elle doit également permettre de faire comprendre les

choix qui sont faits en ce qui concerne les méthodes et données. Elle est composée de deux paragraphes :

- le premier est un exposé sur les impacts des usages des modèles sur le processus de modélisation
- le second, une classification des méthodes de modélisation.

III.1.1 L'IMPACT DES USAGES DES MODELES SUR LA MODELISATION

« Et quoiqu'on en dise, dans la vie scientifique, les problèmes ne se posent à nous d'eux-mêmes. C'est précisément ce qui donne la marque du véritable esprit scientifique... S'il n'y a pas eu de question, il ne peut y avoir de connaissance scientifique. Rien ne va de soi. Rien n'est donné. Tout est construit » G. Bachelard, 1938-1978, cité par Le-Moigne, (1990). Le modèle est le fruit du raisonnement de son auteur qui cherche, à travers lui, à présenter et à expliquer les points saillant d'un phénomène. « Le modèle est une œuvre produite par l'imagination au sens le plus strict puisque c'est la prise en compte d'éléments sélectionnés entre lesquels sont établies des relations pertinentes. L'adéquation du modèle à une partie de la réalité est l'expression de l'imagination scientifique » Raffestin, (1983) in Gumuchian et Marois, (2000). Les modèles sont donc des constructions de l'esprit et comme tels, ils sont objectivés, c'est-à-dire qu'ils visent un ou plusieurs objectifs. Nous retenons de présenter deux types d'objectifs dans le domaine du contrôle des populations vectorielles. Ce choix est lié au fait que ces objectifs déteignent sur le choix de données. En complément à l'assertion de Bachelard sur l'esprit scientifique, nous disons avec les mots de Le-Moigne, (1990) s'appuie lui-même sur le « Discours de l'esprit positif » d'Auguste Comte 1843, que « les données du problème sont données par le problème que rencontre le modélisateur ».

III.1.1.1 Les usages des modèles pour l'élaboration des politiques et définition de stratégies

Les plans de lutte contre les maladies à transmission vectorielle intègrent, bien souvent et d'ailleurs très logiquement, le contrôle, voire l'éradication des vecteurs. La mise au point de tels programmes nécessite de disposer d'information sur la bio-écologie du vecteur et surtout sur la répartition dans l'espace. Ceci explique le recourt fréquent à des modèles spatiaux. Ces modèles servent à la définition de politiques ou de stratégies de contrôle ou d'éradication du vecteur. Ceux utilisés à de telles fins couvrent généralement de grands espaces. Ce sont des modèles couvrant de grands espaces, des étendues continentales à nationales Pin-Diop, (2006). Or, la modélisation s'appuie bien souvent sur des observations. L'observation est une mesure dans le temps et dans l'espace et fait

appel à la notion d'échelle. « Elle est liée à la mesure de l'espace ou du temps. Passer d'une échelle à une autre consiste à changer d'unité de mesure » ⁴⁵Pavé A (1994) cité Mathian et Piron, (2001). La taille des objets géographiques dans l'espace et la taille de l'espace à couvrir ne sont pas sans effets sur l'utilisation des données dans les travaux de recherche, particulièrement en modélisation spatiale. « Le choix de l'échelle géographique oriente et influence les résultats. En effet, les résultats changent selon l'échelle géographique : c'est le problème de l'unité spatiale modifiable bien connu sous l'acronyme MAUP en anglais. Les conclusions déduites à une échelle géographique ne sont pas nécessairement valables à une autre échelle » Gumuchian et Marois, (2000). Les échelles influent aussi sur les coûts financiers, temporels et matériels. Ces modèles destinés à l'élaboration de politiques et de stratégies générales utilisent principalement des données provenant d'images satellites basse résolution comme les AVHRR NOAA, HRR Meteosat, avec des résolutions spatiales de l'ordre de 1 km à 1,5 km. Ils utilisent également des données météorologiques et des modèles numériques de terrain comme GOTO30. Leur objectif en général est d'identifier et de cartographier les zones à risque de manière à ce que les efforts de contrôle et les stratégies d'intervention soient les plus efficaces et ciblés possible » Tran et *al.*, (2005). Ils sont appelés, en raison de la faible résolution spatiale de données, des modèles à petite échelle. Les exemples sans doute les plus cités dans la littérature spécialisée concernent les glossines/la trypanosomiose et la Fièvre de la Vallée du Rift en Afrique de l'Est. Pin-Diop, (2006) fait un exposé assez exhaustif des travaux développés dans le cadre de la lutte contre la trypanosomiose et le contrôle de la population de glossines avec les travaux suivants: Hendrickx et *al.*, (2001); Hendrickx et *al.*, (1989); Robinson et *al.*, (1997); Rogers et Randolph, (1993); Rogers et Williams, (1993). Pour la Fièvre de la Vallée du Rift, elle utilise les travaux de Anyamba et *al.*, (2001); Davies et *al.*, (1992); Linthicum et *al.*, (1999); Linthicum et *al.*, (1987); Linthicum et *al.*, (1990).

III.1.1.2 L'impact des usages des modèles pour l'action

La définition de politiques générales ou de stratégies globales n'est pas la seule utilisation des modèles spatiaux dans le contrôle de populations vectorielles. Les modèles spatiaux sont également recherchés pour l'intervention, pour l'action ou pour les ripostes contre la prolifération de certains vecteurs. Dans ce cas, il est inopportun de se contenter de données à basse résolution spatiale des modèles à petite échelle. L'usage des modèles spatiaux pour l'action requiert des données plus précises et sur des espaces beaucoup plus restreints. Les étendues spatiales comme les villes, les sous bassins versants sont raisonnables. Les résolutions spatiales de l'ordre de quelques dizaines mètres

⁴⁵ PAVE A, *La modélisation en biologie et en écologie*, Edition Aléas, Lyon, 1994.

(30 m) à sub-métriques (60 cm) sont les plus appropriées. Dans ce cas, les données proviennent d'observations de terrain et de traitements d'images satellites, haute résolution à très haute résolution. Les images comme LandSat TM et ETM+, SPOT XS, XI, Ikonos et QuikBird sont adaptées à ce type de modélisation. Les informations recherchées en ce qui concerne les vecteurs de maladies sont par exemple l'occupation des terres, les gîtes de reproduction, les gîtes de repos et des éléments beaucoup plus précis sur la végétation comme la composition floristiques. Les travaux de Mushinzimana et *al.*, (2006) utilisant les Images LandSat 7 au Kenya, Ratana et *al.*, (2005) en Corée avec également des images LandSat 7 tous relatifs au vecteur du paludisme, Pin-Diop, (2006) pour les vecteurs de la fièvre de la vallée du Rift à partir d'images satellites Landsat 7 au Sénégal Septentrional, en sont des illustrations.

Que les travaux se réalisent au niveau local ou global, pour des raisons d'action ou pour des raisons d'élaboration de politiques générales, les méthodes restent les mêmes. C'est principalement le type de données et les usages des résultats qui changent. Les chercheurs utilisent soit des modèles rhétoriques, soit des modèles mathématiques, soit des modèles iconiques, soit enfin un mélange de deux ou des trois. Cette synthèse introduit au paragraphe consacré aux méthodes de modélisation.

III.1.2 LES MODELES CLASSES D'APRES LES METHODES

Il existe de nombreux auteurs qui classent les modèles suivant le critère des méthodes. C'est le cas de Tran et *al.*, (2005) qui proposent une classification en deux groupes : les modèles conceptuels qui « s'attachent à expliquer de manière qualitative un phénomène » et les modèles mathématiques. Les modèles mathématiques comprennent les modèles statistiques / empiriques et les modèles théoriques. Un troisième groupe de modélisation, que l'on pourrait appeler modèle informatique, réunit les programmes développés pour faire de la cartographie « automatique » de la distribution spatiale des espèces. Nous pouvons citer aussi, Guisan et Zimmermann, (2000) qui présentent les « Predictive habitat distribution models in ecology ». Cet article propose une analyse des méthodes de modélisation en écologie et peut inspirer les chercheurs qui s'intéressent à ce domaine.

Cependant, nous nous adosserons dans cette présentation sur celle de Brunet, (2000a) qui semble plus complète et intègre de façon plus ou moins directe la plupart des classifications suivant la méthode. En outre, en tant que géographe, Roger Brunet tient compte de la spécificité de la discipline. Il distingue trois groupes de modèles : les modèles rhétoriques, les modèles mathématiques et les modèles iconiques. Ce paragraphe est structuré autour de ces trois groupes de modèles.

III.1.2.1 Les modèles rhétoriques

Ce type de modèle procède à une description des phénomènes et s'exprime par des textes, des discours. C'est une formalisation par le discours. Il présente les grands traits des phénomènes par des textes. Il réunit par le discours un ensemble de faits et de données qui donnent les caractéristiques essentielles des phénomènes. Il s'appelle aussi modèle conceptuel. En écologie, ce type est très utilisé et consiste à faire un rapprochement entre les éléments du paysage et l'écologie du vecteur. La démarche, dans le cas de la modélisation conceptuelle, s'appuie sur la recherche documentaire. Elle vise à acquérir de l'information pour produire un modèle discursif sur l'habitat du vecteur. Ces modèles rhétoriques sont très usités en sémiologie médicale, en taxonomie et plus généralement en systématique. Ils servent souvent à traduire certaines théories. Par exemple, la loi du minimum de Liebig fournit un modèle pour le raisonnement sur la productivité végétale.

Ce type de modèle est souvent appuyé par d'autres types de modèle, comme les modèles mathématiques.

III.1.2.2 Les modèles mathématiques

Les modèles mathématiques « s'expriment par des formules : quelque chose entre image et texte » (Brunet, 2000a). Ils « visent à décrire de manière quantitative le fonctionnement d'un système, en écrivant sous forme d'équations les lois qui le régissent » Tran et *al.*, (2005). « La modélisation en langage mathématique atteint assurément le degré supérieur d'abstraction, du moins du point de vue du langage, qui n'est pas le seul à prendre en compte » Durand-Dastès, (2001). Elle est considérée par beaucoup d'auteurs comme le plus haut niveau d'abstraction, le niveau le plus élaboré du modèle. « Il existe cependant une difficulté pour réaliser un compromis à partir d'un langage mathématique : il s'agit d'un langage qui se fonde sur l'exactitude et tend à évacuer toute approximation, imprécision ou ambiguïté pour réaliser un compromis avec d'autres disciplines pour l'étude d'un objet naturel, chaque discipline étant porteuse d'un point de vue. (...) L'art de la modélisation (mathématique) est de trouver le meilleur compromis entre la simplicité des formes mathématiques et le respect des éléments saillants et identitaires du phénomène étudié » Franc, (2001). « Parmi ces derniers, on distingue encore les modèles empiriques (ou statistiques), basés sur les données, et les modèles mécanistes (ou théoriques), qui sont basés sur les concepts » Tran et *al.*, (2005). Il convient d'ajouter à la liste des développements actuels de modules informatiques proposant des outils à intégrer sur les logiciels de SIG et visant l'élaboration de carte de distribution spatiale.

Les modèles mathématiques mécanistes ou théoriques

Ce sont des modèles visant à déterminer les relations à partir de la compréhension du fonctionnement du système Tran et *al.*, (2005). Ce type de modèle mathématique s'inscrit dans le cadre de la théorie des modèles. « La notion de modèle est utilisée de manière assez particulière, puisqu'elle utilise le terme « modèle » comme quelque chose qui peut donner sens à un discours théorique, par une correspondance terme à terme entre le modèle et la théorie formelle » Langlois, (2005). « La théorie exprime par la formulation explicite des relations qui existent entre des concepts » Gumuchian et Marois, (2000). Dans cette acceptation du modèle mathématique, le modèle tend à reproduire au plan théorique le phénomène en identifiant tous les déterminants et en définissant pour chaque déterminant sa contribution au fonctionnement du système. Il existe très peu de cas de modèles théoriques en géographie. Ils sont le plus souvent importés d'autres disciplines comme l'économie ou la sociologie. Comme le dit Gumuchian et Marois, (2000), « le géographe qui n'a pas souvent produit ses propres théories, procède à des transferts à partir d'autres disciplines. » Les plus célèbres transferts sont ceux de J.H Von Thünen sur la théorie des aréoles et ceux de W. Christaller sur les places centrales.

Les modèles mathématiques empiriques

Les modèles mathématiques de la distribution spatiale des vecteurs sont pour l'essentiel des modèles empiriques. C'est une mise en ordre dans la réalité empirique Durand-Dastès, (2001). Représenter un ensemble d'individus spatiaux décrits par deux variables ou un nuage de points, puis résumer celui-ci par le calcul d'une courbe d'ajustement relève bien de la modélisation. Il s'agit, en modélisation mathématique empirique, d'intégrer des méthodes statistiques classiques ou géostatistique dans le processus de modélisation. Le développement de ce type de modèle a également été facilité par l'essor récent des systèmes d'information géographique (SIG). Dans ce type d'exercice, l'intérêt est d'établir des corrélations statistiques entre présence/absence ou fréquence de vecteur et le milieu décrit par un certain nombre de paramètres. Ces derniers doivent être importants dans l'écologie du vecteur. Certaines sont déterministes comme les régressions, d'autres par contre sont stochastiques avec l'usage de techniques probabilistes.

Exemple d'utilisation des modèles mathématiques empiriques

Pour illustrer ce type de modèle, nous citons la thèse de Hélène Guis sur « *Géomatique et épidémiologie : Caractérisation des paysages favorables à *Culicoides imicola*, vecteur de la fièvre*

catarrhale ovine en Corse » soutenue en 2007, les travaux de Thomas et al., (2003) sur “*Linking field-based ecological data with remotely sensed data using a geographic information system in two malaria endemic urban areas of Kenya*”, Mushinzimana et al., (2006), (Ratana et al., (2005) “*Identification and characterization of larval and adult anopheline mosquito habitats in the Republic of Korea: potential use of remotely sensed data to estimate mosquito distributions*”.

« L’hypothèse sous-jacente à ces modèles est que la présence (...) des vecteurs dépend fortement de l’environnement : type de végétation, conditions climatiques (température, précipitations), etc. qui déterminent leur habitat. [...] La méthode consiste donc à rechercher des corrélations entre la distribution des vecteurs et des paramètres environnementaux. L’inversion de la relation statistique observée, à condition de pouvoir justifier le lien vecteur/environnement, permet d’extrapoler les résultats pour obtenir des cartes de la répartition des vecteurs » Tran et al., (2005). Dans ces cas, les logiciels de SIG doivent disposer d’autres fonctions de calcul de statistiques spatiales et être soutenus par des logiciels d’analyse statistique de données, lorsque la structure des données l’autorise (par exemple, en absence d’autocorrélation spatiale). Les SIG servent dans ce type de modèle comme outils d’analyse. Mais bien souvent dans ce cas, les logiciels de SIG sont très peu outillés pour prendre en charge certaines analyses Joost, (2006). D’où la création de module informatique intégrant des fonctions pour permettre la réalisation de la modélisation mathématique dans les SIG.

Les modèles mathématiques informatiques

Il s’agit de modèles beaucoup plus génériques consistant à palier les insuffisances des logiciels de SIG comme Arcview de Esri, MapInfo de Claristas ou SAVANE de l’IRD développé sous la supervision de Marc Souris⁴⁶. Ils proposent des modules complémentaires dédiés spécifiquement à la réalisation de la cartographie de distribution des espèces en général. Ce qui pourrait être fort utile dans la modélisation de la distribution spatiale des vecteurs de maladies. Les logiciels SIG offrent des fonctions classiques de traitements de l’information géographique souvent très riches mais dont l’usage spécifique dans la modélisation de la distribution spatiale des espèces en biogéographie et en écologie en général est très peu évidente. Joost, (2006) met ainsi en évidence le faible niveau d’intégration des outils d’analyse spatiale dans les logiciels de SIG : “*the lack of integration between GIS and spatial analysis, by picking out that GIS, despite being widely used in different types of spatially explicit studies, are still deficient in integrating generic statistical procedures for predictive purposes.*”

⁴⁶ http://www.th.ird.fr/publications/publications_auteurs_souris.html

Exemples de modèles mathématiques informatiques

Parmi ces modèles, nous pouvons citer "Biomapper"⁴⁷ » développé au Laboratoire de Biologie de la Conservation de l'Université de Lausanne, par Alexandre H. Hirzel. Ce logiciel est inter-opérable avec Idrisi de l'Université de Clarke aux Etats-Unis. Il est à sa version 4 et est gratuit. Nous pouvons aussi citer OpenModeller⁴⁸ développé par Centro de Referência em Informação Ambiental (CRIA), Escola Politécnica da USP (Poli), et Instituto Nacional de Pesquisas Espaciais (INPE). Comme il est dit sur leur site internet, "*openModeller is a fundamental niche modelling library, providing a uniform method for modelling distribution patterns using a variety of modelling algorithms.*" Cet open source est à sa version 7 et peut être intégré à ArcInfo de ESRI. Dans cette lignée, ComDym⁴⁹, qui est sa version 4, est développé par USGS.

La modélisation mathématique produit souvent des fonctions. L'inversion de ces fonctions permet d'obtenir des cartes de distribution qui font partie d'un autre type de modèle, les modèles iconiques.

III.1.2.3 Les modèles iconiques

Les modèles iconiques « passent par le dessin, qu'il s'agisse d'une courbe, d'un profil ou d'une configuration spatiale. [...] Le modèle iconique offre une vision panoptique, ou synchrone si l'on préfère ; il montre d'un coup les relations réciproques des lieux : il les met en situation. Il permet de comprendre, grâce au dessin, les effets d'interférence des formes, la déformation réciproque des modèles. [...] Il est commode et même fécond ; il n'exclut en rien la quantification, ni même la mathématisation ; il les appelle d'ailleurs, comme moyens de contrôle » Brunet, (2000a). La carte est l'exemple même de modèle iconique. Elle peut atteindre des niveaux d'abstraction très élevés. Par exemple, la carte d'isobare couramment utilisée est fondée sur un calcul de correction des observations et atteint un très haut niveau d'abstraction Durand-Dastès, (2001). Chorématique, chorème et chorotype sont des termes qui évoquent par la schématisation, un niveau d'abstraction plus élevé que celui de la carte.

Les modèles iconiques procèdent à la mise en image des phénomènes pour faciliter leur compréhension et leur présentation. Cette mise en image se traduit lorsque la dimension spatiale entre en jeu, par les cartes ou les chorèmes. La traduction sous forme de carte de phénomène répond à une double exigence : la fidélité et la lisibilité. L'élaboration des cartes relève donc de la modélisation spatiale. Elle s'appuie souvent sur des modèles mathématiques ou rhétoriques/discursifs.

⁴⁷ <http://www2.unil.ch/biomapper/>

⁴⁸ http://openmodeller.sourceforge.net/index.php?option=com_frontpage&Itemid=1

⁴⁹ <http://www.mbr-pwrc.usgs.gov/software/comdyn.html>

Ainsi, la cartographie de distribution de l'espèce peut-elle être la recherche dans l'espace (zone d'étude) des sites qui répondent au modèle décrit de l'habitat. L'information spatiale environnementale provient en général du traitement d'image de télédétection/photographie aérienne, d'observations sur le terrain, de l'interpolation de données climatique etc. En s'appuyant sur un SIG, de simples requêtes permettent de sélectionner des espaces répondant à la description de l'habitat du vecteur et à sa cartographie. Dans certains cas, les SIG servent aux traitements d'informations spatiales (géoréférencement des informations, interpolations et estimations, extrapolations). Comme le dit YÉ, (2005), "The power of GIS is its ability to integrate and manipulate multiple layers or themes of spatial data for a large area and from different sources at different scales [...] For malaria research and control, GIS has great potential since it has the capacity to integrate information on all aspects of malaria including the environmental factors, infrastructure and demography [...] it is a good tool of communication to support the decision making process". C'est donc la capacité des SIG à intégrer des données multisources, multimédia et à croiser ces informations à travers des requêtes spatiales qui est fortement utilisée. Les fonctions d'édition cartographique et d'aide à la décision sont également intéressantes pour la restitution des résultats.

Exemples d'utilisation des modèles iconiques

Ce type de modèle est fortement utilisé en géographie, en épidémiologie et en écologie. Les travaux de thèse Bouyer, (2006), « *écologie des glossines du Mouhoun au Burkina Faso : intérêt pour l'épidémiologie et le contrôle des trypanosomoses africaines* » soutenu en 2006, de Pin-Diop, (2006) "spatialisation du risque de transmission de la Fièvre de la Vallée du Rift en milieu agropastoral sahélien du Sénégal septentrional" soutenu en 2006, Balenghien, (2006), portant sur le thème « *De l'identification des vecteurs du virus West Nile à la modélisation du risque d'infection dans le sud de la France* » soutenu également en 2006 s'inscrivent dans cette dynamique.

La cartographie de la distribution spatiale des espèces peut également être le résultat de l'inversion d'une fonction mathématique. La fonction d'une courbe d'ajustement d'un phénomène spatial à partir de certaines de ces variables permet la production de la carte de distribution spatiale par inversion de la fonction. Les modèles iconiques sont un moyen d'exprimer dans l'espace un modèle rhétorique ou un modèle mathématique.

Il convient de remarquer à l'issue de cet exposé que les objectifs et les méthodes de modélisation ne sont pas exclusifs en soi. Il appartient donc au modélisateur de faire des choix en fonction de sa

problématique, de ses objectifs, des moyens dont il dispose etc. Comment a-t-on fait la modélisation dans ce travail, c'est l'objet de la seconde section de ce chapitre.

III.2 LA DEMARCHE, METHODES ET MATERIELS

En rappel, ce travail a pour objectifs: de décrire l'organisation spatiale des deux formes moléculaires M et S d'*An. gambiae* au Burkina Faso, d'identifier les déterminants de cette organisation et de la modéliser. Un tel objectif implique de combiner des compétences relatives à des domaines divers, les uns liés à la thématique, les autres à la technique ou à la mathématique Sanders, (2001). Au plan thématique, cette étude met à contribution la biogéographie qui oriente et donne la perspective théorique et l'entomologie qui fournit l'information nécessaire à la connaissance des anophèles. Au plan technique, elle met à profit l'analyse spatiale, l'analyse de données statistiques et le système d'information géographique. L'approche pluridisciplinaire s'impose dans une telle situation. Mais si une telle approche est stimulante et enrichissante dans le processus de modélisation, une utilisation désordonnée des contributions des différentes disciplines scientifiques peut porter préjudice à la logique et à la compréhension du raisonnement. D'où la nécessité de cette section destinée à la présentation de l'articulation d'ensemble de la méthodologie et aux usages faites des méthodes et des techniques. Cette section est structurée en cinq paragraphes : le premier s'évertue à montrer la perspective et l'articulation de la recherche et les quatre autres s'appesantissent sur les méthodes utilisées à chaque étape du processus d'investigation.

III.2.1 DEMARCHE D'ENSEMBLE DE LA RECHERCHE

L'objectif de la démarche est de mettre en exergue les différentes articulations de la méthodologie et sa structure logique globale. L'insertion de cette étude dans une approche pluridisciplinaire, faisant intervenir deux groupes de disciplines, justifie davantage cette précision de la démarche.

L'utilisation de ces deux groupes de disciplines s'articule autour des quatre étapes suivantes :

1. la modélisation de la distribution qui s'inscrit dans une perspective mathématique empirique et hypothético - déductive. Sur les bases des connaissances à priori, le paradigme biogéographique en l'occurrence, trois hypothèses ont été formulées. La première postule de l'existence d'une structure dans l'organisation des populations d'*An. gambiae* s.s.. La seconde présume d'une contribution différentielle des facteurs écologiques à la distribution spatiale des populations. La troisième tend à privilégier le fait qu'à partir des facteurs les plus déterminants de la distribution des vecteurs, il est possible de formaliser la relation entre la distribution

- spatiale *An. gambiae* s.s. et l'environnement. Cette formulation des hypothèses constitue la première phase de la modélisation.
2. la perspective étant mathématique / empirique, il s'agit d'identifier sur des observations de terrain les régularités et de les exprimer sous forme simplifiée, sans en trahir l'esprit. Pour cela, il est indispensable de procéder à une conceptualisation qui permet d'identifier des objets pertinents et de procéder à la quantification, c'est-à-dire à associer à chaque objet une liste de variables qui permet de le décrire le plus exactement possible (voir paragraphe « univers de variables »). Il arrive souvent que des modèles soient disponibles et facilitent ce travail de conceptualisation. C'est le cas, dans ce mémoire, de l'utilisation des modèles d'interpolation et d'estimation dans l'ajustement des données aux besoins de l'étude, de l'emploi de modèles d'analyse des configurations et structures spatiales (ellipse de déviation standard, indice d'autocorrection de Morin etc.) et de la mise à contribution de certains modèles de calcul statistiques.
 3. la formalisation est la troisième étape et consiste en l'expression du modèle sous un langage déterminé. Deux types de langages sont visés dans le cadre de ce travail. Il s'agit du langage iconique et du langage mathématique. Les modèles issus de ce travail sont présentés sous forme de fonctions mathématiques et sous forme de cartes par inversion des fonctions mathématiques. L'analyse spatiale et la statistique multidimensionnelle/l'analyse de données sont les deux méthodes utilisées à cette étape.
 4. la dernière étape de cette démarche consiste en l'évaluation de la qualité du modèle. Il s'agit de contrôler la qualité du modèle et, si besoin, de procéder à son amélioration par retour sur une ou plusieurs étapes précédentes. Deux méthodes sont utilisées pour l'évaluation de la qualité du modèle : une évaluation à partir de l'analyse des résidus du modèle et une évaluation par comparaison des données produites par les modèles à des données empiriques externes aux modèles.

La mise en œuvre de cette démarche nécessite des données de bonne qualité. Encore faut-il avoir un état des besoins en données bien élaboré !

III.2.2 UNIVERS DES VARIABLES ET ENTITES / OBJETS GEOGRAPHIQUES

Les hypothèses et les indicateurs devant permettre de les vérifier, mettent en jeu un certain nombre de concepts. Il faut pouvoir les rendre opératoires à travers une définition de leurs caractéristiques. C'est le rôle de la variable que Jambou M., (1999), définit comme « toute caractéristique d'une personne ou d'une chose qui peut être exprimée par un nombre ». L'univers des variables est

l'ensemble des informations visant à caractériser le plus possible les phénomènes étudiés. Il répond à la question suivante : quelles sont les informations nécessaires pour renseigner les indicateurs ? Ce qui implique aussi de s'interroger sur les objets sur lesquels les informations doivent être collectées (objets géographiques, périodes, acteurs).

L'objectif de l'étude est, rappelons-le, de modéliser la distribution spatiale des membres du complexe *An. gambiae* au Burkina Faso pour comprendre et faire comprendre la distribution des membres de ce complexe d'espèces à travers l'identification des facteurs environnementaux les plus déterminants. En d'autres termes, nous cherchons à expliquer la variabilité des fréquences dans des lieux donnés, compte tenu de la variabilité de l'environnement.

En nous appuyant sur le paradigme biogéographique, la bio-écologie du complexe *An. gambiae* (section I, chapitre II), nous construisons le diagramme de cause à effet (ou diagramme en arête de poisson ou encore diagramme Ishikawa), pour inventorier l'ensemble des variables à prendre en compte (Figure 15).

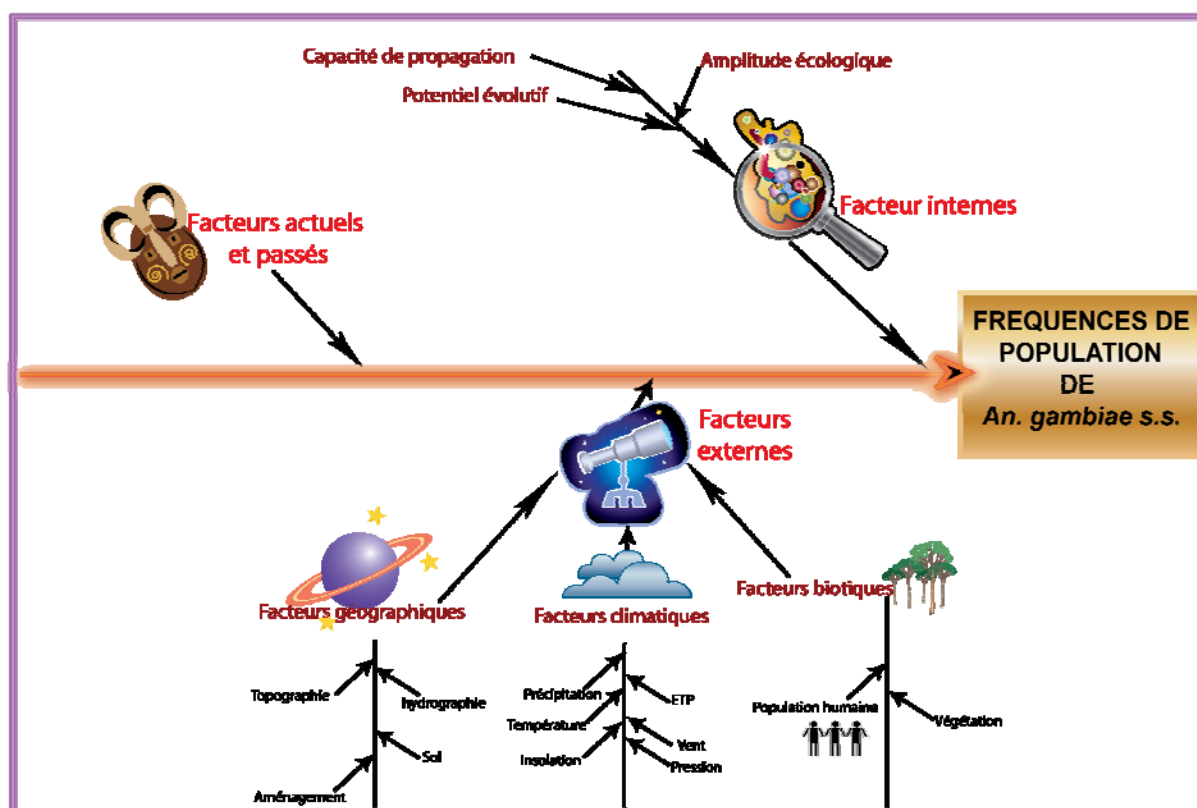


Figure 15 : Diagramme d'Ishikawa des facteurs environnementaux influençant les fréquences d'*Anopheles gambiae* s.s., réalisation SOME Y. S. Corentin

Ce diagramme présente trois grands groupes de facteurs qui contribuent à la distribution spatiale des espèces : les facteurs actuels et passés, les facteurs internes à l'espèce et les facteurs externes à l'espèce.

Dans le cadre de cette étude, l'intérêt est porté sur les facteurs externes. Ces derniers constituent le milieu dans lequel vivent les membres du complexe. Ces facteurs externes sont les facteurs environnementaux.⁵⁰ Ils peuvent être subdivisés en trois classes : les facteurs géographiques, les facteurs climatiques et les facteurs biotiques.

III.2.3 L'INVENTAIRE DES VARIABLES

D'après l'état actuel de l'art sur la bio-écologie des anophèles (voir chapitre précédant), le climat à travers des paramètres tels que la température, les précipitations et l'insolation, conditionne énormément le taux et la durée d'éclosion des œufs. Il influence donc le potentiel reproductif des anophèles. Il est donc d'un intérêt certain de retenir ces paramètres importants comme variables. Aussi, la température, les précipitations, les pressions, l'insolation, l'évapotranspiration potentielle (ETP) et les vents ont-ils été retenus comme variables sur le plan climatique.

Au niveau géographique, le relief est retenu comme un facteur limitant à partir de certaines altitudes. La ponte se faisant dans les eaux de surface, la disponibilité et l'accessibilité à cette ressource par les anophèles est de ce fait d'une grande importance. Les hommes, à travers l'aménagement du territoire, surtout en matière de mobilisation des ressources en eau, peuvent être en densité très élevée par endroit. Dans les implantations humaines, les aménagements peuvent créer des conditions locales, quelques fois temporaires, favorables à la prolifération des anophèles. Il semble que la composition chimique des eaux,⁵¹ en particulier leur pH intervient dans la distribution des vecteurs. L'hydrographie, l'aménagement, la topographie (altitude) et les sols constituent des variables à retenir comme facteurs géographiques.

Les anophèles ont besoin de sang humain pour la maturation de leurs œufs. Les adultes se nourrissent également de jus sucré prélevé sur la végétation qui leur sert également de site de repos et de protection contre les prédateurs. La présence humaine et la végétation constituent des facteurs biotiques retenus comme variables.

En somme, les variables retenues sont au nombre de 14, réparties en deux groupes : les variables à expliquer et les variables explicatives. Les variables à expliquer sont au nombre de deux : les fréquences d'*An. gambiae* forme moléculaire M et les fréquences d'*An. gambiae* forme moléculaire

⁵⁰ Utilisation équivalente de facteurs éco-géographiques et de facteurs environnementaux.

⁵¹ Il n'a pas été possible de réunir l'information spatiale sur la qualité des eaux et sur la chimie des sols.

S, chacune exprimant la population d'une forme moléculaire dénombrée en un lieu. Ce sont les variables entomologiques. Les variables explicatives sont les variables environnementales ou écologiques. Nous les appelons aussi variables éco-géographiques. Elles peuvent être classées en trois sous groupes :

- les variables climatiques qui regroupent la température, les précipitations, l'ETP, les vents, l'insolation, la pression de vapeur ;
- les variables géographiques qui réunissent l'aménagement, l'hydrographie, le relief, les localités ;
- les variables biotiques constituées de la présence humaine et de la végétation.

Certaines d'entre elles sont d'évaluation si courante que leurs unités sont connues et les méthodes de leur évaluation ne sont pas nombreuses. D'autres ont des contenus d'interprétations peu évidentes par rapport à l'objectif de l'étude. D'où l'intérêt de préciser les unités. Dans le Tableau 3 sont consignées les variables retenues ainsi que leurs contenus.

Tableau 3 : Les variables, leurs unités de référence et leurs définitions

Variables	Unités	Contenus	Apport dans l'écologie du vecteur
Température	Degré Celsius	Degré de chaleur en un lieu	Limite de température tolérable, qualité des gîtes
Précipitation	Millimètre	Hauteur d'eau tombée dans un lieu	Quantité et qualité de gîtes larvaires
Insolation	Fraction	Pourcentage de temps d'ensoleillement au cours d'une journée En temps totalement couvert, la fraction d'ensoleillement est égale à 0	Qualité des gîtes larvaires
Vent	Mille/second	Vitesse du vent	Qualité des sites de repos
ETP	Millimètre	La vapeur d'eau potentiellement produite du fait de la transpiration et de l'évaporation	Limite d'aridité tolérable (indicateur d'humidité de l'atmosphère)
Pression	Hectopascal	Tension de vapeur	Qualité de l'atmosphère
Hydrographie	Mètre	Distance en tout point des cours d'eau. Elle exprime l'accessibilité à l'eau	Disponibilité de site des gîtes larvaires
Relief	Mètre	Altitude (hauteur par rapport au niveau moyen de la mer)	Difficulté de survie à partir de certaines altitudes, effet sur les températures et les pressions
Aménagement	Type d'occupation des terres	La diversité de l'occupation des terres	Disponibilité de gîtes larvaires de ponte, en gîtes de repos et repas sanguin
Localités	Mètre	Accessibilité par rapport aux implantations humaines	Disponibilité en repas sanguin
Végétation	Proche infrarouge/rouge	Intensité de la reflectance chlorophyllienne	Lieu de repos, alimentation des adultes surtout mâle
Présence humaine	Nombre d'habitants/kilomètre carré	Densité de la population	Disponibilité du repas sanguin
Population de M	Nombre de M	Nombre de M en un lieu donné	Expression de l'adéquation aux exigences écologiques de M
Population de S	Nombre de S	Nombre de S en un lieu donné	Expression de l'adéquation aux exigences écologiques de S

III.2.4 ENTITES / OBJETS GEOGRAPHIQUES

Où et sur quoi évaluer les variables inventoriées ? C'est l'objet de ce paragraphe. L'habitat est l'expression par excellence des relations entre une espèce et son environnement. L'habitat est donc l'unité élémentaire de dénombrement des fréquences de vecteurs qui cadre avec l'environnement. Cependant, l'habitat des vecteurs varie au cours d'une journée en fonction de l'état physiologique et des préférences écologiques (endophilie, exophilie, ponte, repos pour maturation des œufs, etc.).

La capture des anophèles est réalisée par maison, un des habitats préférentiels des anophèles. La maison correspond à une habitation humaine. Elle est le premier niveau géographique de dénombrement des anophèles capturés. La maison est un objet géographique appartenant à une localité.⁵² Les moustiques dénombrés au niveau des maisons d'une même localité sont cumulés au niveau de la localité. La sommation des captures permet la caractérisation la population d'anophèles au niveau des localités. Le niveau géographique d'analyse de la population d'anophèles est donc la localité. La localité est à une résolution plus fine que le village qui est l'entité administrative de base de l'organisation du territoire burkinabé.

La caractérisation de l'environnement des anophèles dans une localité nécessite également la définition d'une unité de référence. Cette unité peut acquérir une grande performance si elle s'appuie sur une échelle spatiale d'intégration des écosystèmes pour reprendre les termes de Lacoste et Salanon, (2006). Par référence au niveau spatial pertinent de collecte de données climatiques, le niveau microécosystémique correspond au niveau stationnel où interagissent et peuvent être caractérisés tous les facteurs climatiques Lacoste et Salanon, (2006). Ce niveau doit répondre au critère fondamental d'homogénéité. Il doit également être en adéquation le plus possible avec l'habitat des vecteurs. Pour ce faire, un carré d'un kilomètre de côté, soit une superficie de 1 km², a été retenu comme surface de référence pour la caractérisation de l'environnement. Cette superficie correspond à l'aire de compétence des vecteurs et est suffisamment réduite pour garantir une homogénéité confortable. L'objet géographique de référence est donc une cellule de 1km sur 1km.

⁵² La localité correspondant à tout regroupement humain, doté ou non d'un statut juridique au niveau national. Elle est différente du village qui lui est une implantation humaine reconnu comme tel par l'administration publique.

En somme, deux entités/objets géographiques par type de données, sont retenues :

1. pour les données entomologiques, l'objet géographique de référence pour la capture des moustiques est la maison. La maison est donc le niveau géographique de collecte. L'objet géographique pour l'analyse des informations entomologiques est la localité. Le passage du niveau maison au niveau localité est basé sur la sommation
2. Pour les données environnementales, l'objet géographique pour l'observation est la station. L'environnement est défini sur une maille de 1km sur 1km ; les variables associées à cette entité sont les suivantes : pluie, température, ETP, pression, vent, insolation, relief, hydrographie, aménagement, localité, population humaine, végétation. L'objet géographique pour l'analyse des données environnementales est la maille de 1km². Le passage de la station à la maille de 1km sur 1km, est basée essentiellement sur l'interpolation et l'estimation. L'objet géographique de référence est une maille de 1km² implantée en polygone (pixel).

III.2.5 LES METHODES DE COLLECTE DE DONNEES

Deux types de données sont collectés :

- 1- les données primaires correspondant aux données entomologiques ;
- 2- les données secondaires constituées des données éco-géographiques.

Deux méthodes sont utilisées pour la collecte de ces données : le transect et la consultation documentaire.

Pour réunir les fréquences des différents membres du complexe *An. gambiae*, le recours à une enquête géographique par sondage s'est imposé compte tenu du protocole du projet, de la durée de l'étude et des ressources. Cet échantillonnage s'est appuyé sur la méthode du transect.

III. 2.5.1 Le transect

En vue de la collecte des données entomologiques, un transect long de 3 183 km et de 15 km de large de part et d'autre de l'axe du transect est défini. Il couvre une superficie totale de 107 875 km² sur 274 200 km², soit plus de 35% du territoire burkinabé. Il recoupe l'ensemble de la diversité écologique de la zone d'étude. Ce transect permet de sélectionner 5 150 sur 10 044⁵³ localités que compte le Burkina Faso, soit plus de 50% de la population statistique.

Ce transect est subdivisé en trois sections ou circuits : le premier circuit, long de 547 km, le second de 1 303 km et le troisième de 1 333 km. Ce transect constitue la première étape d'un échantillonnage aléatoire spatial stratifié.

⁵³ Les localités se définissent comme toute implantation humaine sans lien avec son statut juridique (village, ville, hameau de culture lieu dit etc.)

La seconde phase consiste en un maillage du site, avec des cellules de 5 x 5 km⁵⁴, soit 11 321 mailles. Ces mailles sont indexées. Parmi elles, 4 315 répondent au critère de distance par rapport à l'axe du transect. De celles-ci 300 sont tirées au sort. Le tirage est stratifié de manière proportionnelle à la longueur du circuit. Ainsi, sur le circuit de 547 km, 50 cellules sont tirées au sort et sur les deux autres circuits, 125 cellules sont choisies au hasard.

La troisième phase consiste à tirer au sort, une localité et une seule, dans chacune de 300 cellules. Les coordonnées de ces localités sont remises à l'équipe d'entomologistes du Centre National de Recherche et de Formation sur le Paludisme (CNRFP) en vue de la capture et du dénombrement d'*An. gambiae* s.s.. Les méthodes de capture et d'identification des formes moléculaires d'*An. gambiae* sont présentées en annexe 1 et 2. La Figure 16 présente les différents aspects de cet échantillonnage. Les captures et les identifications ont été effectuées par l'équipe d'entomologique du CNRFP.

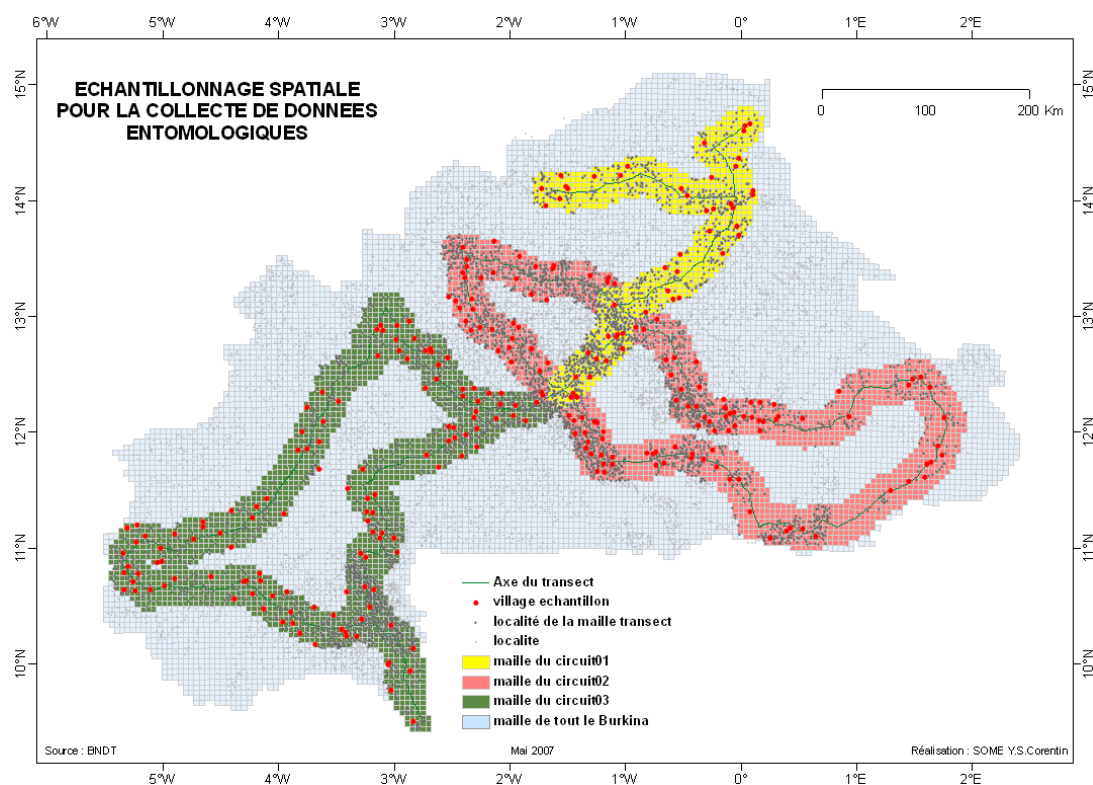


Figure 16 : Transect, circuits et points de collecte des données entomologiques

III. 2.5.2 La revue de littérature scientifique

La revue de littérature est utilisée pour faire le point sur les données relatives aux variables éco-géographiques, à la méthodologie, à la bio-écologie des vecteurs et aux différentes techniques de

⁵⁴ Certains paramètres liés à cet échantillonnage ont été définis dans le document du projet NIH. Il s'agit du nombre de sites de dénombrement (300) et de la taille des sites de prélèvement (5km x 5km).

traitement de l'information spatiale et statistique. Elle consiste essentiellement en l'inventaire des bases de données (bases de données géographiques/cartographiques, base de données climatiques), en la lecture d'un certain nombre de documents (ouvrages de référence, thèses, manuels, revues scientifiques, etc.), en des discussions avec les experts sur des thèmes très variés (biologie et écologie des vecteurs, méthodologie en modélisation, analyse spatiale, biogéographie, etc.).

III.2.6 EVALUATION DE LA QUALITE DES DONNEES

La perspective de la modélisation est de partir d'observations de terrain pour formaliser le phénomène et s'en servir comme moyen de son explication et de sa compréhension. C'est donc une perspective empirique de la modélisation basée sur l'observation du réel pour reprendre les termes de Langlois et Reguer, (2005). Dans ces conditions, la qualité des données revêt une importance toute particulière. Elle détermine, en effet, la qualité du modèle.

L'objectif de cette analyse de la qualité est d'apprécier, le plus objectivement possible, la capacité des données collectées, particulièrement les données secondaires, à répondre aux exigences de l'analyse. Il ne s'agit nullement de porter un jugement de valeur absolue sur la qualité liée à la production des données, tant les objectifs de la production peuvent être très éloignés des nôtres et justifier certaines décisions techniques. Le but de cet exercice est de répondre aux questions essentielles suivantes : est-ce que les données collectées sont adaptées aux besoins ? Sinon, est-il possible de les traiter pour les adapter ? Dans ce cas, quels sont les traitements applicables ?

III.2.6.1 Méthode d'évaluation utilisée

La norme ISO 19113 résumée par Devilliers et Jeansoulin, (2005), portant sur la qualité de l'information géographique, fournit une clé intéressante à l'appréciation de la qualité interne et externe des données dans les SIG.

Sur le plan de la qualité interne des données, c'est-à-dire le niveau de similitude entre les données collectées et celles « parfaites qui auraient dues être collectées » et appelées encore « terrain nominal » ou « univers du discours », nous nous intéressons à :

1. l'exhaustivité qui se rapporte à la présence ou à l'absence d'un élément, d'un attribut ou d'une relation ;
2. la cohérence logique qui s'intéresse à la pertinence par rapport à des règles logiques de structure de données, d'attributs ou de relations ;

3. l'exactitude spatiale relative à la véracité du positionnement de l'objet dans l'espace ;
4. l'exactitude temporelle qui revisite les attributs temporels et les relations temporelles des objets ;
5. l'exactitude des attributs qui fait allusion aux attributs quantitatifs, à la justesse des attributs qualitatifs et à la classification des objets et des relations.

Pour la qualité externe des données, il s'agit plutôt d'une appréciation contextuelle des données. Elle s'intéresse à l'adéquation, à une utilisation particulière et intègre des paramètres tels que le rapport coût/qualité, temps de production/durée de l'étude. C'est comme le définit la norme ISO, « l'ensemble des propriétés et caractéristiques d'un produit ou service qui lui confère l'aptitude à satisfaire des besoins implicites ou explicites ».

L'évaluation de la qualité externe peut être réalisée aussi sur un canevas proposé par Bedard et Valliere, (1995), in Rodolphe et Robert, (2005), et qui s'appuie sur six paramètres :

1. la définition c'est-à-dire l'adéquation de la nature des données et des objets aux besoins ;
2. la couverture ou la compatibilité de l'extension spatiale et temporelle de la donnée aux besoins ;
3. la généalogie, c'est la correspondance entre l'origine des données, les méthodes de production et les besoins ;
4. la précision sémantique, temporelle et spatiale des objets et des attributs par rapport aux besoins ;
5. la légitimité qui exprime la reconnaissance officielle, la portée légale et le respect des normes de la donnée par rapport aux besoins ;
6. l'accessibilité, c'est-à-dire la facilité d'accès aux données (coût, délais, format, confidentialité, droit d'auteur).

L'application d'une telle grille d'évaluation requiert la définition d'un « terrain nominal » qui représente les données attendues.

III.2.6.2 Terrain nominal

L'idéal est de réunir sur les lieux de prélèvement, soit sur les 300 points de prélèvements, toutes les caractéristiques du milieu et du complexe *An. gambiae*. Les données relatives aux différentes variables doivent se superposer parfaitement au plan spatial pour permettre la comparabilité. Au plan temporel, la caractérisation du milieu doit être faite sur tous les 300 sites en même temps. L'ensemble des informations doit être cohérent avec le site par conformité à la description officielle

et à la littérature scientifique. La précision temporelle est l'heure avec une répétitivité, toutes les semaines à la même heure de prélèvement. Le choix de la semaine s'appuie sur le cycle gonotrophique qui est d'un maximum de 7 jours. Cette répétitivité devrait permettre de suivre la variabilité spatio-temporelle de population des vecteurs et du milieu au cours d'une année. La précision spatiale est le kilomètre. En effet, le kilomètre correspond à la distance de vol de l'anophèle. Les sources, à moins d'être produites dans le cadre direct de cette étude, doivent provenir au moins de structures officielles.

III.2.6.3 Appréciation de la qualité des données

Ce terrain nominal permet d'apprécier les données et de prendre les décisions en matière de recevabilité, compte tenu de la réalité objective de terrain (contrainte temporelle, logistique, budgétaire etc.), et le cas échéant, de décider des traitements visant à rapprocher le plus possible les données au terrain nominal.

Le Tableau 4 présente l'appréciation des informations collectées par variable et selon les différents paramètres retenus dans l'inventaire des variables en III.2.3 et le terrain nominal.

Du point de vue de la qualité interne, les données ont une bonne cohérence logique, une bonne exactitude spatiale et une bonne exactitude des attributs. Elles ont cependant d'énormes lacunes en termes d'exhaustivité et d'exactitude temporelle. Les données relatives aux facteurs climatiques sont des données historiques et sont collectées à travers un réseau de stations totalement différentes des lieux d'observations. Il en est de même pour ce qui concerne les données relatives aux facteurs géographiques. Ce sont des informations historiques agrégées qui ont parfois plus de 20 ans. C'est le lieu de citer les informations sur l'hydrographie et la topographie (altitude) qui relèvent des cartes topographiques et dont les plus récentes datent de 1984. L'ensemble des données ne répond donc pas à l'exhaustivité et à l'exactitude temporelle.

Sur le plan de la qualité externe, la limite observable sur les données concerne particulièrement la couverture et la précision temporelle et spatiale.

Pour rapprocher le plus possible le terrain nominal et la nature des données collectées, deux types de traitement sont appliqués: le premier vise à résoudre la question de la temporalité et le second apporte la solution à la question de la spatialité.

Tableau 4 : Evaluation de la qualité des données collectées

Paramètres	Données à évaluer													
	Précipitation	Température	vent	insolation	pression	ETP	Hydrographie	Topographie (Altitude)	Aménagement	localité	population humaine	végétation	population de M	population de S
	Qualité Interne													
Exhaustivité	Non	Non	Non	Non	Non	Non	Non	Non	Non	Non	Non	Non	Oui	Oui
Cohérence logique	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui
Exactitude spatiale	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui
Exactitude temporelle	Non	Non	Non	Non	Non	Non	Non	Non	Non	Non	Non	Non	Non	Non
Exactitude d'attributs	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui
	Qualité externe													
Définition	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui
Couverture	Non	Non	Non	Non	Non	Non	Non	Non	Non	Non	Non	Non	Non	Non
Généalogie	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Non	Oui	Non	Oui	Oui
Précision	Non	Non	Non	Non	Non	Non	Non	Non	Non	Non	Non	Non	Non	Non
Légitimité	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui
Accessibilité	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Non	Non

III.2.7 LES METHODES DE TRAITEMENT DES DONNEES

Les méthodes de traitement recouvrent l'ensemble de opérations appliquées aux données pour les rapprocher le plus possible du terrain nominal. Les principales lacunes observées relèvent de l'exactitude et de la précision temporelle, de la couverture et de la précision spatiale. C'est pour résoudre les questions de temporalité et de spatialité que ces traitements sont appliqués. Ceux-ci peuvent être regroupés en types de traitements : les traitements de type spatial et les traitements de type temporel.

III.2.7.1 les traitements de type spatial

Les données disponibles ne couvrent pas l'ensemble des sites et présentent aussi des référentiels techniques différents. Pour résoudre ces problèmes, nous appliquons des méthodes de transformation géométrique/topologique et des méthodes d'interpolation et d'estimation spatiales. Ces traitements sont appliqués lorsque les données ont des référentiels techniques hétérogènes ou lorsqu'elles ne couvrent pas l'ensemble des points échantillonnés (Cf. tableau 4).

Les méthodes de transformation géométrique

Elles consistent à inscrire l'ensemble de données dans un même système de projection. Et pour conserver la légitimité, toutes les données spatiales ont été inscrites dans le référentiel cartographique officiel du Burkina Faso : Ellipsoïde Clark 1880, Projection Univers Transverse Mercator (UTM), Zone 30 étendue à 31.

Les données spatiales, n'ayant pas de système de coordonnées ont fait l'objet de géoréférencement et de transformation. Pour les transformations et le géoréférencement, la méthode de transformation affine est utilisée. La validation de la transformation est réalisée pour une erreur quadratique moyenne très faible ($\leq 0,04$). Les données ayant déjà un système mais différent du référentiel officiel ont été re-projetées.

Les méthodes d'interpolation et d'estimation

Ces méthodes sont utilisées pour combler les lacunes constatées en termes de couverture spatiale. La plupart des données étant collectées en certains points échantillonnés, les valeurs prises en des points non échantillonnés sont souvent ignorées. Le but de l'utilisation de ces méthodes est donc d'obtenir des valeurs en tout point du site, soit par estimation soit par interpolation.

Cinq méthodes d'interpolation ont été utilisées : Global Polynomial Interpolation (GPI), Inverse Distance Weighthing (IDW), Local Polynomial Interpolation (LPI), Radial basis fonction (RBF) et Krigeage. Certaines sont exactes et déterministes comme IDW, d'autres sont stochastiques comme le Krigeage. Pour chaque variable considérée, nous appliquons par essais successifs les cinq méthodes puis nous choisissons celle qui a l'erreur prédite la plus faible. La résolution retenue pour l'interpolation est de 1km sur 1km, correspondant approximativement à l'aire de compétence en un vol des anophèles.

L'interpolation est utilisée pour les altitudes ainsi que pour les différents paramètres climatiques, température, précipitation, vent, pression, insolation, ETP, Arnaud et Emery, (2000); Claritas, (2002); Droesbeke et *al.*, (2006); ESRI, (2006); Zaninetti, (2005)

Deux autres méthodes ont été utilisées pour pallier les insuffisances liées à la couverture et à l'exhaustivité. Il s'agit du calcul des densités et du calcul des distances. L'estimation des distances a été faite pour deux variables : l'hydrographie et les localités. Cela consiste à calculer la distance en tous points, par rapport aux cours d'eau, et la distance en tout point par rapport aux localités. Le type de distance utilisé est linéaire. La maille est un carrée de 100 m de côté. La réduction de la taille de la maille à 100 m, donc l'augmentation de la résolution spatiale permet de prendre en compte un minimum de diversité par rapport au centre de collecte.

La méthode d'estimation des densités est utilisée pour la variable population. Elle consiste à répartir des valeurs (valeurs mesurées) attachées à un ensemble de points sur une maille. La fonction de répartition est, soit l'égalité, soit la proportionnelle à la distance. Nous obtenons ainsi une population par unité de surface.

Dans le cadre de cette étude, les paramètres utilisés sont les suivants :

Type de maille : carré

Taille des mailles de la grille : 1km sur 1km

Type de densité : pondérée à la distance

Rayon de compétence : 10 000 m

Unité de surface : km²

Ces traitements ont permis d'approcher le terrain nominal en proposant par estimation les caractéristiques concernées au niveau spatial. Si la question de la couverture et de l'exhaustivité spatiale trouve ainsi une solution acceptable, la question de l'exactitude et de la précision temporelle reste à traiter.

III.2.7.2 Les traitements de type temporel

La simultanéité souhaitée pour les variables éco-géographiques et les fréquences des vecteurs et la répétitivité de la caractérisation constituent un niveau de précision et d'exactitude temporelle extrêmement difficile à atteindre. Aussi, le recours à des valeurs moyennes caractérisant les différents facteurs est-il une solution !

L'usage des données les plus récentes

Certains paramètres connaissent une grande stabilité et des mesures à forte répétitivité dans le temps n'apporteront pas de grands changements. C'est le cas du relief, de l'hydrographie et des localités. C'est pourquoi, pour ces variables, les données n'ont subi aucun traitement au niveau temporel. D'autres sont d'estimation tellement coûteuse que leur production est difficilement envisageable à la périodicité souhaitée. C'est le cas des statistiques sur la population et de la cartographie de l'occupation des terres. Pour ce qui est de la population, une estimation a été faite pour l'année de collecte des données entomologiques à partir du taux d'accroissement naturel. Pour ce qui concerne l'occupation des terres, la donnée cartographique la plus récente a été retenue comme significative pour la période de collecte des données entomologiques.

Les méthodes statistiques

Concernant les autres variables, pour lesquelles il existe une production journalière des informations et qui intéressent les facteurs climatiques et la végétation, nous considérons les états moyens. Pour les facteurs climatiques, le recul statistique nécessaire est de 30 ans. Cette condition n'a pas toujours pu être respectée. Dans de tels cas, nous prenons en compte le recul maximum pour les stations disponibles.

Ainsi, lorsque la série le permet, les valeurs moyennes pour tous les paramètres ont-elles été calculées sur 30 ans. Sinon, l'ensemble de la série est pris en compte. Pour ce qui concerne la végétation, la moyenne des indices décennaux normalisés de végétation, calculée sur l'ensemble de la période de collecte qui s'étend de juillet à octobre 2005, a été retenue.

En somme, sur le plan temporel, la solution a été d'utiliser les caractéristiques moyennes. Cet état moyen a été obtenu soit par calcul soit par prise en compte des caractéristiques propres de la variable considérée.

III.2.8 MATERIELS / LOGICIELS UTILISES

Le déploiement de cette méthodologie a été possible grâce à l'utilisation d'un certain nombre de logiciels. Nous regroupons les logiciels en cinq classes : SIG et analyse spatiale, télédétection/traitement numérique d'images, analyse de données, modélisation et gestion des GPS.

Logiciels de SIG et d'analyse spatiale.

Ils ont servi à l'application des traitements de type spatial sur les données : transformation géométrique, projection, interpolation et estimation. Ces logiciels ont également été utilisés pour le calcul de certains indicateurs choisis pour la vérification des hypothèses. Il s'agit des indices d'autocorrélation spatiale et des indices locaux d'association spatiale. Après la définition des modèles de répartition en fonction des paramètres environnementaux, ces logiciels ont permis d'inverser les fonctions et de passer des modèles mathématiques aux modèles iconiques. Certains éléments de géométries ont été également calculés comme la distance moyenne entre les localités. En sus de ces fonctions de traitement et de calcul, les logiciels de SIG sont également utilisés comme logiciel de cartographie en permettant l'édition cartographique à partir des données qui ont été collectées et restructurées en bases de données. Les logiciels utilisés sont : *Arcview 3x* avec l'extension *Spatiale Analyst de ESRI* et extension complémentaire et *Arcgis 9 x* avec les extensions *Spatiale Analyst* et *Géostatistical Analyst de ESRI*. La raison de l'utilisation des deux logiciels est liée au fait que nous avons démarré le processus avec la version 3X avant d'avoir accès à la version supérieure. De plus, parce que la version 3X existait depuis longtemps, elle avait une plus grande disponibilité en *shareware* avec les fonctions très utiles dans les traitements et inexistantes dans la version supérieure.

Logiciels de télédétection / traitement numérique d'image

Les logiciels de télédétection ont été utilisés principalement pour le traitement des images NDVI. Ces traitements ont consisté à géoréférencer les images, à les découper et à calculer un NDVI moyen sur 12 images obtenues. L'usage également de deux solutions logicielles est lié également à des fonctionnalités spécifiques que nous recherchions. Ainsi, le géoréférencement et le découpage des images ont été effectués avec le logiciel *ERDAS Imagine*. Pour le calcul de l'images arithmétique correspondant au NDVI moyen sur les 12 images, il a été effectué sur *Winchips* qui offre un module de calcul très simples et très accessibles.

Logiciels d'analyse de données

Les logiciels d'analyse de données ont servi à la fouille des données. Il s'agit principalement des traitements liés à l'ACP, aux différentes régressions, à la réalisation de certains tests et calculs comme le test de Kolmogorov-Smirnov et le calcul de pourcentage des différentes formes moléculaires par site de capture. Les logiciels utilisés sont XL-Stat 2003, en extension à Microsoft Excel 2003, de Microsoft Corporation et Tanagra de Université de Lyon II.

Logiciel de modélisation

Ce logiciel a servi à l'élaboration des modèles conceptuels de données et de traitement. C'est le logiciel *Power AMC-Designor de Sybase* qui a été utilisé.

Logiciels pour GPS

La collecte de données entomologiques nécessitait que l'équipe de terrain disposât des coordonnées géographiques sur les localités échantillons pour y aller. De plus, de retour du terrain, il fallait également télécharger les données du GPS aux ordinateurs. Pour cela, les logiciels suivants ont été mis à contribution : *GPS track maker*, de Odilon Ferreira Junior, Individual resident in Belo Horizonte – MG – Brazil, et *Mapsource de Garmin*.

CONCLUSION DU CHAPITRE III

L'approche méthodologique utilisée dans le cadre de cette étude est pluridisciplinaire. La biogéographie y est positionnée comme discipline charnière de formulation des hypothèses. Elle a permis d'intégrer des résultats fournis par l'utilisation des apports d'autres disciplines scientifiques et techniques. Les méthodes utilisées peuvent être regroupées en trois grandes classes :

- La collecte des données, effectuée à travers deux types de méthodes que sont la revue de littérature et le transect,

- L'analyse de qualité des données qui appuie sur la norme ISO 19113 reformulée par Devilliers et *al.*, (2005) avec pour objectif, l'identification du terrain nominal à partir d'une connaissance *a priori* des besoins en information et en traitement,
- Les méthodes de traitement qui regroupent les méthodes de traitement spatial et a-spatial et les méthodes de calcul des indicateurs.

Ces méthodes se sont appuyées essentiellement sur la statistique/l'analyse de données, la télédétection, l'analyse spatiale et le système d'information géographique. Les traitements fournissent des résultats dont l'interprétation permettra la vérification des hypothèses de départ, le niveau d'atteinte des objectifs spécifiques.

CONCLUSION DE LA PREMIERE PARTIE

Cette première partie a permis de délimiter le champ d'investigation de ce travail de recherche. Elle l'a positionné comme un travail de géographie et plus précisément de biogéographie qui procède d'une approche pluridisciplinaire pour contribuer à la réflexion sur une question de santé, celui de la répartition spatiale des principaux vecteurs du paludisme au Burkina Faso, les formes moléculaires M et S d'*Anopheles gambiae*. Sans renier son caractère géographique, ce travail a bénéficié des contributions indispensables de l'entomologie, de l'analyse des données, de l'analyse spatiale et des systèmes d'information géographique pour la mise en place d'une démarche méthodologique qui permet à chaque méthode d'apporter sa contribution dans le processus d'investigation. Cette étude ambitionne de modéliser la distribution spatiale d'*Anopheles gambiae* s.s. c'est-à-dire d'en donner une représentation formelle et fidèle.

Les contours de l'étude sont définis au plan scientifique. Le champ d'application est connu à travers une présentation de l'anophèle et du Burkina Faso. La démarche méthodologique et l'ensemble du processus ont été exposés. A quels résultats aboutit-on ? Quelle interprétation faire de ces résultats ? Quelles appréciations donner par rapport aux hypothèses et aux objectifs de ce travail. C'est le contenu de la seconde partie de ce travail.

DEUXIEME PARTIE

STRUCTURATION DES DONNEES, ELABORATION ET EVALUATION DES MODELES

Cette partie fait une synthèse sur les données collectées et leurs traitements. Elle fait aussi le point de l'organisation spatiale d'*An. gambiae* s.s. au Burkina Faso et des facteurs environnementaux qui déterminent les structures et configurations spatiales mises en exergue.

Aussi, a-t-elle pour but d'évaluer les indicateurs pour vérifier les hypothèses de recherche et apprécier le niveau d'atteinte des objectifs. Elle est composée de trois chapitres :

- le chapitre 4 est consacré à la présentation de la base de données à travers les modèles conceptuels de données et de traitement, ainsi que les traitements appliqués aux données éco-géographiques /environnementales ;
- le chapitre 5 s'intéresse à l'organisation spatiale des formes moléculaires M et S d'*An. gambiae* et l'analyse des facteurs déterminants de la distribution spatiale ;
- le chapitre 6 est destiné à la modélisation de la distribution spatiale des formes moléculaires M et S d'*An. gambiae* au Burkina Faso.

CHAPITRE IV : ORGANISATION ET STRUCTURATION DE LA BASE DE DONNEES

Recueillir des données est une chose, les structurer en fonction de la logique d'un domaine, des objectifs de stockage, de l'accessibilité et des traitements, en est une autre. La collecte de données, même en suivant une méthodologie bien définie, n'est pas suffisante pour en assurer une bonne organisation et une bonne structuration. Cela est particulièrement vrai en bases de données numériques où des contraintes techniques imposent des logiques de conception de la structure organisationnelle des données et des traitements qui s'appuient sur un formalisme bien défini. L'utilisation efficace et efficiente des SIG requiert qu'on s'intéresse à la structuration des données et des traitements. Aussi, ce chapitre est-il consacré à la modélisation des données ainsi qu'à leur traitement. Le formalisme utilisé dans le cadre de cette étude est MERISE. Deux modèles ont été réalisés : le modèle conceptuel de données et le modèle conceptuel de traitement. La présentation suivante est structurée selon les deux modèles auxquels s'ajoute le traitement des données éco-géographiques.

IV.1 LE MODELE CONCEPTUEL DE DONNEES

Le processus d'opérationnalisation de la méthodologie a conduit à la collecte d'un ensemble de données qu'il est nécessaire de bien structurer pour pouvoir en faire une exploitation optimale à l'aide des systèmes de gestion des bases de données (logiciel). Cela a conduit à l'élaboration d'un modèle conceptuel de données.

Le modèle conceptuel de données (MCD) « représente la structure logique globale d'une base de données, indépendamment du logiciel ou de la structure de stockage des données. Un modèle conceptuel contient toujours des données qui constituent une représentation formelle des données nécessaires au fonctionnement d'une entreprise. Il permet de représenter l'organisation des données sous forme graphique, de vérifier la validité des données de conception et générer le modèle physique de données, qui spécifie les modalités physiques de mise en oeuvre de la base de données »⁵⁵. Le MCD est une représentation du domaine étudié à travers l'identification de des objets qui le constituent et des relations fonctionnelles. La Figure 17, ci-après, présente la modélisation des données utilisées dans le cadre de ce travail de recherche.

⁵⁵ Aide PowerAMC Version 6.1

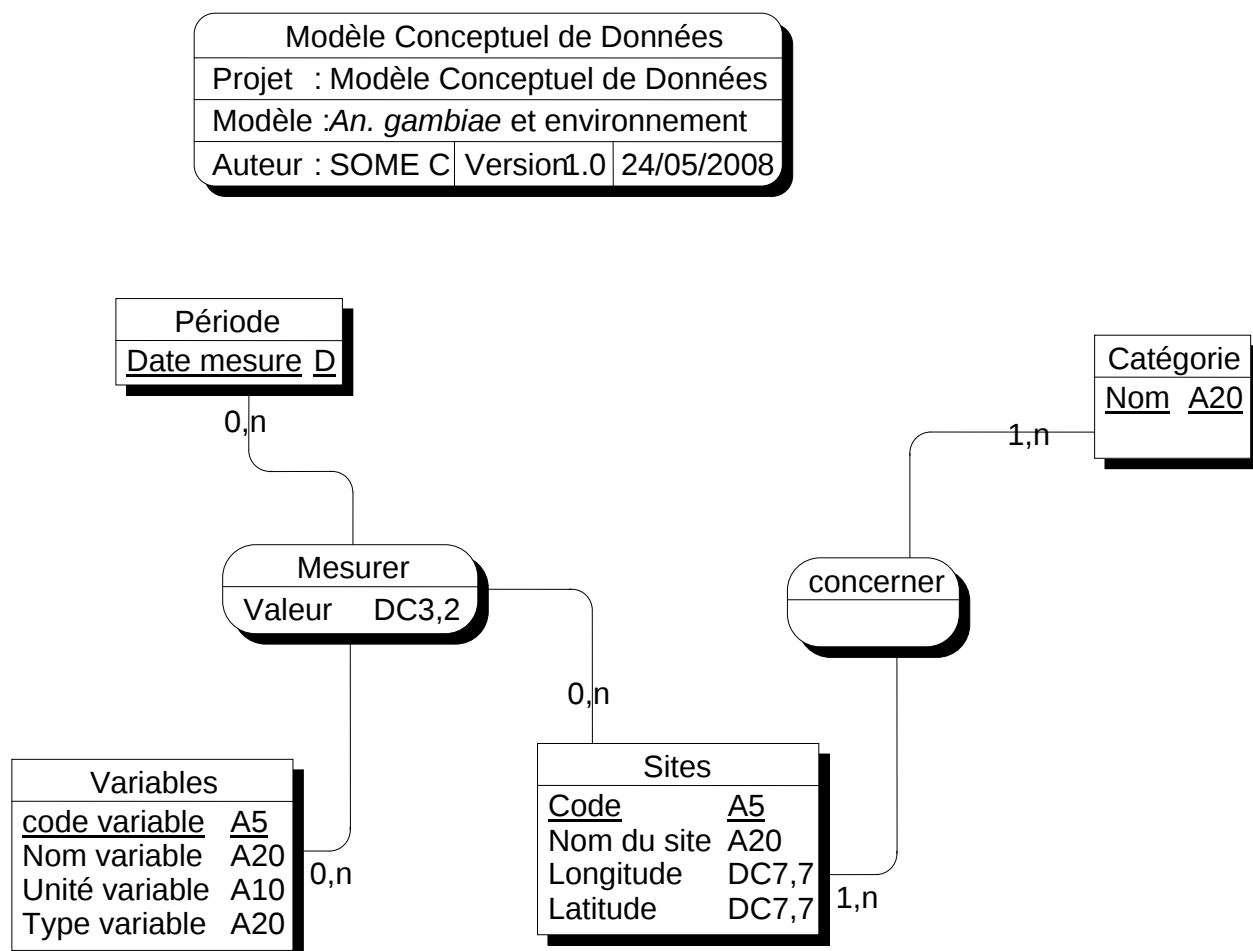


Figure 17 : Modèle conceptuel de données

Le modèle conceptuel contient 14 groupes d'information (Tableau 5) repartis entre les entités et les associations.

Tableau 5 : Liste des informations

Nom	Code	Type	Contenu
Catégorie	CATEGORIE	A15	Classification des sites selon les variables mesurées
Code	CODE	A5	Code de la catégorie
code variable	CODE_VARIABLE	A5	Code de la variable
Date mesure	DATE_MESURE	D	Date de collecte des données
Latitude	LATITUDE	DC7,7	Latitude du lieu de collecte
Longitude	LONGITUDE	DC7,7	Longitude du lieu de collecte
Nom	NOM_CATEGORIE	A20	
Nom du site	NOM_SITE	A20	Nom du lieu de collecte
Nom variable	NOM	A20	Nom de la variable collectée
Type variable	TYPE_VARIABLE	A20	Répartition des variables entre variables explicatives variables à expliquer. Type de variable (entomologique, environnementale)
Unité variable	UNITE	A10	Unité de la variable
Valeur	VALEUR	DC3,2	Valeur prise par la variable

L'entité est un « objet défini au sein du système d'information et au sujet duquel vous voulez conserver des informations » Sysbase, (1994). Notre système en compte quatre (Tableau 6).

Tableau 6 : Liste des entités

Nom	Code	Signification
Catégorie	CATEGORIE	La catégorie permet de distinguer/classer les sites de collecte de données par variables
Période	PERIODE	Période pendant laquelle les données ont été collectées
Sites	SITES	Lieu de collecte des données
Variables	VARIABLES	Les variables

Ces entités sont mises en relation à travers deux associations (Tableau 7)

Tableau 7 : Liste des associations

Nom	Code	Nombre
Concerner	CATEGORIE_SITE	0
Mesurer	VARIABLES_SITE	0

Sur la base de ce modèle, les variables sont mesurées sur des sites et à plusieurs dates durant une période. Ces variables sont regroupées par catégorie.

Chaque entité a des propriétés qui sont des éléments d'information liés à l'entité. C'est l'ensemble des éléments qui permettent de décrire les entités.

IV.1.1 L'ENTITE « CATEGORIE »

La catégorie correspond à un jeu de sites décrivant une même variable. Par exemple, les sites de mesure de la température vont correspondre à la catégorie température. Elle est décrite par la propriété nom de catégorie. La catégorie est associée à l'entité site. Dans cette association, un site relève d'une catégorie et une catégorie peut concerner un ou plusieurs sites. C'est une relation de 1 à plusieurs qui lie l'entité catégorie et l'entité site.

IV.1.2 L'ENTITE « SITES »

Le site est le lieu d'évaluation d'une variable. Il correspond par exemple au poste pluviométrique, à la localité de collecte des données entomologiques, ou à la densité de végétation sur une surface d'un km². Les propriétés du site, c'est-à-dire les éléments qui le caractérisent dans le système, sont : le nom du site, son code et ses coordonnées géographiques en longitude et en latitude. Le nom du site correspond au nom du lieu de collecte et le code est un identifiant qui lui est affecté. Ces coordonnées permettent d'identifier sa position dans le référentiel géographique.

Le site est en relation avec deux entités : l'entité catégorie qui permet de classer le site selon la

variable qui y est mesurée et l'entité variable qui est l'ensemble des éléments mesurés sur le site. Ainsi, sur un site, peuvent être mesurées plusieurs variables et une variable peut être mesurée sur plusieurs sites. L'association entre l'entité site et l'entité variable est de type plusieurs à plusieurs. Ces mesures peuvent être effectuées à plusieurs dates durant une certaine période, d'où l'existence d'une entité période.

IV.1.3 L'ENTITE « PERIODE »

Cette entité contient les informations relatives aux dates d'évaluation ou de mesure des variables sur les sites. Par exemple, pour les températures, les précipitations, l'insolation ou l'indice de végétation, il est nécessaire de savoir quand et où les observations / mesures ont été effectuées. L'entité période a deux propriétés : la date et le code. Elle est en relation avec deux entités que sont : l'entité site et l'entité variable.

La relation entre l'entité période et l'entité variable est de type plusieurs à plusieurs. Une variable peut donc être mesurée en zéro ou plusieurs dates et à une date, zéro ou plusieurs variables peuvent être mesurées.

La relation entre l'entité période et l'entité site est également de type plusieurs à plusieurs. En effet, plusieurs dates de mesure peuvent être effectuées sur un site et plusieurs sites peuvent être des lieux de mesure à la même date.

IV.1.4 L'ENTITE « VARIABLES »

La variable est un facteur qui rentre dans l'évaluation de la qualité de l'environnement par rapport aux exigences écologiques d'*An. gambiae* s.s. ou dans le dénombrement des fréquences d'*An. gambiae* s.s.. C'est par exemple, la température, l'évapotranspiration potentielle (ETP), la densité de population, le vent, la pression, l'hydrographie, le relief, la présence humaine, la population de S.

L'entité variable est décrite par quatre propriétés (Tableau 8) : le nom de la variable, le code de la variable, l'unité d'évaluation de la variable et le type de variable.

Tableau 8 : Propriété de l'entité variable

Nom	Code	Contenu
Code variable	CODE_VARIABLE	Code de la variable
Nom variable	NOM	Nom de la variable
Unité variable	UNITE	Unité de mesure de la variable
Type variable	TYPE_VARIABLE	Rôle de la variable

L'entité variable est en relation avec les entités période et site. La relation entre l'entité variable et l'entité période est de type plusieurs à plusieurs. Une variable peut être mesurée à plusieurs dates et à une même date plusieurs variables peuvent être mesurées. Il en va de même de la relation entre l'entité variable et l'entité site. Plusieurs variables peuvent alors être mesurées sur un même site et sur plusieurs sites peut être mesurée une même variable.

Ainsi modélisées, les données servent à des traitements et le modèle conceptuel de traitement permet de les anticiper.

IV.2 LE MODELE CONCEPTUEL DE TRAITEMENT

Le modèle conceptuel de traitement (MCT) « a pour objectif de représenter formellement les activités exercées par le domaine de gestion et les échanges de ce domaine avec son environnement » Sysbase, (1994). Il permet de modéliser les activités exercées par le domaine et d'identifier les échanges du domaine avec son environnement et les activités induites pour structurer le fonctionnement du domaine (Figure 18). En d'autres termes, le modèle consiste en l'articulation de l'ensemble des processus envisagés dans le cadre du travail.

Dans le cas de cette étude, les traitements peuvent être subdivisés en trois processus :

- le processus « collecte et traitement des données entomologiques » ;
- le processus « collecte et traitement de données environnementales » ;
- le processus « modélisation et élaboration de carte de prédiction de distribution ».

IV.2.1 LE PROCESSUS DE COLLECTE ET DE TRAITEMENT DES DONNEES ENTOMOLOGIQUES

Ce premier processus est composé de trois opérations :

1. L'opération sélection du site concerne l'ensemble des dispositions pratiques devant permettre la collecte des données primaires sur le terrain, particulièrement les données entomologiques. Elle consiste en quatre actions : la définition du transect, l'échantillonnage des sites de collecte, la définition des circuits de collecte, l'identification des lieux au moyen de coordonnées géographiques. La mise en œuvre de cette opération est conditionnée par la disponibilité des ressources humaines et financières ainsi qu'à l'accord des décideurs. A l'issue de l'opération, les lieux de collectes et de la logistique (les fiches de collecte, nécessaires à l'opération de collecte des données entomologiques) sont disponibles.

2. L'opération de collecte de données entomologiques est l'exécution pratique de la collecte de données sur le terrain. Elle consiste à recueillir sur les sites échantillonnés les informations sur les données relatives aux deux formes moléculaires M et S d'*An. gambiae* s.s.. Les annexes 1 et 2 présentent, respectivement, les méthodes de collecte des moustiques et les méthodes d'identification des formes moléculaires M et S d'*An. gambiae*. Elle est mise en œuvre seulement lorsque l'opération de sélection de sites est réussie et fournit la liste des localités et la fiche de collecte et que la période est propice à la collecte de données (période de prolifération de moustiques). Cette opération est constituée de quatre actions : la capture des moustiques, l'identification des formes moléculaires M et S d'*An. gambiae* s.s., le dénombrement des fréquences utiles (femelles gravides et semi-gravides, ou gorgées), un premier renseignement des fiches (lieu de collecte, fréquences de membres du complexe). Cette opération produit des fiches renseignées et des carcasses d'*An. gambiae* s.s. à analyser au laboratoire. Si cette opération est réussie, elle permet d'enclencher l'opération de dépouillement.

3. L'opération dépouillement/analyse de laboratoire. L'identification des formes moléculaires n'est pas totale sur le terrain et il faut des analyses au laboratoire pour l'achever. C'est la finalité de l'opération dépouillement/analyse de laboratoire. L'identification des membres du complexe achève

par le tri et le regroupement les anophèles capturés par unité spatiale de collecte (site) le renseignement d'une fiche synthétique. Cette fiche permet de connaître, par site d'observation, le nombre d'*An. gambiae* forme moléculaire S, le nombre d'*An. gambiae* forme moléculaire M, le nombre d'*An. arabiensis*, les coordonnées en longitude et en latitude des sites d'observation. Ces actions de l'opération « dépouillement/analyse de laboratoire » clôturent le processus collecte de données entomologiques.

IV.2.2 LE PROCESSUS DE COLLECTE ET DE TRAITEMENT DES DONNEES ENVIRONNEMENTALES

Il regroupe l'ensemble des opérations dédiées à la collecte de données secondaires sur la bio-écologie des anophèles, sur les caractéristiques du milieu physique et humain du Burkina Faso, le site de l'étude. Un accent particulier a été mis sur les données secondaires spatiales permettant la caractérisation du Burkina Faso en rapport avec la bio-écologie des anophèles dans ce modèle. Ce processus est composé de trois opérations :

1. L'opération collecte des données secondaires : elle consiste en la revue de littérature : manuels, précis, dictionnaires, planches, cartes, publications scientifiques, bases de données géographiques ou non, etc. C'est une opération de fouilles des travaux déjà réalisés et dont les informations peuvent constituer des apports intéressants en termes de données dans le test des hypothèses de départ. Cette opération se résume à la seule action d'inventaires et de collectes des informations déjà existantes. Il en résulte toujours un jeu de données.

2. L'opération caractérisation des données secondaires ; c'est un travail d'analyse et d'évaluation du jeu de données issues de l'opération précédente. Elle se décline en six actions : la description et le renseignement des données, la vérification de l'adéquation entre les différents systèmes de projection, la vérification de l'adéquation des échelles spatiales et temporelles par rapport aux objectifs de l'étude, la structuration des données et la création d'une base de données. Cette opération produit toujours une base de données et un rapport sur les données collectées.

3. Les travaux complémentaires consistent en une opération de mise en phase des données par rapport à l'ensemble de l'étude, du point de vue de leur adéquation aux objectifs et à la méthodologie. Cette opération est composée de cinq actions : l'harmonisation des systèmes de projection, la définition d'un référentiel spatial, l'harmonisation des échelles spatiales et temporelles et le complètement des données manquantes par estimation ou par interpolation ou encore par extrapolation. Cette opération permet également de produire, en sortie, une table complétant la table

issue du processus de collecte de données entomologiques en fournissant, pour chaque site de collecte, des informations relatives au milieu et intervenant dans la bio-écologie des anophèles.

Ces deux premiers processus peuvent se dérouler en même temps. Par exemple, en même temps que se déroule l'opération « sélection des sites », la collecte de données peut avoir lieu. Le troisième processus quant à lui n'est possible que lorsque les résultats des deux premiers processus sont disponibles.

IV.2.3 LE PROCESSUS DE MODELISATION ET DE REALISATION DE CARTE DE DISTRIBUTION PREDICTIVE

Ce processus vise l'élaboration de carte de distribution des deux formes moléculaires M et S d'*An. gambiae* au Burkina Faso, en s'appuyant sur les observations de terrain, les caractéristiques du milieu et la bio-écologie des anophèles. Pour cela, ce processus doit permettre, par identification des régularités, de formaliser la relation entre la distribution des différentes formes moléculaires et l'environnement. Cette formalisation sert de support pour la production de carte de distribution. Ce processus compte de cinq opérations :

1. L'opération « formatage du tableau de synthèse ». Il s'agit de faire le formatage des tableaux de données à analyser en regroupant en un seul les synthèses des processus « collecte de données entomologiques » et « collecte de données secondaires ». Il s'agit également de vérifier l'exhaustivité de l'information sur les différents sites de collecte.

2. L'opération « analyse spatiale descriptive » consiste en l'analyse de la structure spatiale à travers l'utilisation d'indicateurs statistiques (centroïde, barycentre, ellipse de déviation standard) et d'indices (Moran, Getis-Ord General G, Anselin Local Moran et Getis-Ord Gi*).

3. L'opération « ACP », analyse en Composantes Principales, avec variables qualitatives complémentaires, est un traitement statistique que l'on réalise pour avoir la structure de données statistiques à travers les corrélations entre les variables et l'identification des axes factoriels. Cette opération permet également une première appréciation de la contribution des différentes variables à la distribution spatiale des deux formes moléculaires d'*An. gambiae* s.s.

4. L'opération « ANCOVA » (analyse de covariance) est utilisée dans le cas où l'ACP révèle que les variables qualitatives peuvent améliorer la qualité de la modélisation. Si l'ANCOVA est concluante, il en résulte plusieurs fonctions de la distribution selon les modalités des variables qualitatives.

L'évaluation de la contribution des variables qualitatives est faite avec des tests statistiques.

5. L'opération « régression » suppose que l'ANCOVA montre une faible contribution des variables qualitatives. Dans ce cas, toutes ces variables qualitatives sont éliminées. A cela suit le calcul de différents coefficients et certains tests statistiques : tests de Fisher, coefficient de détermination, analyse de résidus, composition d'une fonction de régression etc. Par ces actions, l'analyse de résidus permet une évaluation de la qualité du modèle. La fonction de régression est inversée au niveau spatial pour produire une carte de distribution. Une comparaison est ensuite réalisée entre des données issues du modèle et des observations empiriques. Elle permet de juger de la qualité du modèle et confirmer l'évaluation par l'analyse des résidus.

IV.3 TRAITEMENT DES VARIABLES ECO-GEOGRAPHIQUES

Dans le processus « traitement », le traitement des données éco-géographiques est pris en charge à travers les activités d'interpolation, d'estimation et de classification. Le traitement de variables éco-géographiques présente quelques spécificités relatives surtout à l'exhaustivité spatiale de l'information qui est indispensable à la modélisation. Elles subissent des traitements spécifiques dont les résultats sont réintégrés à la base de données comme entité. C'est pourquoi, cette section leur a été spécialement dédiée.

L'étude des relations entre la distribution d'*An. gambiae* s.s. et les facteurs de l'environnement requiert que l'on dispose des informations sur chaque partie de la relation comme le montre le processus de modélisation du MCT. Les données éco-géographiques constituent les facteurs externes de la répartition spatiale des espèces. En fonction de leur nature et de leur impact sur la bio-écologie des membres du complexe *An. gambiae*, elles ont été regroupées en trois types : les facteurs géographiques, les facteurs biotiques et les facteurs climatiques. Les données collectées ne sont pas homogènes par groupe de facteurs et ne présentent pas non plus le même niveau de conformité par rapport au besoin de l'étude (terrain nominal). Aussi, les données éco-géographiques ont-elles été réparties en trois catégories en fonction des traitements effectués.

IV.3.1 INTERPOLATION DES DONNEES

Elle concerne : les précipitations, les températures, les vents, l'insolation, l'ETP, la pression, les altitudes.

IV.3.1.1 Les précipitations : des données sources aux données interpolées

Les données de précipitation proviennent de deux sources : la base de données LoClim de la FAO et la base de données de la Direction Générale de la Météorologie du Burkina. Les données utilisées couvrent la période de 1975-2005 et concernent 157 stations dont 140 à l'intérieur du Burkina Faso et 17 dans les pays voisins (Niger, Mali, Togo, Bénin, Côte d'Ivoire, Ghana). Les valeurs archivées sur ces stations qui ont été utilisées sont les précipitations annuelles. Les stations sont implantées en point et sont repérées par des coordonnées angulaires sexagésimales (degré, minute, seconde) en longitude / latitude.

Les traitements effectués sur les données de précipitation sont :

- la spatialisation et le changement de système de projection avec transformation des coordonnées d'origine en coordonnées rectangulaires dans le système de projection de référence du Burkina défini en méthodologie (UTM, ellipsoïde de Clarke 1880, Zone 30 étendue à 31) ;
- le calcul de la pluviosité de chaque station ;
- l'interpolation par la méthode du Krigeage simple avec une fonction d'ajustement trétrasphérique après analyse de la dérivabilité de la variable régionalisée et une analyse de l'erreur (Figure 19 et Figure 20).

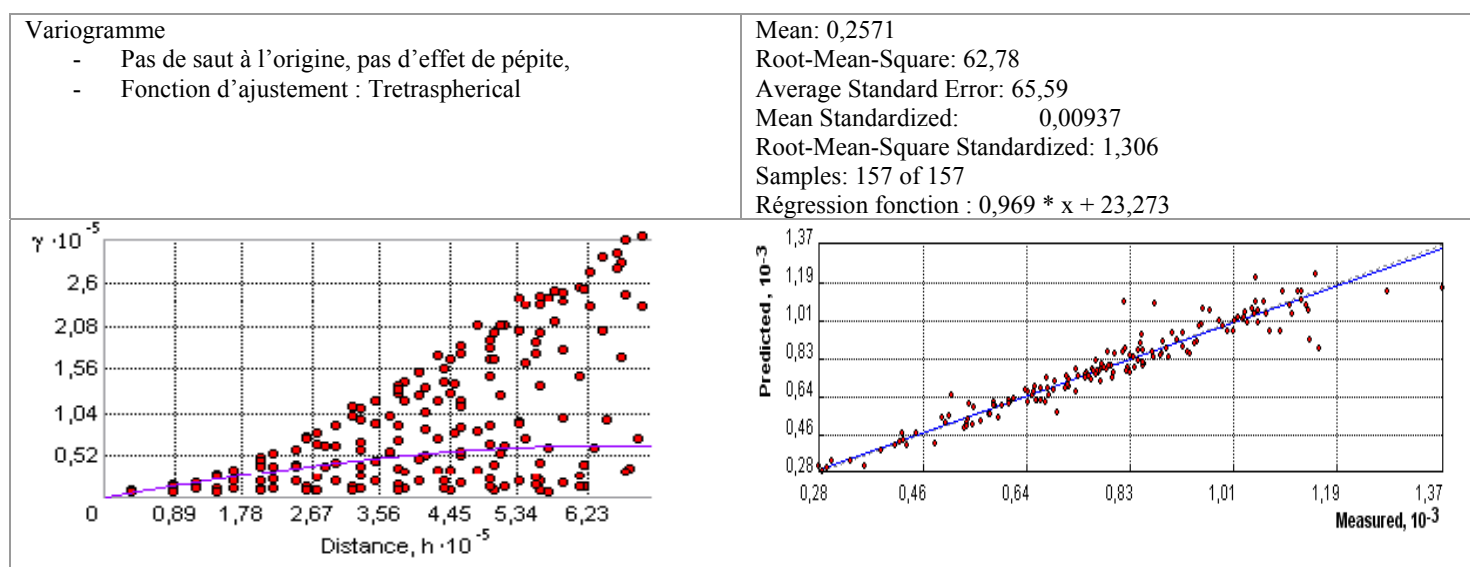


Figure 19 : Variogramme et fonction de régression pour l'interpolation des précipitations

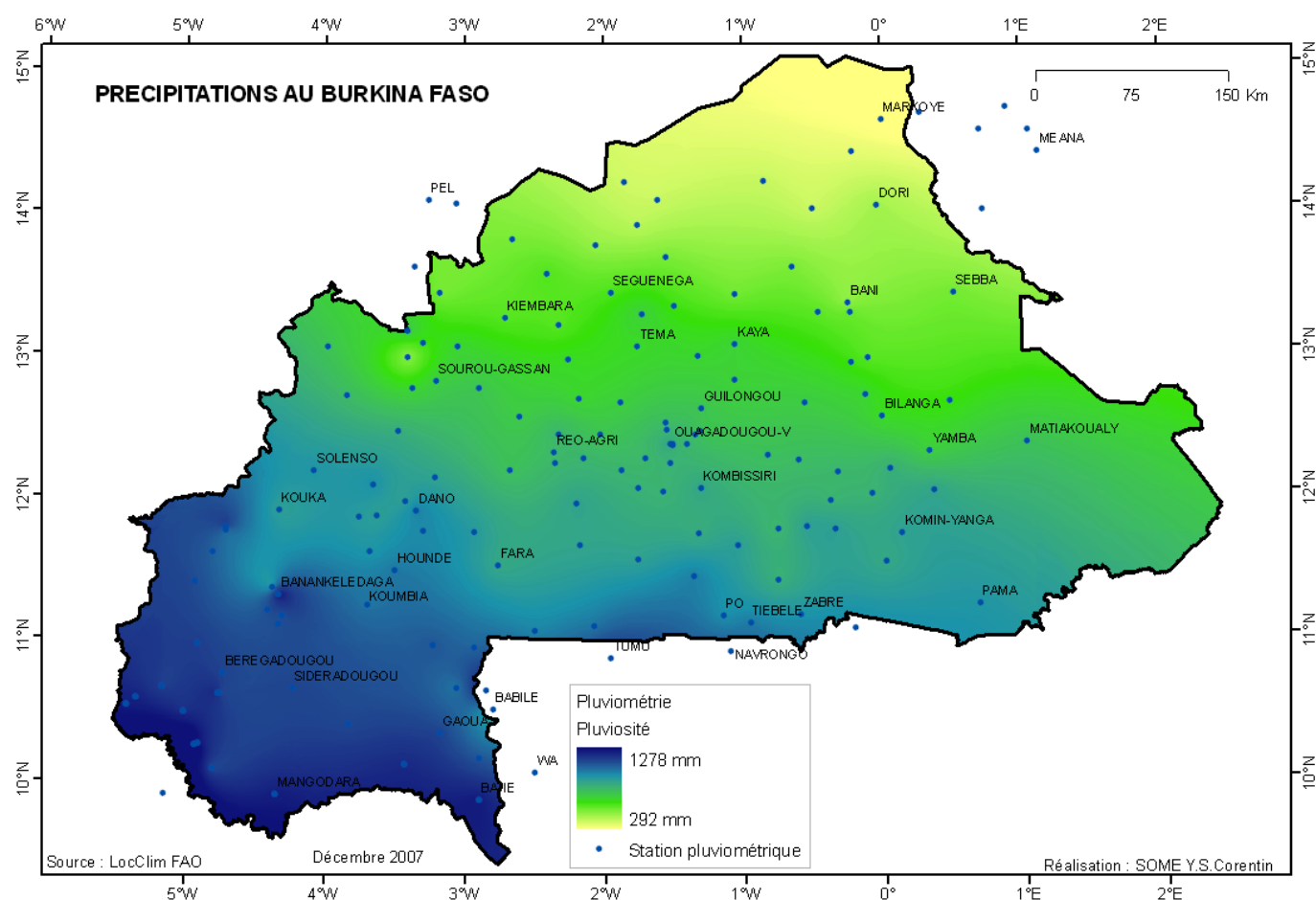


Figure 20 : Répartition des précipitations au Burkina Faso

IV.3.1.2 Les températures : des données sources aux données interpolées

Les données de température proviennent de la base de données LocClim de la FAO et la base de données de la Direction de la Météorologie Nationale. Elles couvrent la période de 1975 à 2005. Le réseau d'observation utilisé compte 68 points d'observation. Les données enregistrées sont les températures moyennes minimales mensuelles et les températures moyennes maximales mensuelles. Sur les 68 points d'observation, 31 sont situés sur le territoire burkinabé, soit moins de 50% des stations d'observations. Le reste se trouve dans les pays voisins limitrophes. Ces stations sont toutes repérées en coordonnées géographiques (latitude / longitude), en degrés sexagésimaux.

Les traitements effectués sur les données de températures sont :

- la spatialisation et le changement de système de projection avec transformation des coordonnées d'origine en coordonnées rectangulaires dans le système de projection de référence du Burkina défini en méthodologie (UTM, ellipsoïde de Clarke 1880, Zone 30 étendue à 31) ;
- le calcul de la température moyenne annuelle de chaque station ;
- l'interpolation par la méthode IDW après analyse de la dérivabilité de la variable régionalisée et une analyse de l'erreur (Figure 21 et Figure 22).

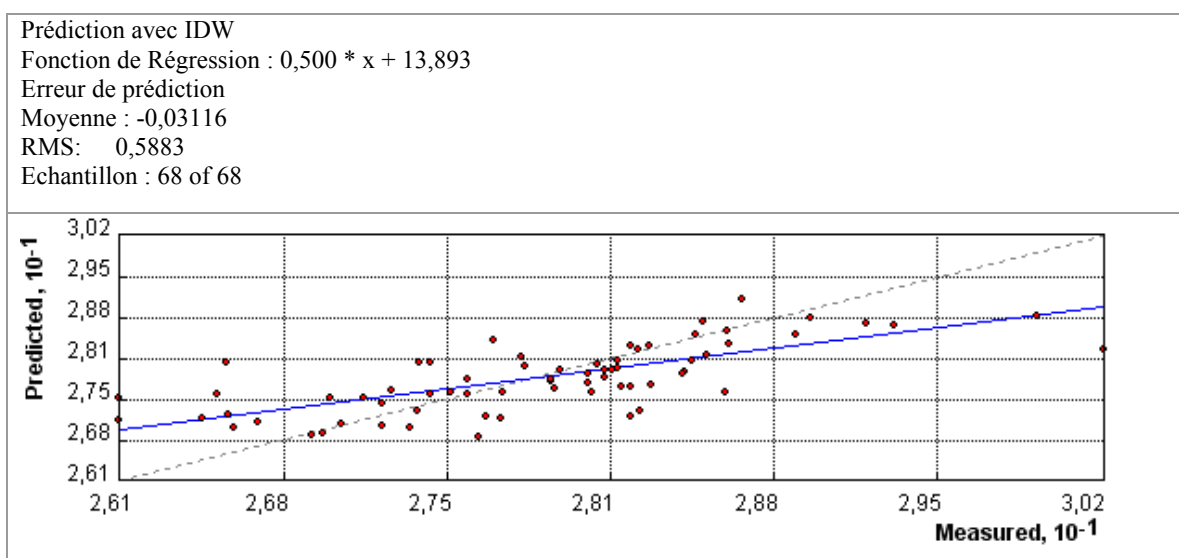


Figure 21 : Fonction de régression de l'interpolation des températures moyennes annuelles

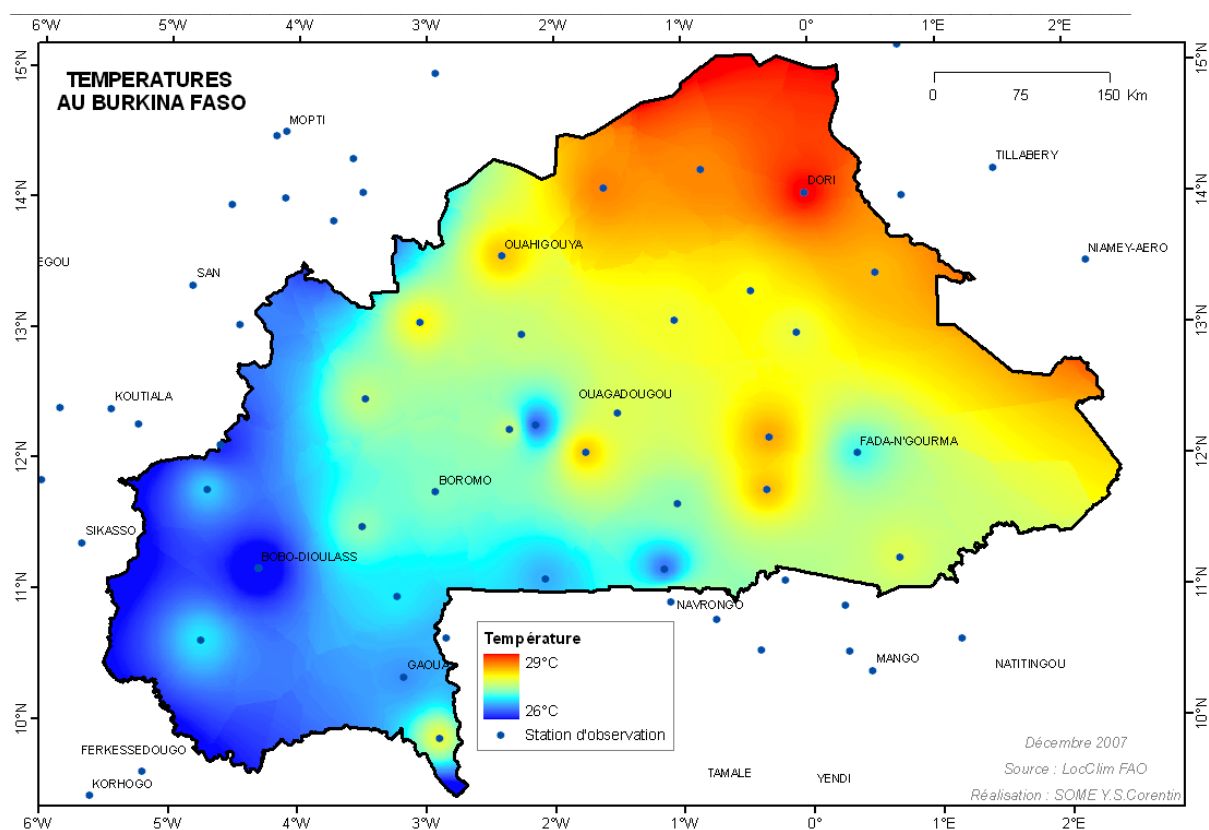


Figure 22 : Répartition des températures moyennes annuelles au Burkina Faso

IV.3.1.3 L'Évapotranspiration potentielle : des données sources aux données interpolées

Les données de l'évapotranspiration potentielle proviennent de la base de données de la Direction Générale de la Météorologie Nationale du Burkina et la base de données Loclim de la FAO. Elles couvrent la période allant de 1975 à 2005. Les observations ont été effectuées sur 38 stations dont 7 à l'intérieur du Burkina Faso. Les points d'observations sont positionnés en coordonnées géographiques en longitude/latitude et exprimés en unités sexagésimales.

Les traitements ont consisté :

- en la spatialisation et le changement de système de projection avec transformation des coordonnées d'origine en coordonnées rectangulaires dans le système de projection de référence du Burkina défini en méthodologie (UTM, ellipsoïde de Clarke 1880, Zone 30 étendue à 31) ;
- au calcul de l'ETP moyenne annuelle de chaque station ;
- En l'interpolation par la méthode du krigeage ordinaire après analyse de la dérivabilité de la variable régionalisée et une analyse de l'erreur (
- Figure 23 et Figure 24).

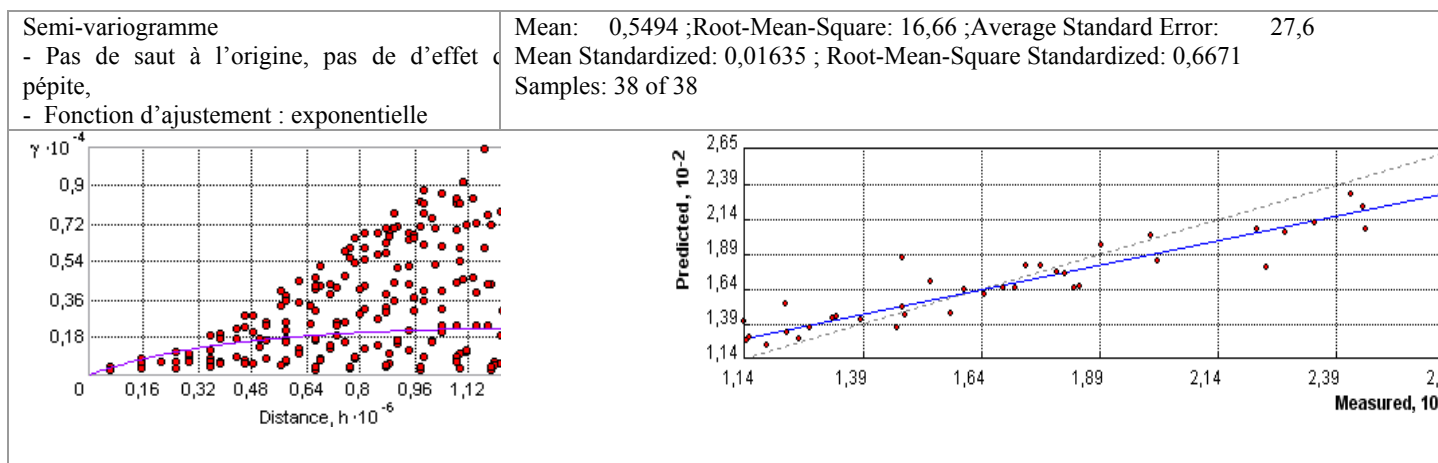


Figure 23 : Variogramme et droite de régression de l'interpolation de l'évapotranspiration potentielle moyenne annuelle

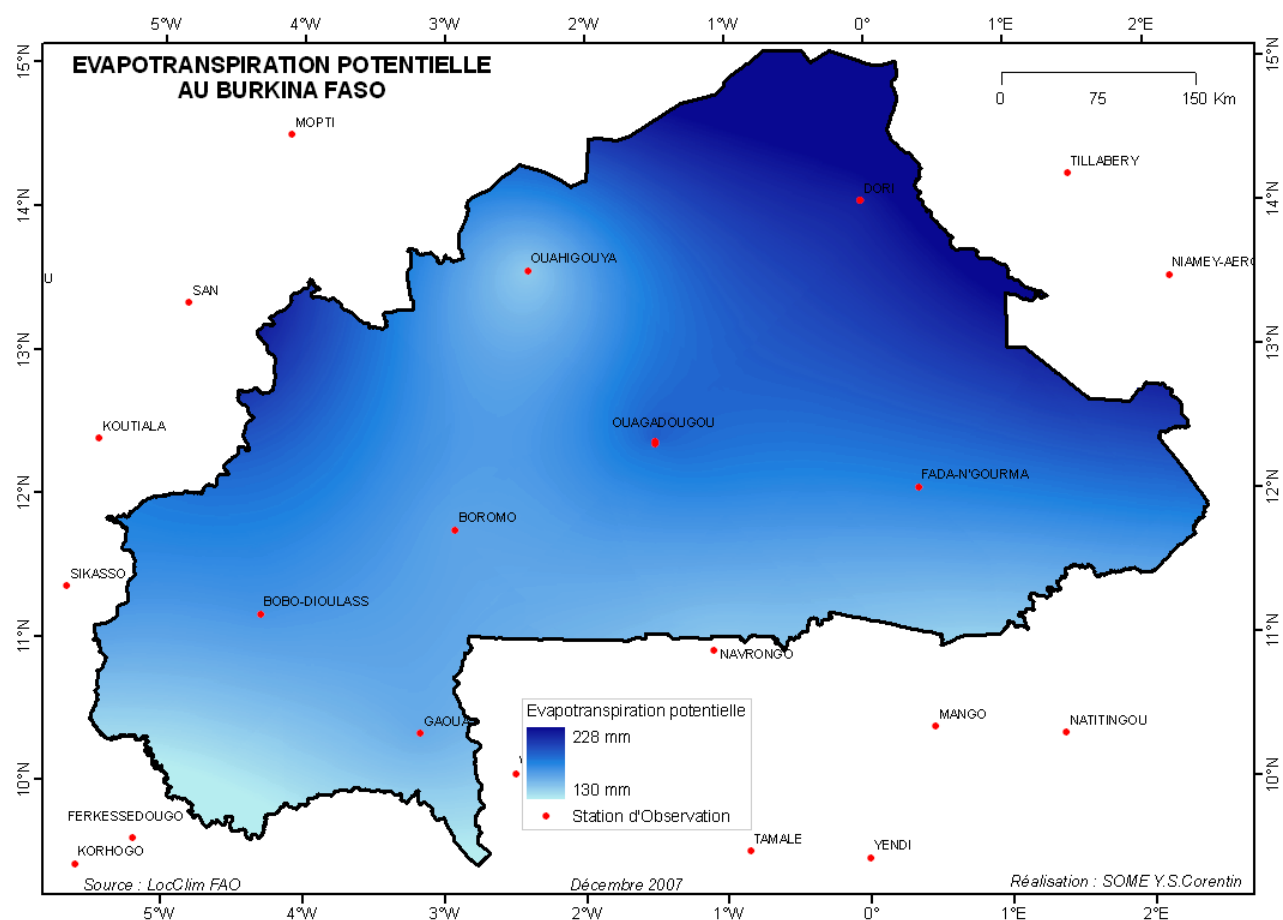


Figure 24 : Répartition de l'évapotranspiration potentielle annuelle au Burkina Faso

IV.3.1.4 La pression de vapeur : des données sources aux données interpolées

Les données de pression de vapeur proviennent de la base de données LocClim de la FAO. Elles couvrent également la période allant de 1975 à 2005. Pour notre zone, 40 points d'observations dont 7 à l'intérieur du site sont disponibles. Le géoréférencement de ces points est fait en coordonnées géographiques avec les unités en degrés sexagésimaux.

Les traitements effectués sur ces données sont :

- la transformation du système de coordonnées et leur inscription dans le référentiel technique retenu ;
- le calcul des valeurs moyennes annuelles des pressions sur l'ensemble des sites d'observation ;
- l'interpolation par la méthode « *Global polynomial Interpolation* ». Cette méthode a eu les meilleurs taux d'erreur et le meilleur ajustement (Figure 25 et Figure 26).

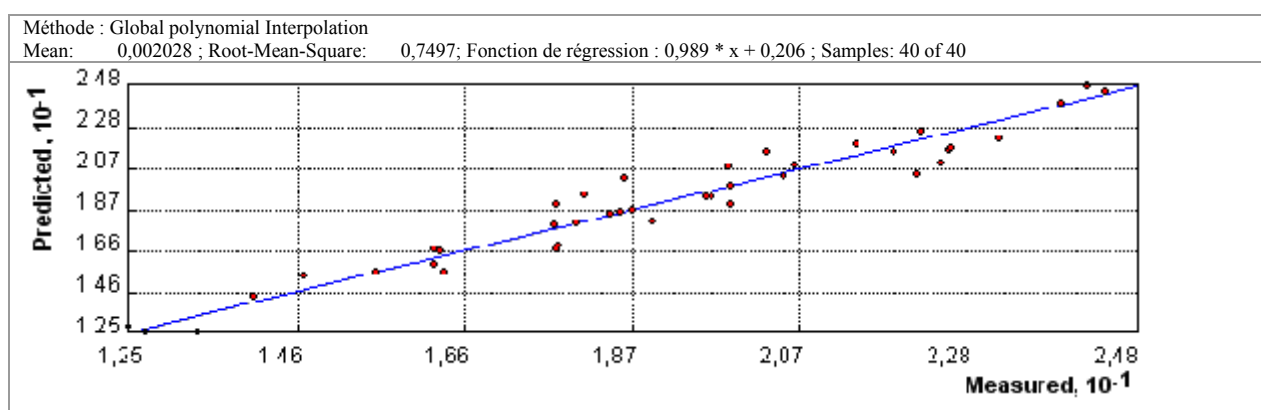


Figure 25 : Droite de régression des pressions de vapeur moyennes annuelles au Burkina Faso

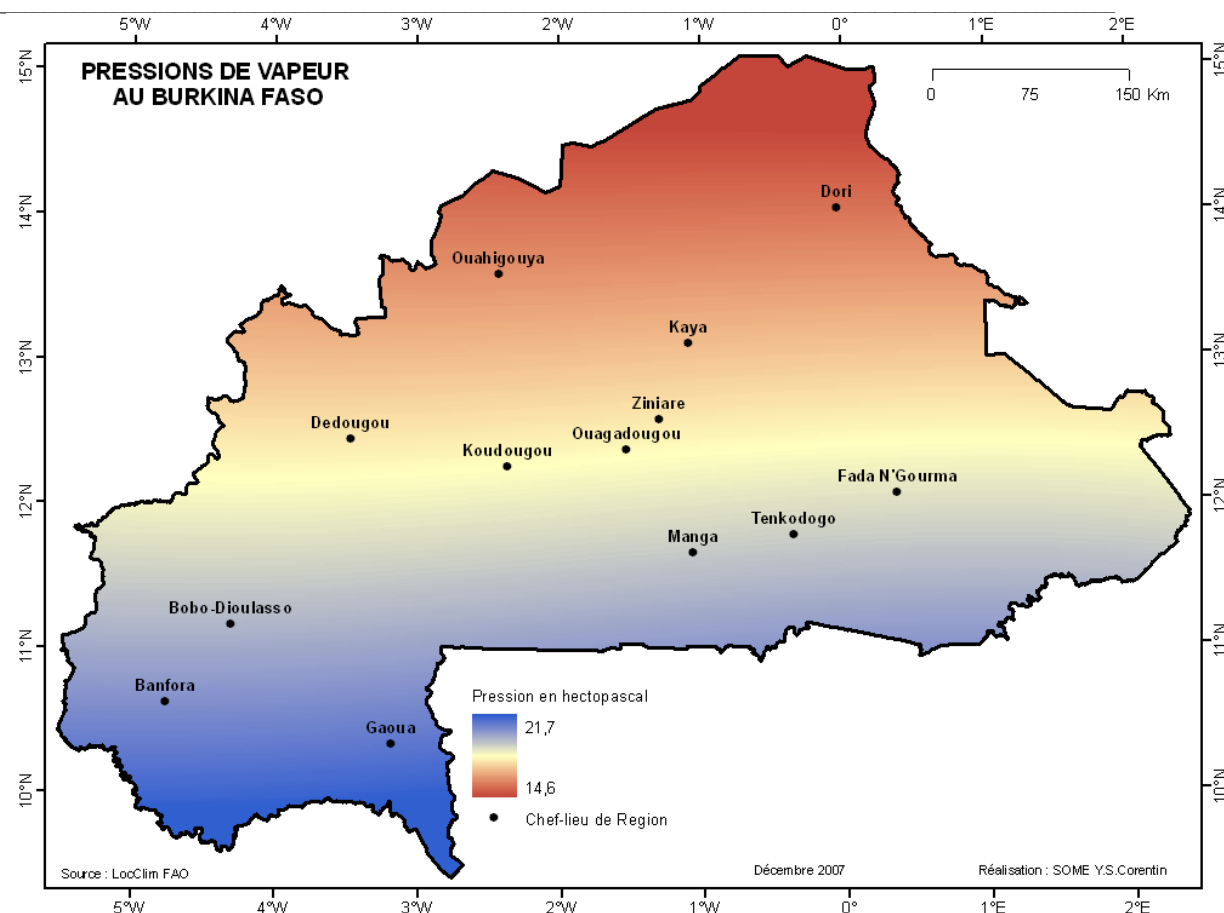


Figure 26 : Répartition des pressions de vapeurs moyennes annuelles au Burkina Faso

IV.3.1.5 L'insolation : des données sources aux données interpolées

Les données d'insolation proviennent de la base de données LoClim de la FAO. Les données d'observation couvrent 41 stations, dont 7 sont situées sur le territoire burkinabé. Ces stations sont repérées par des coordonnées géographiques en degrés sexagésimaux. Les relevés couvrent la période 1975-2005 et sont en moyennes mensuelles. Les traitements subis par ces données concernent à la fois l'adaptation au référentiel technique géographique et au niveau d'échelle temporelle. Après la spatialisation, une re-projection des données en UTM Clarke 1880, Zone 30 étendue à 31 est appliquée. Au niveau temporel, l'insolation moyenne annuelle par station a été calculée. Pour la question de couverture et d'exhaustivité, une interpolation avec la méthode *Global Polynomial Interpolation* a été adoptée (Figure 27, Figure 28).

Méthode : Global polynomial Interpolation ; Mean: 0,00113 ; Root-Mean-Square: 0,06316
 Fonction de régression : $0,647 * x + 0,229$; Samples: 41 of 41

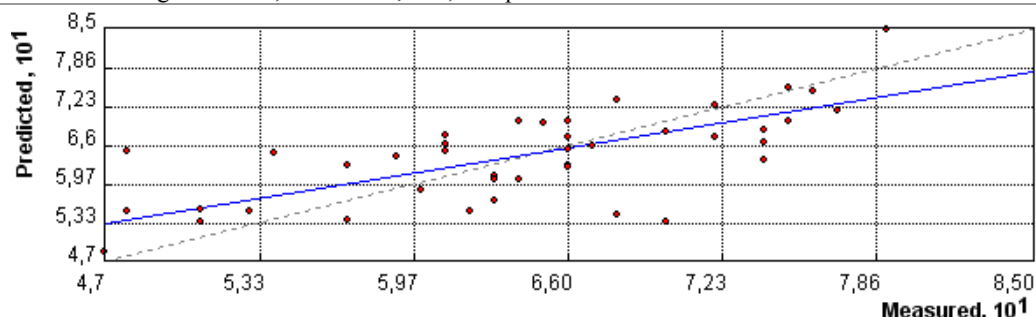


Figure 27 : Droite de régression de l'insolation

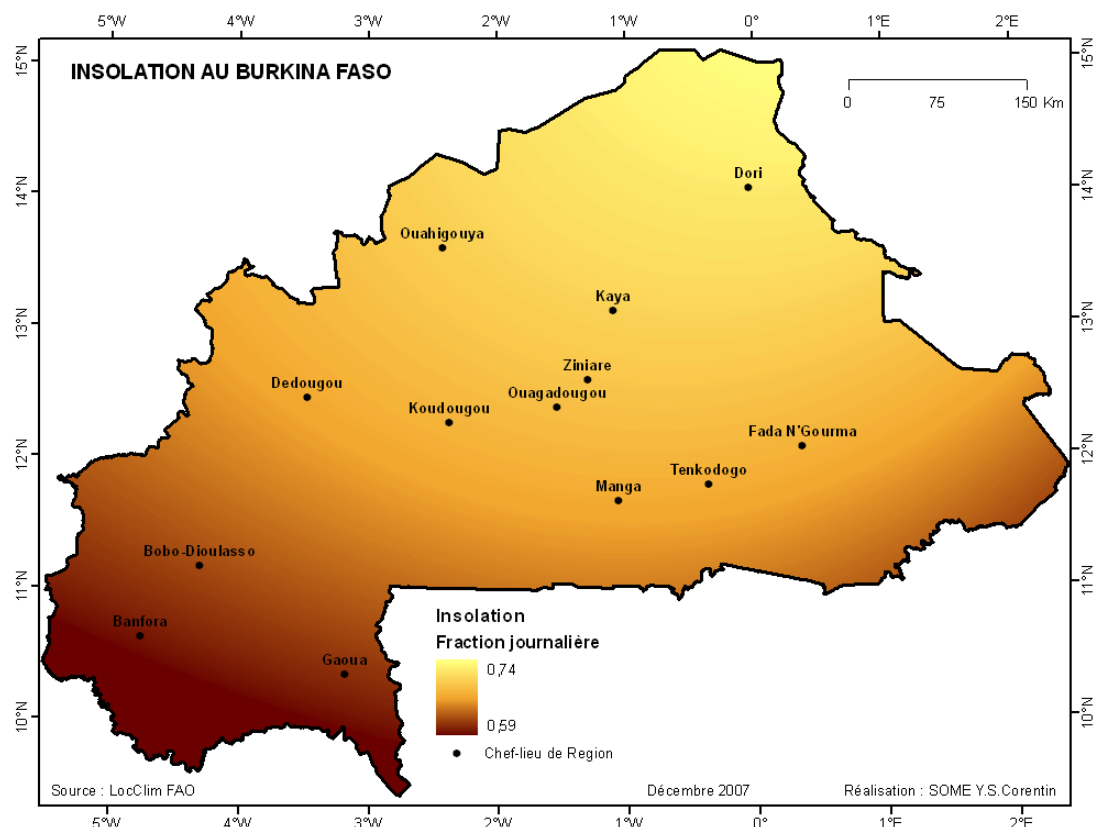


Figure 28 : Répartition de l'insolation moyenne annuelle au Burkina Faso

IV.3.1.6 La vitesse du vent : des données sources aux données interpolées

Les données relatives à la vitesse du vent ont été extraites de la base de données LocClim de la FAO. 39 stations d'observation ont été mises à contribution dont 7 au Burkina Faso. Le géoréférencement de ces stations est fait avec des coordonnées géographiques en unités sexagésimales. Les données enregistrées sont des vitesses moyennes mensuelles. Les traitements appliqués aux données extraites sont :

- la transformation du système de coordonnées des stations et leurs inscriptions dans le système de référence retenu pour l'étude ;
- l'adaptation au niveau d'échelle temporelle retenue (qui est l'année) par le calcul des vitesses moyennes annuelles par station ;
- l'application de la méthode d'interpolation *Radial Basis Fonctions* pour résoudre le problème de l'exhaustivité et d'étendue spatiale (Figure 29, Figure 30).

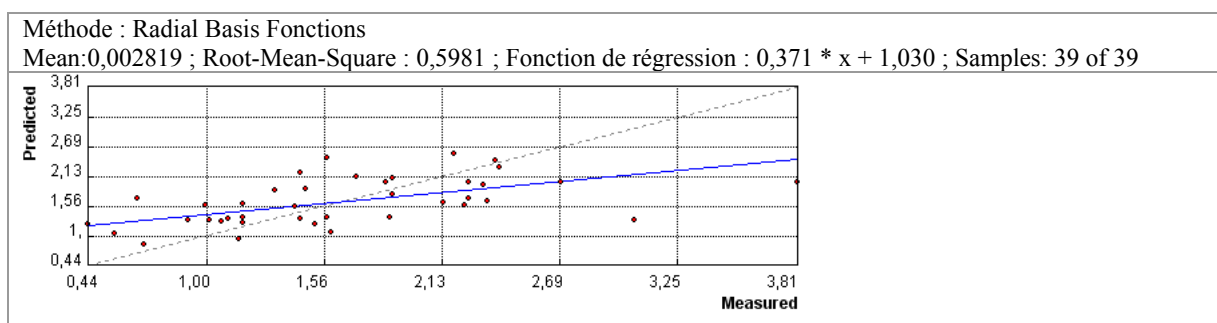


Figure 29 : Droite de régression de la vitesse des vents

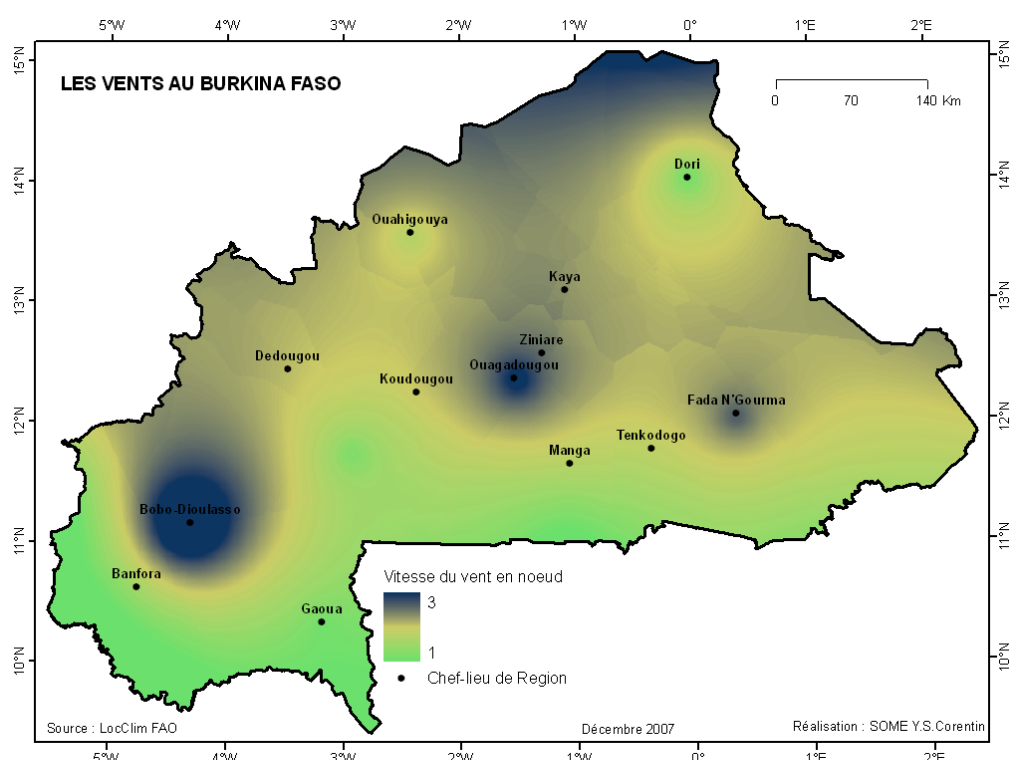


Figure 30 : Répartition des vitesses moyennes annuelles des vents au Burkina Faso

IV.3.1.7 L'altitude: des données sources aux données interpolées

Les données d'entrée des altitudes proviennent de la Banque Nationale de Données Topographiques (BNDT) du Burkina Faso. Elles sont constituées de l'ensemble des points cotés de la carte topographique au 1/200 000. 5 381 points cotés ont été utilisés et tous se situent sur le territoire

burkinabé (Figure 31 et Figure 32). En dehors de l'interpolation avec la méthode *Global Polynomial Interpolation*, aucun traitement n'a été effectué.

Méthode : Global polynomial Interpolation ; Mean: 1,936 ; Root-Mean-Square: 39,82 ;
Fonction de regression : $0,606 * x + 121,359$
Samples: 5381 of 5381

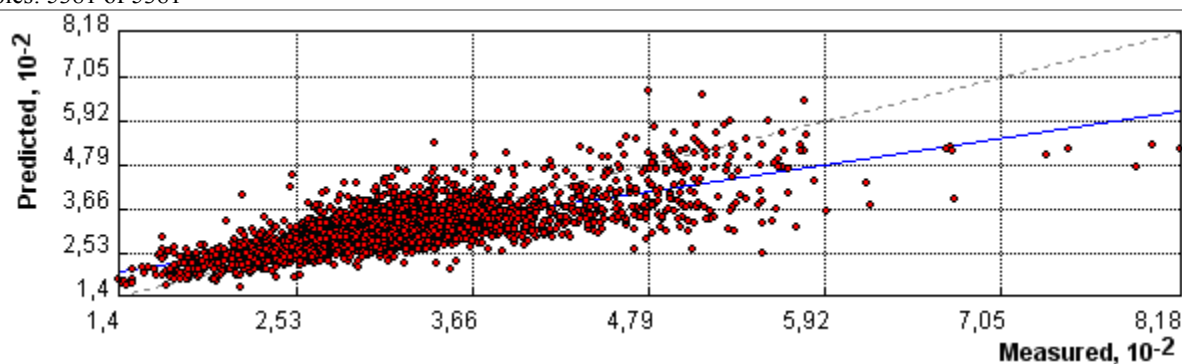


Figure 31 : Droite de régression des altitudes

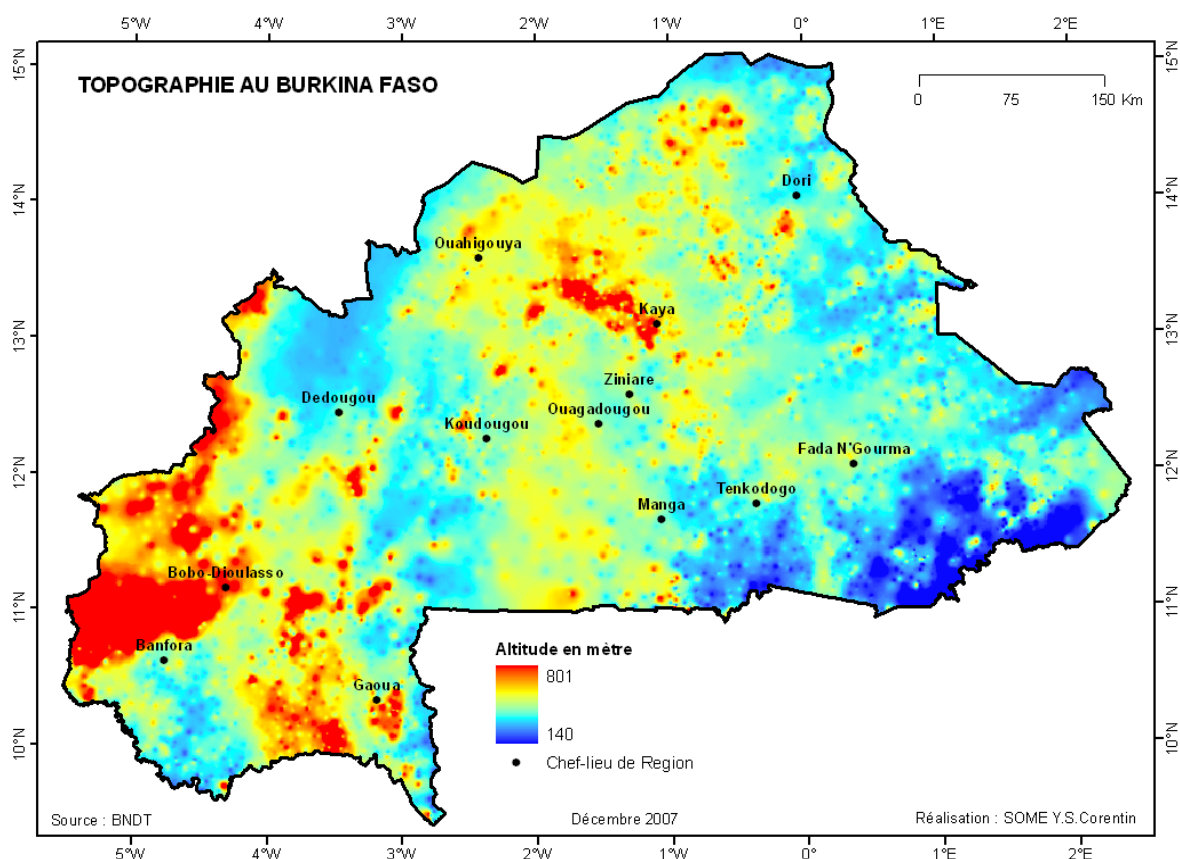


Figure 32 : Répartition des altitudes au Burkina Faso

IV.3.2 ESTIMATION DES DONNEES

L'interpolation s'adapte mal aux ensembles finis, comme par exemple les cas de dénombrement de population. Pour ces cas, des méthodes d'estimation particulières ont été adoptées. Elles sont mieux

adaptées à ce type de données et permettent d'avoir une meilleure précision. L'estimation concerne trois données éco-géographiques : les populations, l'hydrographie et les localités. Pour les populations, ce sont les densités qui sont recherchées. Pour l'hydrographie et les localités, les distances en tous points du site par rapport à ces objets géographiques sont requises.

IV.3.2.1 Population : des effectifs de population aux densités en tout point

Les données de population proviennent du recensement général de la population et de l'habitat de 1996. Ce sont des données sur les effectifs des populations humaines par villages. Elles ne sont pas géoréférencées et ne couvrent que les villages administratifs du Burkina en 1996. Pour le géoréférencement de ces données de recensement, une jointure avec les localités de la base de données topographiques a été effectuée. Au total, 7 733 villages ont été identifiés et joints aux localités de la base de données topographiques. Les statistiques attachées à ces points datent de 1996. Elles ont été réévaluées en appliquant le taux d'accroissement naturel du pays qui est de 3,3% pour la période 1985 à 1996. Pour répondre à la question d'exhaustivité et d'étendue, une estimation des *Densités lissées* par désagrégation des effectifs de population affectés aux différents points représentant les villages a été appliquée. Les statistiques sur les distances entre deux villages sont les suivantes :

- distance minimum⁵⁶ entre deux localités voisines : 427 m soit environ 0,5 km ;
- distance moyenne entre deux localités voisines : 4 148 m soit environ 4,1 km ;
- distance maximum entre deux localités voisines : 22 471 m soit environ 22,5 km.

Le diamètre moyen des localités est donc de 8 396 m soit 8,5 km. C'est cette distance qui est retenue comme estimateur de la zone de répartition des effectifs autour de chaque point. La Figure 33 donne le résultat de ces traitements.

⁵⁶ Les distances considérées sont les distances euclidiennes.

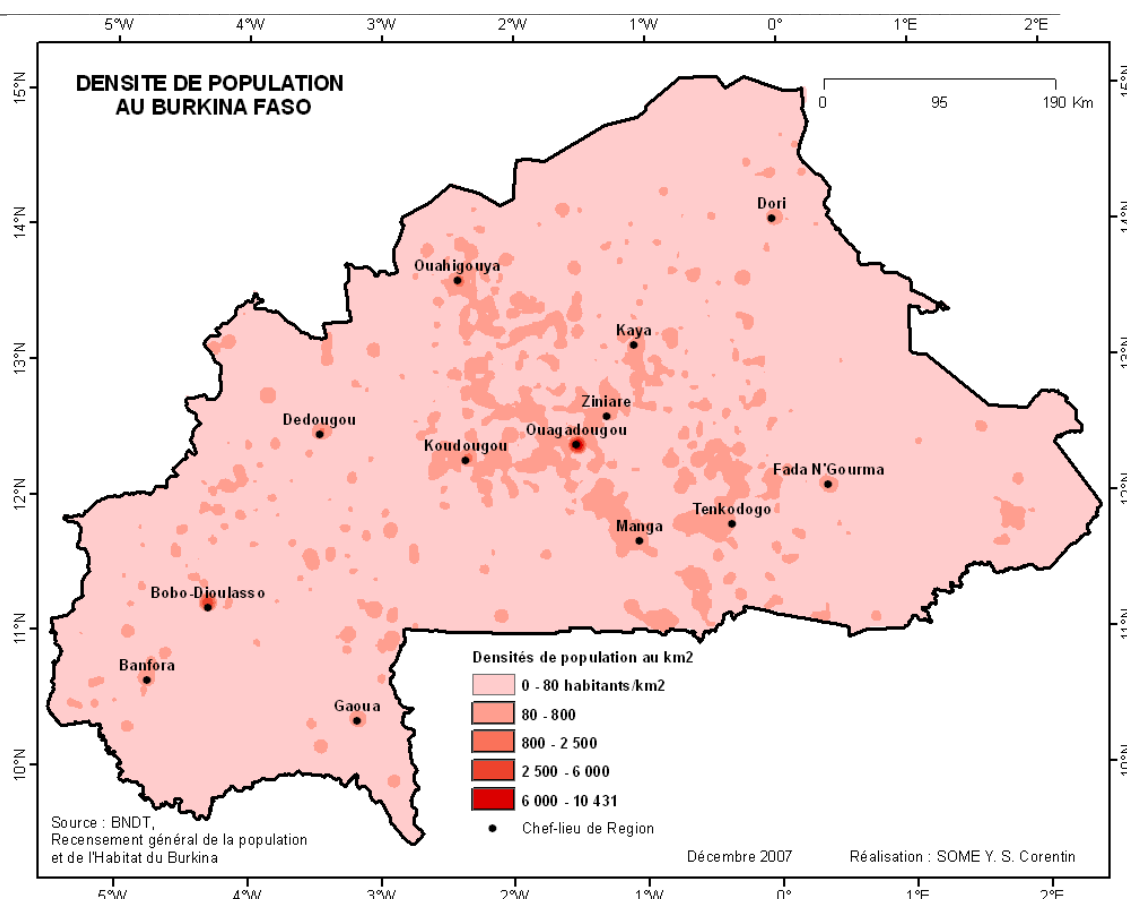


Figure 33 : Répartition des densités de population au Burkina Faso en 2005

IV.3.2.2 Des NDVI à la densité de la végétation

La quantification de la végétation s'est appuyée sur les indices décennaux normalisés de la végétation de SPOT Végétation. Les images NDVI de l'Afrique pour la période allant du 1^{er} juillet 2005 au 31 octobre 2005 correspondant à la période de capture ont été retenues. Au total, 12 images NDVI ont été collectées sur l'Afrique. Habituellement, les NDVI ont des valeurs comprises entre -1 et 1. Les images NDVI exploitées fournies par SPOT Végétation sont recodées sur 8 bits. Ce recodage est effectué en multipliant la valeur du NDVI par 127 et ajoutant 127 au résultat $((NDVI \times 127) + 127)$. Ce recodage conduit à des valeurs sur les pixels allant de 0 à 255.

Les traitements appliqués aux données NDVI concernent leur géoréférencement en coordonnées géographiques, la création d'une image arithmétique par calcul du NDVI moyen, le découpage de la zone d'étude et enfin la transformation du système de coordonnées dans le référentiel technique retenu (Figure 34).

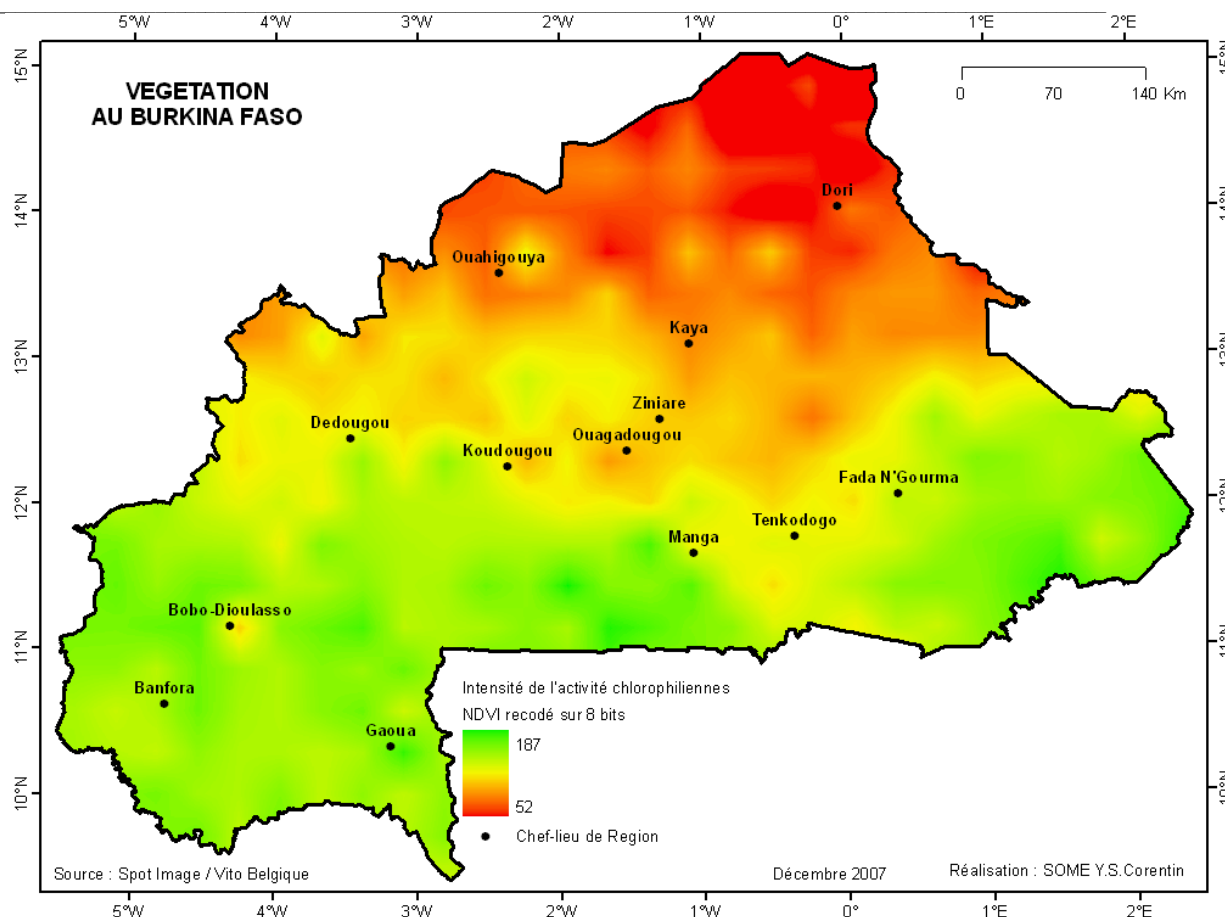


Figure 34 : Répartition des densités végétales au Burkina Faso

IV.3.2.3 Les localités et l'hydrographie : la distance en tout point par rapport à ces objets géographiques

Les données sur l'hydrographie et les localités proviennent de la Banque Nationale de Données Topographiques (BNDT). Aucun autre traitement n'a été effectué sur ces couches en dehors du calcul des distances en tous points du Burkina Faso. Les Figure 35 et Figure 36 ci-après présentent les résultats de ces traitements pour l'hydrographie et pour les localités.

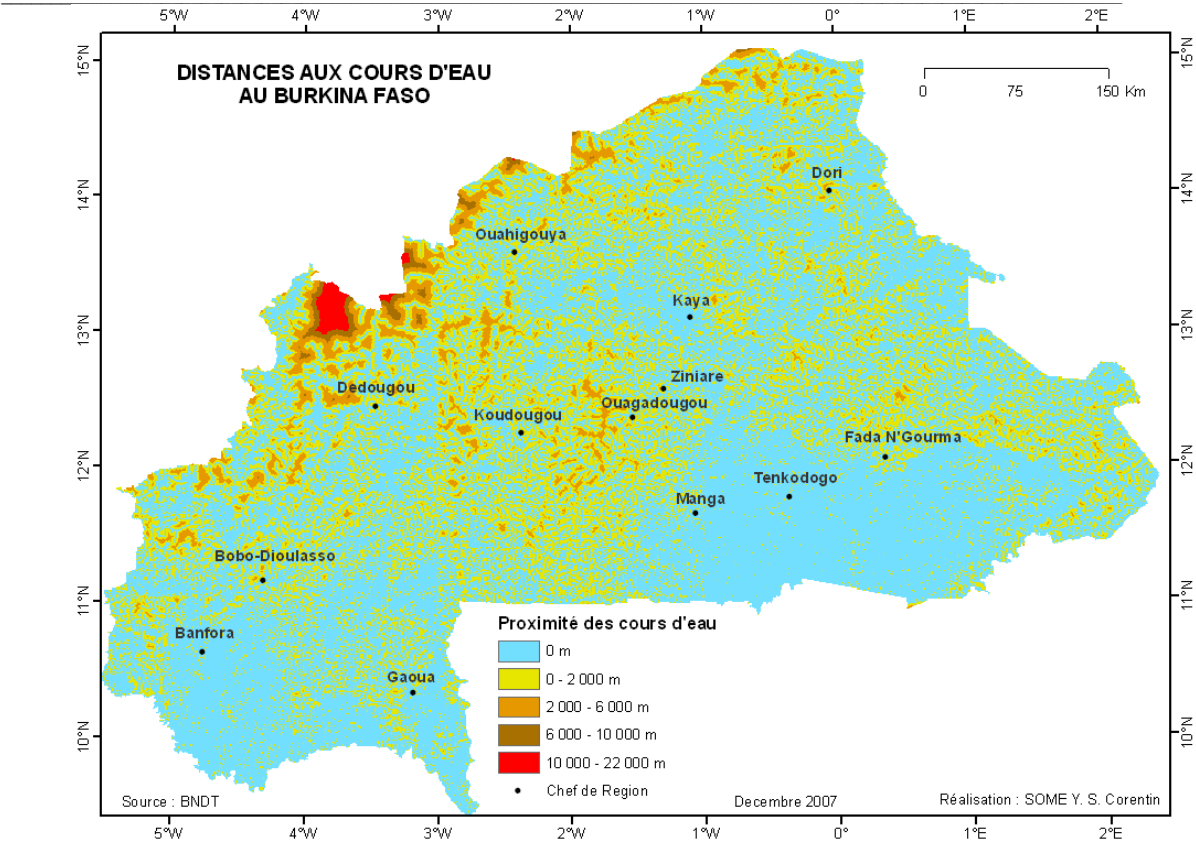


Figure 35 : Répartition des distances en tout point par rapport à l'hydrographie

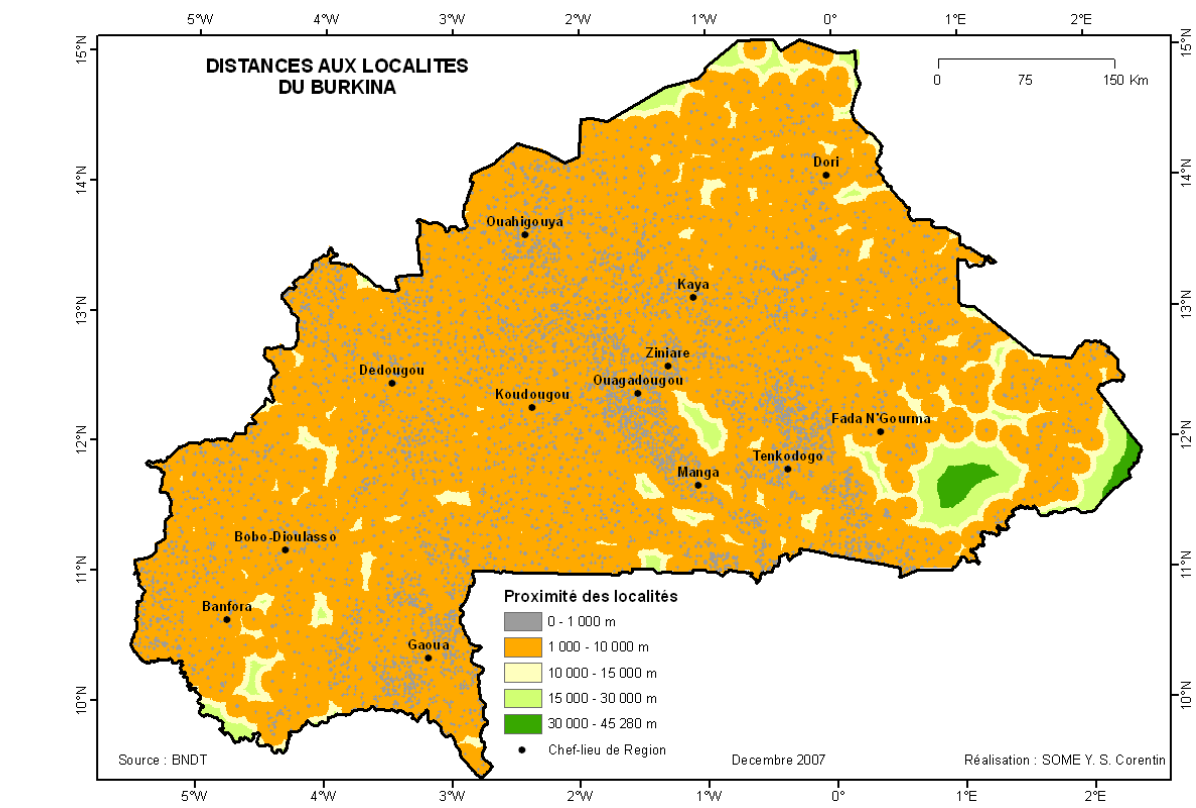


Figure 36 : Répartition des distances par rapport aux localités

IV.3.3 LES DONNEES CLASSIFIEES

Les données classifiées concernent les données qualitatives. Ce sont les données relatives aux sols, à la végétation et à l'occupation des terres. Elles proviennent respectivement de la carte de reconnaissance pédologique du Burkina, de la carte de végétation du Burkina et de la Base de Données sur l'Occupation des Terres (BDOT) du Burkina. La nomenclature utilisée dans le cadre de ces bases de données est très détaillée. Ces données sont de production très coûteuse et il n'y a pas une série. Elles n'ont par conséquent subies que deux types de traitement : le changement de projection et la classification. Ce dernier traitement a consisté en la réduction du nombre de modalités.

CONCLUSION AU CHAPITRE IV

La base de données utilisée a un modèle conceptuel de données assez simple qui comprend quatre entités et trois processus. Elle stocke cependant une grande quantité d'informations en raison de l'entité variable qui regroupe l'ensemble des informations qui seront utilisées dans le cadre de cette étude. L'entité variable et le processus de traitement des données entomologiques engendrent 13 couches d'information géographique supplémentaires et autant en données tabulaires non géographiques.

A partir de ce riche jeu de données, rien de plus normal que d'entreprendre l'analyse de la structure et des configurations spatiales. Il s'agira de calculer nos indicateurs et évaluer nos deux premières hypothèses (Chapitre 1) qui sont la distribution non aléatoire des formes moléculaires M et S d'*An. gambiae* au Burkina Faso et la contribution différentielle des facteurs environnementaux à l'organisation spatiale de ces vecteurs du paludisme. C'est l'objet du chapitre V.

CHAPITRE V : DISTRIBUTION SPATIALE DES MEMBRES DU COMPLEXE *AN. GAMBIAE* : STRUCTURES GLOBALES ET CONFIGURATIONS LOCALES

Comment se distribuent, au Burkina Faso, les formes moléculaires M et S d'*An. gambiae* s.s. ? Cette distribution est-elle aléatoire ? Sinon, peut-on mettre en exergue une direction préférentielle ? Y a-t-il une influence de la position des points de collecte les uns par rapport aux autres sur les fréquences observées ? Y a-t-il une influence de la distance entre les points de collecte les uns par rapport aux autres sur les fréquences observées ? Ce sont autant de questions auxquelles cette section répond. L'objectif de ce chapitre est de tester les deux premières hypothèses, à savoir l'existence d'une structure spatiale et la contribution différentielle des facteurs environnementaux à l'organisation spatiale. Ce chapitre est composé de trois sections. La première est consacrée à l'analyse de l'organisation spatiale d'*An. gambiae*s.l., la seconde à l'analyse des configurations spatiales locales et la troisième à l'analyse des corrélations entre les distributions spatiales et les facteurs environnementaux.

V.1 L'ORGANISATION SPATIALE D'*An. gambiae* s.s.

Cette étude postule de l'existence d'une structure dans la distribution spatiale des espèces et de l'impact de quelques facteurs environnementaux majeurs dans cette distribution. En d'autres termes, l'espace géométrique (la proximité, distance,...) ne peut pas expliquer entièrement la distribution observée. Pour le vérifier, trois statistiques spatiales comme indicateurs sont retenues. Il s'agit de statistiques descriptives (centrographique et de dispersion) et de l'autocorrelation spatiale. La statistique spatiale descriptive est destinée à résumer l'essentiel des observations. La seconde, l'autocorrelation spatiale vise à étudier l'organisation et la structuration d'*An. gambiae* s.s.. L'organisation spatiale d'*An. gambiae* s.s. au Burkina est abordée d'abord globalement, puis chaque forme moléculaire est analysée de façon indépendante. Au préalable, une présentation des données utilisées pour les analyses est nécessaire. Cette section est subdivisée en quatre paragraphes : le premier est consacré à la présentation des données, les trois autres à la structure spatiale des deux formes moléculaires prises individuellement et *An. gambiae* s.l.

V.1.1 LES STATISTIQUES GENERALES SUR LES DONNEES ENTOMOLOGIQUES COLLECTEES.

La collecte de données a été effectuée de juillet à octobre 2005 dans trois cents localités du Burkina Faso. Au total, 5 496 moustiques capturés ont été identifiés par la PCR Fanello et al, (2002). Une moyenne de 18,32 individus a été identifiée par localité avec un écart type de 6,01. Le maximum de capture identifié est de 39 et a été réalisé à Pimpindiang au nord du Burkina. Le minimum a été réalisé à Mati et à Télékaya, également dans le nord Burkina.

Ces indicateurs globaux se décomposent au plan spécifique conformément à la Figure 37 ci-dessous.

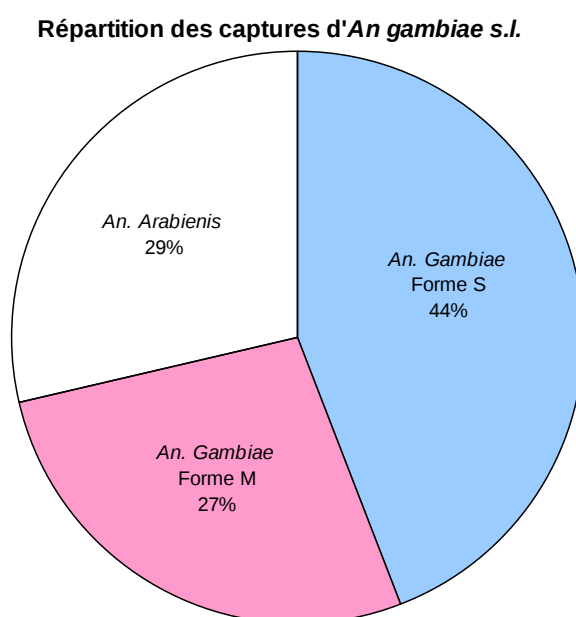


Figure 37 : Répartition des captures d'*Anopheles gambiae*s.l. au Burkina Faso de juillet à octobre 2005

Sur les 5 496 anophèles capturés, 2 422 appartiennent à *An. gambiae* forme moléculaire S, soit 44%, 1 505 à *An. gambiae* forme moléculaire M, soit 27% et enfin 1 569 à *An. arabiensis* soit 29%. D'après ces statistiques, *An. gambiae* forme moléculaire S représente presque la moitié des captures dans le complexe. *An. arabensis* et *An. gambiae* forme moléculaire M sont en proportion presque égale. 3 929 soit 71 % des captures appartiennent à *An. gambiae* s.s..

V.1.2 LA STRUCTURE SPATIALE D'*Anopheles gambiae* s.l.

Au plan spatial, quelques calculs de statistiques descriptives permettent une première caractérisation des données collectées (Figure 38). Il s'agit du barycentre, du centroïde et de l'ellipse de déviation standard. Le barycentre de la distribution des points de collecte se situe pratiquement au centre du pays à quelques kilomètres de Ouagadougou. Il se superpose presque parfaitement avec le centroïde des points de collecte, à une distance inférieure à 3 km et largement inférieure à la distance moyenne entre deux points d'observation qui est de 15 km. Cette quasi superposition du barycentre de distribution et du centroïde suggère une faible influence de la localisation sur les fréquences.

La distance standard permet d'observer une dispersion de 221,1 km autour du barycentre. L'ellipse de déviation standard, quant à lui, est étirée en direction nord-est/sud-ouest et sur une superficie de 119 290,49 km². Le grand rayon a une longueur de 282,35 km contre 134,5 pour le petit, soit pratiquement le double du second, pour un angle de rotation de 65,41°. Ces statistiques indiquent une orientation dans la distribution spatiale d'*An. gambiae*s.l. qui est étirée dans la direction nord-est/sud-ouest. 175 points de collecte sur les 300 sont recouverts par l'ellipse de déviation standard, soit 58% des points de collecte contre 42% qui se situent en périphérie. La dispersion des fréquences des moustiques est donc normale.

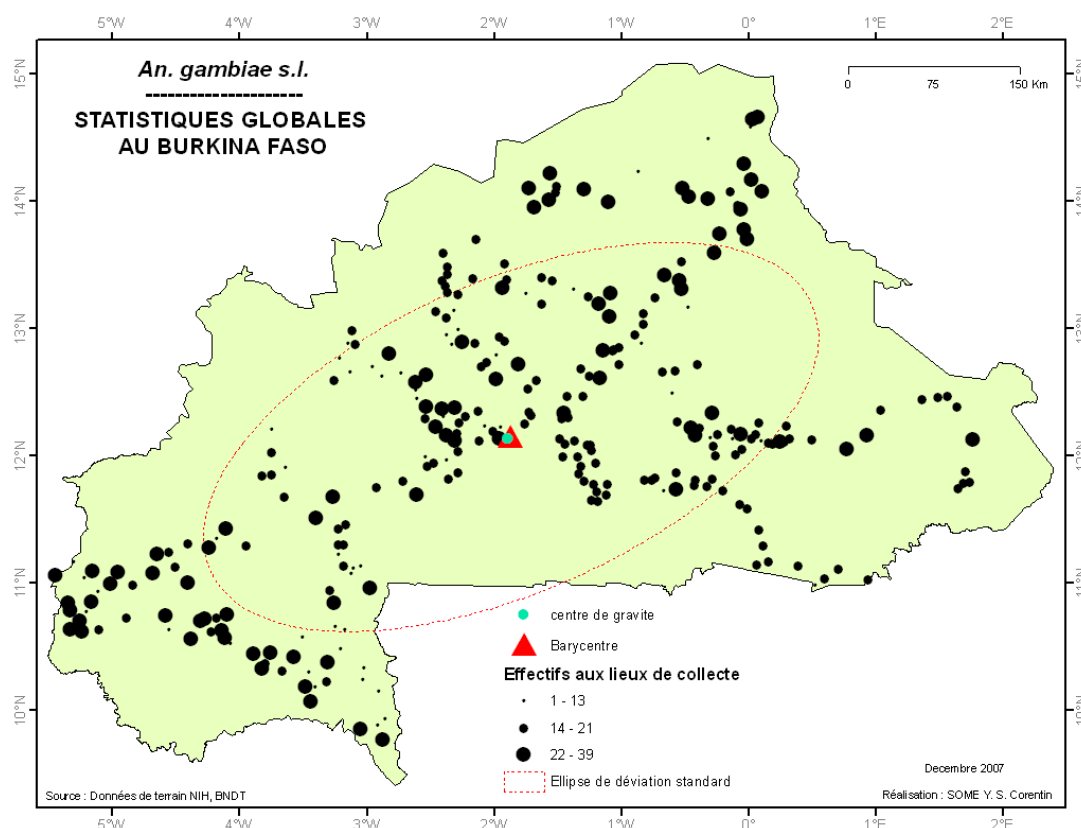


Figure 38 : Structure spatiale de distribution d'*An. gambiae* s.s.

Si la distribution semble normale au regard des indicateurs statistiques présentés, des interrogations demeurent : quelle est le rôle de l'espace dans la distribution des fréquences sur les lieux de collecte ? Existe-t-il des agrégats spatiaux ? L'existence d'une orientation mise en exergue par l'ellipse de déviation standard pose également une question sur l'existence d'une anisotropie. En somme, deux préoccupations restent en suspend. La première concerne l'autocorrelation spatiale et la seconde la configuration spatiale.

Le calcul⁵⁷ de l'indice d'autocorrelation spatiale effectué sur les 300 points de collecte et les fréquences observées sur ces points présentent les résultats suivants : Indice d'autocorrelation spatiale de Moran = 0,014316 ; Indice attendue = -0,003344 ; Variance = 0,000074 ; Z Score = 2,046.

L'indice de Moran est très proche de 0. La réalisation de la variable aléatoire régionalisée est donc indépendante. Cela signifie que les fréquences d'*An. gambiae*.l. en un point de collecte ne sont pas influencées par les fréquences observées dans les points voisins. Il n'est donc pas possible de prédire les fréquences en un point, connaissant les fréquences observées en un autre point, appuyant sur la distance qui les sépare.

Cette indépendance observée dans la réalisation peut être le fruit du hasard et n'a de valeur que pour l'échantillon qui a servi au calcul. Il est donc nécessaire de faire un test pour vérifier la validité de cette indépendance spatiale pour la variable spatiale dans sa totalité. Ce test statistique d'indépendance de la variable régionalisée donne un Z score égale à 2,05 au seuil de confiance de 95%. Il se situe en dehors de l'intervalle de confiance [-1,96 ; 1,96]. Il est difficile d'en déduire l'absence totale d'autocorrelation spatiale sur l'ensemble de la variable. Il existe sans doute des agrégats spatiaux exprimant l'existence locale d'autocorrelation spatiale.

Le calcul des concentrations des valeurs fortes et des valeurs faibles de Getis-Ord General G présente les résultats suivants : Général G Observé = 0,000007 ; Général G attendus = 0,000007 ; Variance de Général G = 0,000000 ; Z Score = -1,35. Ces indicateurs statistiques montrent que les agrégats observés sont le fait du hasard ; l'hypothèse nulle de concentration des faibles et fortes valeurs aléatoires étant acceptée au seuil de 95%.

⁵⁷ Le type de distance utilisé dans le calcul de l'indice d'autocorrelation spatiale de Moran est la distance euclidienne. La conception des relations spatiales entre les fréquences d'*An. gambiae* l.s. dénombré sur les différents points de collecte, s'appuie sur la distance inverse.

L'indice de Getis-Ord General G montre qu'il y a indépendance spatiale dans la distribution d'*An. gambiae* s.l., c'est-à-dire que les fréquences observées en un point ne sont pas influencées par la valeur des fréquences observées dans les points voisins. L'orientation spatiale mise en exergue par l'ellipse de déviation standard ne trouve pas d'explication dans les relations de voisinage. Il est donc important de recourir à d'autres facteurs qui expliquent cette orientation observée. De plus, l'infirmité statistique de la dépendance spatiale de la variable aléatoire régionalisée observée avec l'indice de Moran, impose de pousser l'analyse en vérifiant les relations entre les fréquences observées et l'environnement.

Mais avant, il faut se demander si ces conclusions au niveau *An. gambiae* s.l. sont exactement les mêmes lorsqu'on s'intéresse à l'une ou l'autre forme moléculaire du complexe *An. gambiae* : *An. gambiae* forme moléculaire S et *An. gambiae* forme moléculaire M.

V.1.3 LA STRUCTURE SPATIALE D'*Anopheles gambiae* FORME MOLECULAIRE S

Au plan de la statistique spatiale descriptive, le barycentre de la distribution spatiale d'*An. gambiae* forme moléculaire S se situe complètement au sud du pays à une distance d'environ 125 km au sud-ouest du centroïde. Les fréquences d'*An. gambiae* de forme moléculaire S ont donc fortement influencé la position du barycentre. Ce décalage met également en exergue une tendance à la concentration d'*An. gambiae* de forme moléculaire S dans la partie sud du site d'observation. La distance standard est de 236,04 km. L'ellipse directionnelle a un rayon de 96,45 km observable en abscisse, un rayon en ordonnée de 319,57 km et un angle de rotation 75,41°. L'ellipse de déviation standard couvre une superficie de 96813,72 km² et recouvre 155/300 points d'observation, soit 51,66% de la distribution. Cette répartition des points d'observation par rapport à l'ellipse de déviation standard montre une distribution assez normale (Figure 39).

L'indice d'autocorrélation spatiale de Moran utilisé pour mesurer le niveau de dépendance spatiale donne les résultats suivants : Indice de Moran = 0,28 ; Expected Index = -0,0033 ; Variance = 0,000075 ; Z score = 32,73.

La valeur de l'indice de Moran est très proche de 0. Il n'y a donc pas autocorrélation spatiale dans la distribution observée.

Cependant, le test statistique devant valider cette indépendance spatiale pour l'ensemble de la variable aléatoire régionalisée donne un Z score de 32,72 qui relativise l'interprétation de l'indice

d'autocorrélation de Moran. Ce Z score indique, en effet, l'existence de noyau d'autocorrélation spatiale avec un seuil de confiance de 99%.

Le calcul de l'indice de concentration des valeurs fortes et des valeurs faibles de Getis-Ord General G confirme les statistiques de Moran et donne les valeurs suivantes : Observed General G = 0,000008 ; Expected General G = 0,000007 ; General G Variance = 0,000000 ZScore = 3,87. Il y a sans doute existence d'une autocorrélation spatiale mais elle est locale.

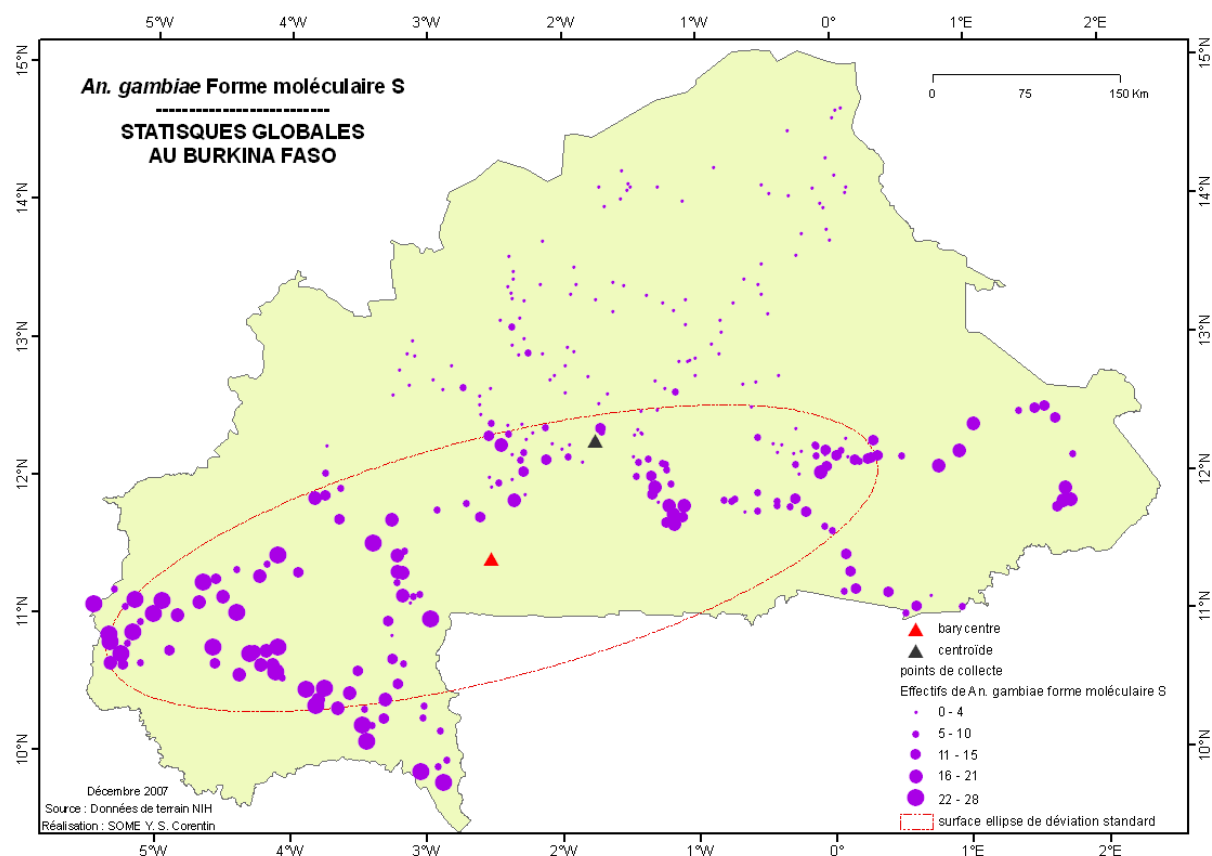


Figure 39: Structure spatiale de distribution d'*Anopheles gambiae* forme moléculaire S

En somme, d'*An. gambiae* forme moléculaire S a tendance à se concentrer au sud du Burkina Faso, dans les régions des Cascades, du Sud-Ouest et des Hauts Bassins. Cette zone correspond à la partie où les précipitations sont les plus abondantes, les températures basses, la végétation la plus dense. C'est la partie la plus humide du pays. Nous pensons donc à une préférence par rapport à l'humidité. La distribution des fréquences se dégrade fortement du sud-ouest vers le nord-est. Nous percevons une similitude entre le front de dégradation des fréquences et celui de l'humidité (Figure 8, Chapitre 2).

Les fréquences observées en un point ne sont pas dépendantes des points à leur voisinage. Il n'est pas convenable de dériver les fréquences en tous points sur la base de la position de points connus. Toutefois, la concentration au sud crée des agrégats, des zones dans lesquelles les fréquences peuvent avoir des similarités. Le sud-ouest et l'ouest du Burkina Faso présentent sans doute des conditions écologiques plus favorables à *An. gambiae* forme moléculaire S. Qu'en est-il pour la forme moléculaire M d'*An. gambiae* ?

V.1.4 LA STRUCTURE SPATIALE D'*Anopheles gambiae* FORME MOLECULAIRE M

La distribution spatiale d'*An. gambiae* forme moléculaire M (Figure 40) se caractérise par les statistiques suivantes : un barycentre décalé par rapport au centroïde de 112 km au nord-est, pour une distance standard de 176,7 km. Il y a donc tendance à une concentration dans la partie nord du site. L'ellipse de direction a un rayon en abscisse de longueur égale à 122 km, un rayon de 218 km en ordonnée et un angle de rotation de 56,84°. L'ellipse couvre 138 points d'observation sur 300 soit 44% des fréquences sur une superficie de 83695,12 km². Nous en déduisons l'existence d'une direction dans la distribution des fréquences d'*An. gambiae* forme moléculaire M. Cette orientation dans la direction permet de soupçonner une dépendance spatiale dans la distribution des fréquences d'*An. gambiae* forme moléculaire M, d'où la nécessité de tester l'autocorrelation spatiale.

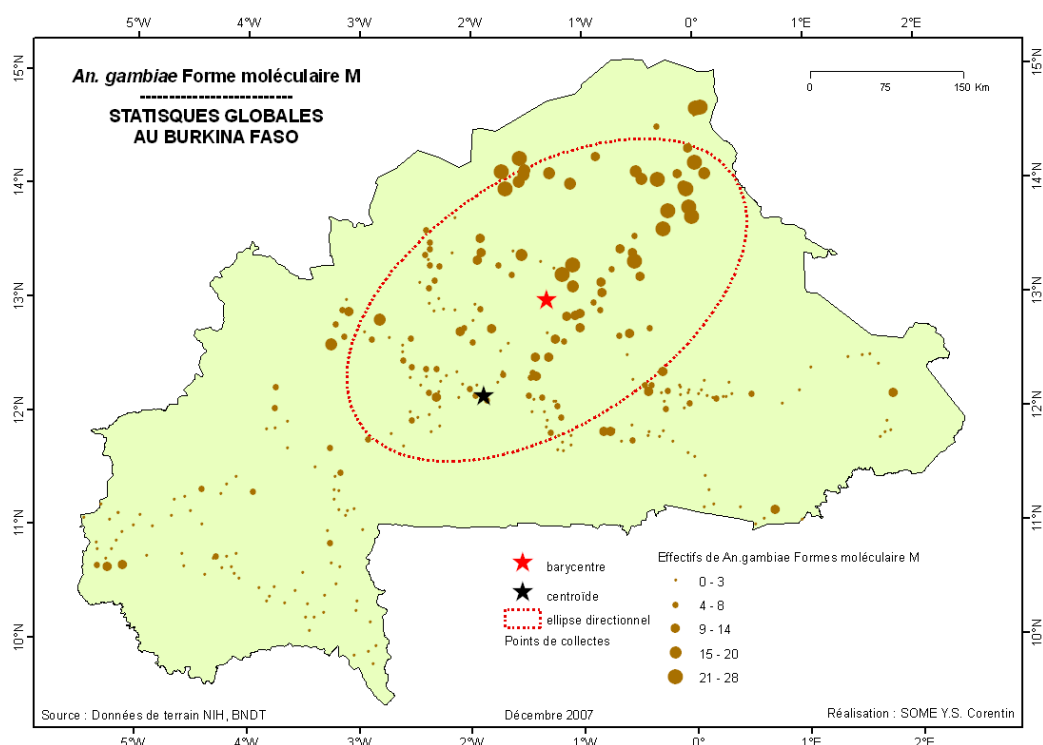


Figure 40 : Structure spatiale de distribution d'*An. gambiae* forme moléculaire M

L'évaluation de l'indice d'autocorrelation de Moran donne les résultats suivants : Indice de Moran = 0,18 ; Indice attendu = -0,0033 ; Variance = 0,000074 ; Z Score = 22,08. L'indice d'autocorrelation de Moran étant proche de 0, il y a indépendance spatiale pour cette réalisation de la variable aléatoire régionalisée. Le Z score indique l'existence de corrélation spatiale localisée à travers des agrégats spatiaux. L'existence de ces agrégats spatiaux est acceptable au seuil de confiance de 99%. L'indice de concentration des valeurs fortes et des valeurs faibles de Getis-Ord General G, avec les résultats suivants, renforce l'interprétation du Z score de l'indice de Moran : Général G Observé = 0,000009 ; Général G attendu = 0,000007 ; Variance du Général G = 0,000000 ; Z Score = 6,71. D'après ces statistiques, il y a indépendance spatiale sur les observations effectuées. Il existe toutefois des agrégats spatiaux à l'intérieur desquels peut s'exprimer une autocorrelation spatiale.

En somme, *An. gambiae* forme moléculaire M a tendance à se concentrer au nord du Burkina. Ses fréquences diminuent du nord-est vers le sud-ouest. Cette partie du Burkina correspond à celle où les précipitations sont les plus faibles, les températures les plus élevées, la végétation la moins luxuriante. C'est le domaine des steppes. C'est la partie la plus aride du pays. Nous pensons donc à une préférence par rapport à l'aridité. Il y a une similitude entre le front d'aggradation des fréquences et celui de l'aridité.

Malgré cette orientation, il n'y a pas d'autocorrélation spatiale même s'il existe des agrégats spatiaux. Puisqu'il y a indépendance spatiale doublée d'une forte concentration des fréquences au nord-est, le nord du pays semble avoir de meilleures conditions écologiques d'*An. gambiae* forme moléculaire M que le sud.

Nous retenons de l'analyse des structures spatiales une tendance à l'allopatrie entre les deux formes moléculaires M et S. La forme S présente de fortes concentrations au sud-ouest et à l'ouest du pays, dans les régions des Cascades, du Sud-Ouest et des Hauts Bassins. La forme moléculaire M, quant à elle, est plus dominante au nord et au nord-est, dans les régions du Sahel, du Nord et du Centre-Nord. Ce résultat d'occupation différentielle de l'espace est en conformité avec celui de Costantini et al., (2009)⁵⁸ qui a été réalisé au Burkina Faso dont nous sommes coauteurs. Il est aussi en phase avec celui de Torre et al., (2005) basé sur des observation au Ghana et au Burkina Faso, d'après lequel "*M and S have different geographical and micro-geographical distributions. The differentiated micro-geographical distribution in dry savannah regions of west Africa may be associated with a diverse ability of the two forms to exploit man-made breeding sites.*" Pour Onyabe et al. (2003), parlant de la distribution spatiale des formes moléculaire M et S *An. gambiae* au Nigeria, "*both molecular forms*

⁵⁸ Voir l'intégralité être trouvé à <http://www.biomedcentral.com/1472-6785/9/16>

occurred throughout the country, with no apparent relationship to the ecological transition from dry savannah in the north to humid forest in southern Nigeria. In all localities, however, 1 form or the other occurred virtually exclusively. Cette forte prédominance d'une forme moléculaire sur l'autre évoque également des zones d'occupation préférentielle.

L'analyse de la structure spatiale a également montré l'absence d'autocorrelation spatiale. Toutefois, elle laisse entrevoir, malgré l'indépendance spatiale globale, la présence d'agrégats spatiaux, témoignant de l'existence, au niveau local, d'autocorrélation spatiale. En d'autres termes, cette analyse suggère l'existence de lieux où les fréquences observées semblent très proches. Il convient donc d'identifier ces agrégats et rechercher les facteurs de ces concentrations de valeurs similaires en certains points.

V.2 L'ANALYSE DES CONFIGURATIONS SPATIALES LOCALES

Où se situent les sous ensembles spatiaux ? Répondre à cette question est l'objectif de cette seconde section. Elle s'inscrit dans la continuité de la précédente pour identifier les agrégats spatiaux. Les indices locaux d'association spatiale d'Anselin Local Moran et Getis-Ord G_i^* sont retenus comme méthodes d'identification des agrégats spatiaux. L'indice d'Anselin Moran permet de quantifier le niveau de similarité/dissimilarité entre points voisins et met en exergue les zones à valeurs homogènes ou hétérogènes. Le G_i^* de Getis-Ord quant à lui permet de faire l'analyse des concentrations (hot spot). Cette section est composée en 6 paragraphes. Chaque forme moléculaire est analysée avec l'indice local d'Anselin Moran et du point de vue de l'indice G_i^* de Getis-Ord après une étude globale de *An. gambiaes.l.*

V.2.1 ANALYSE DES CONFIGURATIONS SPATIALES LOCALES D'*Anopheles gambiaes.l.*

Le calcul de l'indice Local d'Anselin et Moran est effectué sur les fréquences de *An. gambiaes.l.* en vue d'identifier les zones de similarité. Les statistiques sur les valeurs de Z score de l'indice local d'Anselin et Moran pour *An. gambiaes.l.* sur les points d'observation se présentent de la façon suivante : effectif total : 300 ; minimum : -9,08 ; maximum : 5,95 ; total : 62,21 ; moyenne : 0,2 ; déviation standard : 1,62.

La classification de ces valeurs en seuil naturel de Jenks en trois classes et leur spatialisation permet d'obtenir la Figure 41.

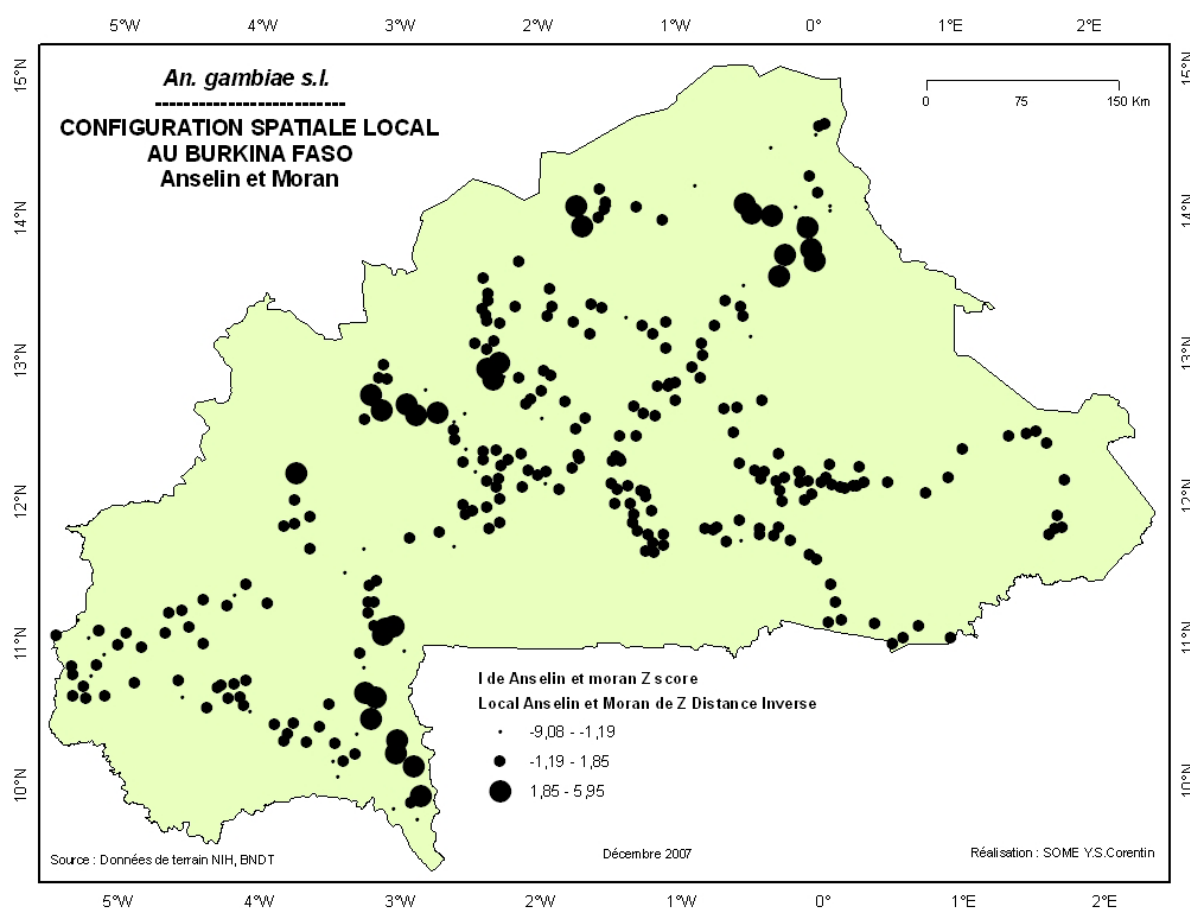


Figure 41 : configuration spatiale d'*Anopheles gambiae*.l.

Les valeurs fortes du Z score représentent les plus fortes similarités des fréquences. Il y a quatre zones de concentration de valeurs fortes. Ces zones de concentration sont dispersées sur l'ensemble du territoire burkinabé. La plupart des points ont des valeurs de Z score moyennes et occupent l'essentiel du pays. Il y a donc un écart entre le zonage et le gradient nord-est/sud-ouest révélé par l'ellipse de déviation standard au niveau de l'analyse globale. Ce même écart s'observe entre le gradient nord-est/sud-ouest mis en exergue dans l'analyse du milieu physique. Le facteur déterminant des concentrations doit alors être fortement localisé et lié à des conditions très particulières. Cette conclusion n'est pas forcément valable pour les formes moléculaires M et S d'*An. gambiae* s.s., prises individuellement.

V.2.2 ANALYSE DES CONFIGURATIONS SPATIALES LOCALES D'*Anopheles gambiae* FORME MOLECULAIRE S

Les valeurs de Z scores de l'indice local d'Anselin et Moran pour *An. gambiae* forme moléculaire S pris sur les points d'observation, se caractérisent par un effectif total de 300 individus, un Z score minimum de -4,34 et un Z score maximum de 20,76 pour une moyenne de 3,52, un écart type de 4,08

et un cumul total de 1058,25. La classification des 300 points d'observation en fonction des Z scores en 3 classes au seuil naturel de Jenks révèle des zones de forte ressemblance (Figure 42).

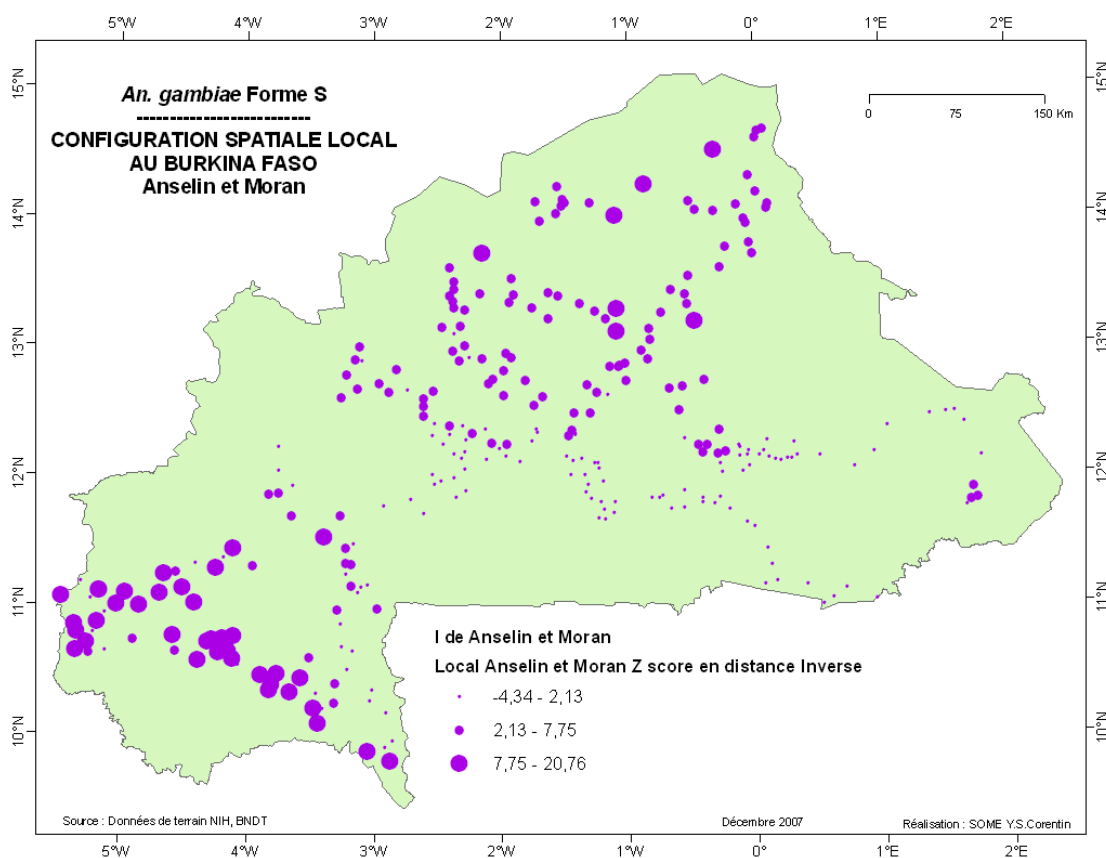


Figure 42 : configuration spatiale d'*Anopheles gambiae* de forme moléculaire S

Les Z score qui ont les valeurs les plus élevées (Z score compris entre 7,75 et 20,77) constituent un agrégat dans l'extrême sud-ouest du pays avec quelques points isolés dans la partie nord. Il y a donc une forte similarité entre les valeurs du Z score dans la partie sud-ouest de la zone d'étude. Ces points isolés à Z score élevés sont noyés dans des Z scores moyens, (Z score compris entre 2,13 et 7,75) dominant dans la partie nord du Burkina. Dans cette partie du pays, la similarité des valeurs est moyenne. Les Z scores les plus faibles abondent dans le sud, le centre et l'est du pays. Ces parties constituent donc une zone très hétérogène. Les valeurs entre les points proches y sont très différentes. Cette configuration indique que le sud-ouest présente des conditions assez homogènes par rapport aux exigences d'*An. gambiae* de forme moléculaire S. Dans la moitié nord du Burkina, il y a une variabilité relativement plus élevée par rapport au sud-ouest. Dans cette zone, les conditions écologiques sont moins homogènes par rapport aux exigences d'*An. gambiae* forme moléculaire S. Dans le centre, les fréquences sont très différentes entre deux points proches. La configuration de la distribution spatiale d'*An. gambiae* forme moléculaire S est donc plus visible par rapport à celle du complexe *An. gambiae*.

An. gambiae de forme moléculaire M a-t-il lui aussi des agrégats très distincts ?

V.2.3 ANALYSE DES CONFIGURATIONS SPATIALES LOCALES D'*Anopheles gambiae* FORME MOLECULAIRE M

A quelques variantes près, la configuration spatiale d'*An. gambiae* forme moléculaire M est l'opposée de celle d'*An. gambiae* forme moléculaire S (Figure 43).

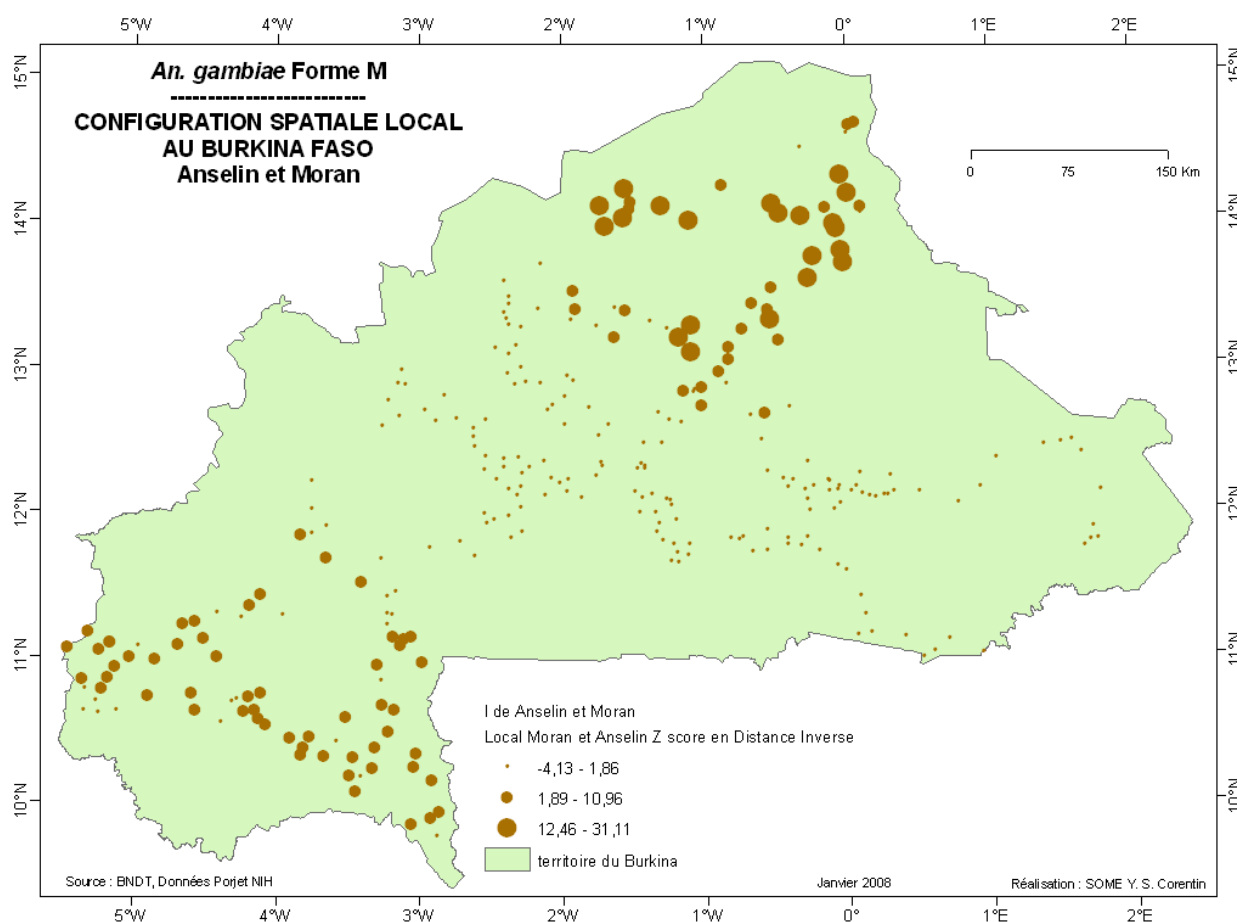


Figure 43 : configuration spatiale d'*Anopheles gambiae* forme moléculaire M

Le nombre d'individus ayant fait l'objet de l'évaluation du Z score est de 300. Les valeurs prises par le Z score sur ces 300 points varient de -4,13 (le minimum) à 31,11 (le maximum) pour une moyenne de 2,5 et un écart type de 5,46. Le cumul de Z score sur l'ensemble des 300 points est de 752,28. La classification en trois classes en seuil naturel de Jenks dégage trois grandes zones de concentrations des valeurs faibles et fortes. Dans le nord, les Z scores sont les plus élevés. Ce qui indique une zone homogène. L'extrême sud-ouest du pays présente les valeurs moyennes de Z score. L'homogénéité des valeurs des Z scores est moins forte que dans le nord. Le centre, le sud et l'est sont très hétérogènes. Les Z scores y sont les plus faibles.

Les agrégats d'*An. gambiae* forme moléculaire M sont également plus marqués que ceux d'*An. gambiae*.l. La position spatiale des agrégats est l'inverse d'*An. gambiae* forme moléculaire S. L'opposition des configurations spatiales d'*An. gambiae* forme moléculaire S et *An. gambiae* forme moléculaire M permet de comprendre l'absence d'agrégats marqués pour l'ensemble du complexe. L'expression localisée d'agrégats doit provenir de l'influence très peu marquée d'*Anopheles arabiensis*, un autre membre du complexe.

Il convient cependant d'analyser de plus près les agrégats. C'est ce que vise l'utilisation de l'indice de GI* de Getis-Ord. Avec cet indice, nous analysons le niveau d'homogénéité des agrégats à travers les similarités qui existent entre deux points voisins d'observés.

V.2.4 CONFIGURATIONS SPATIALES SELON L'INDICE GI* DE GETIS-ORD POUR *Anopheles gambiae*.l.

Le Z score de GI* de Getis-Ord présente les statistiques suivantes pour les 300 points d'observation pour *An. gambiae* sl. Le Z score de GI* minimum est de -2,87, son maximum de 3,43 pour une moyenne de 0,0019 et un écart type de 1,001 ; le cumul de l'ensemble des valeurs étant de 0,57. La classification des Z de GI* en trois classes au seuil naturel de Jenks et la spatialisation donnent la Figure 44.

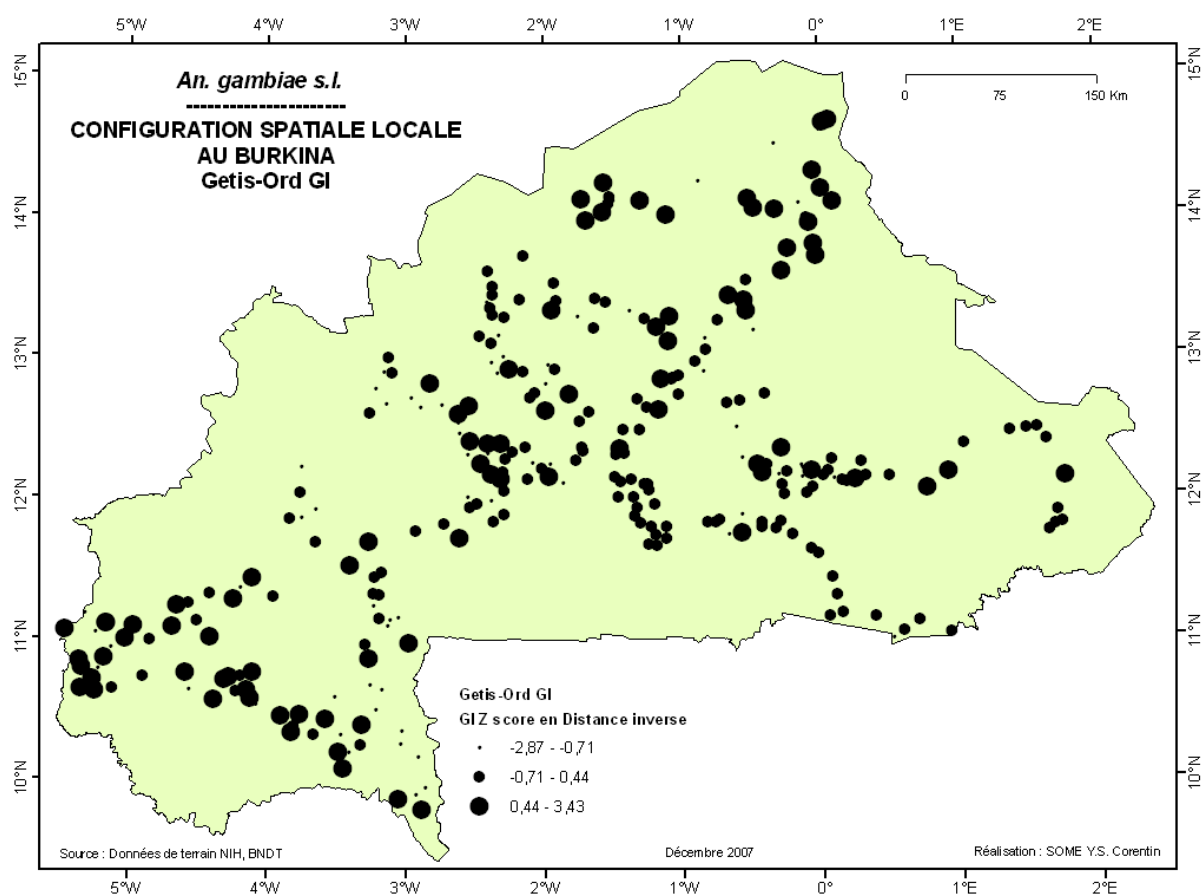


Figure 44 : configuration spatiale d'*Anopheles gambiae*s.l. selon l'indice Getis et Ord

Cette carte montre un fort mélange des valeurs fortes et des valeurs faibles, de sorte qu'il y a une grande mosaïque de petits agrégats spatiaux. La distribution des agrégats est aléatoire et conforte l'interprétation faite de l'analyse effectuée avec l'indice local d'Anselin et Moran. Les agrégats spatiaux ne sont pas marqués et il est difficile, compte tenu de l'organisation du milieu, d'envisager une relation entre l'environnement d'*An. gambiae*s.l. à ce niveau de précision (échelle). Cette analyse pourrait être envisagée à un niveau d'échelle beaucoup plus fin, comme les gîtes. La diversité des exigences spécifiques des membres du complexe *An. gambiae*s.l. est certainement la cause de cette absence de structuration spatiale bien marquée. La réduction de la diversité, en considérant que chaque membre du complexe, pourrait présenter une structure spatiale plus visible et des conditions de mise en relation plus favorables, compte tenu de la diversité du milieu.

V.2.5 CONFIGURATIONS SPATIALES SELON L'INDICE GI* DE GETIS-ORD POUR *Anopheles gambiae* FORME MOLECULAIRE S

Pour *An. gambiae* forme moléculaire S, les résultats de l'évaluation de l'indice GI* de Getis-Ord donnent les statistiques suivantes pour les 300 points d'observation : Minimum : -2,87 ; Maximum : 3,43 ; Moyenne : 0,001907 ; Ecart type : 1,001 et cumul de 0,57. La classification de Z score de GI*

en trois classes au seuil naturel de Jenks et la spatialisation des résultats donnent la Figure 45.

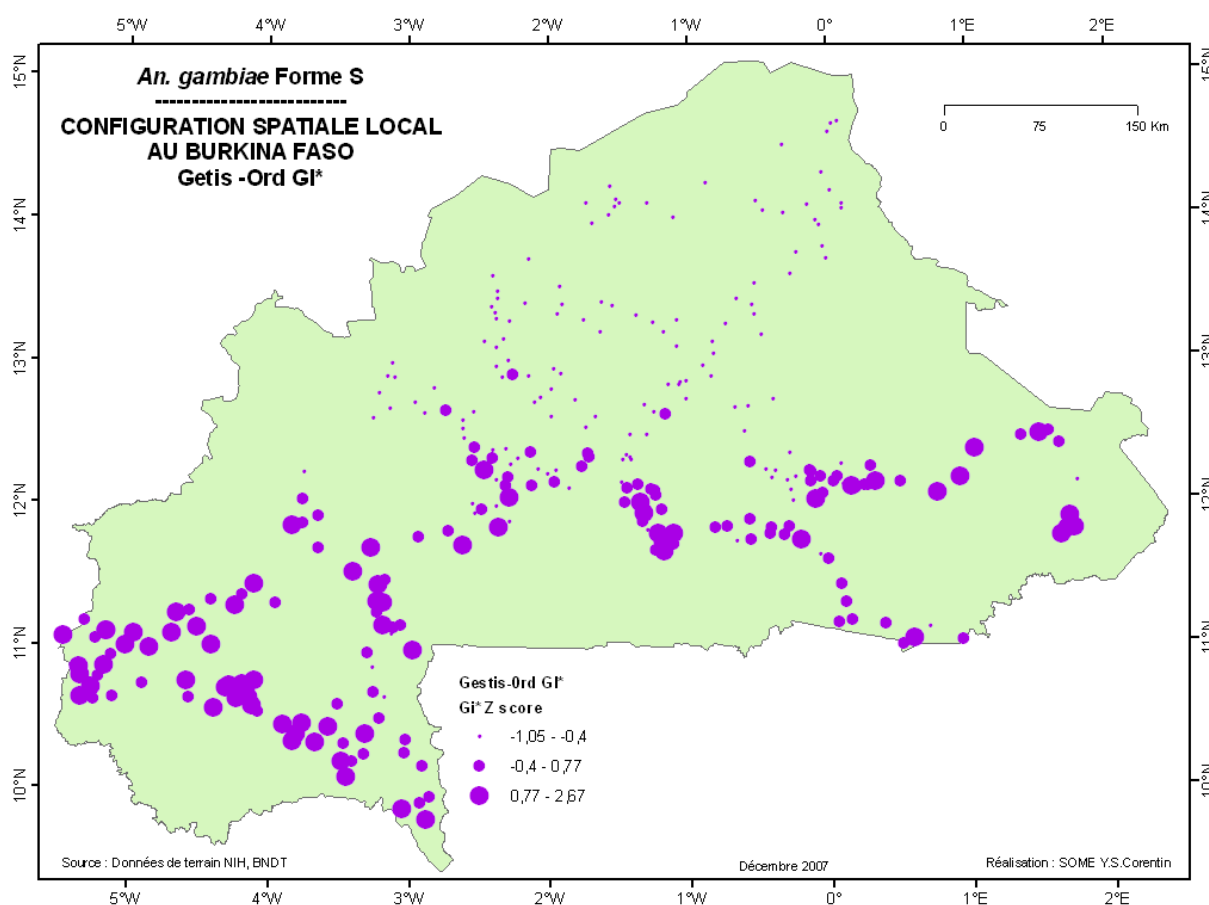


Figure 45 : Configuration spatiale d'*Anopheles gambiae* forme moléculaire S selon l'indice Getis et Ord

Cette carte met en exergue deux grands ensembles spatiaux avec une démarcation presque brusque le long de la latitude 12°30'. En dessous de cette latitude, se concentrent les valeurs fortes et moyennes de Z de GI*. C'est donc une zone de forte concentration *An. gambiae* forme moléculaire S. Au-dessus de 12°30', les valeurs Z de GI* sont très faibles et expriment une forte dispersion des fréquences. Deux grands ensembles bien marqués se distinguent clairement. Cette démarcation observée correspond *grosso modo* à la démarcation entre le domaine soudanien et le domaine sahélien. De plus, les valeurs sont orientées dans le même sens que celui de l'ellipse de déviation standard et l'essentiel des paramètres du milieu biophysique. L'hypothèse de l'existence d'une structuration dans la distribution spatiale des fréquences de la forme moléculaire M est donc vérifiée. Il existe à la fois une orientation dans la distribution spatiale des fréquences et, mieux, une orientation dans la distribution spatiale des agrégats. La distribution n'est donc pas le fait du hasard. Cette hypothèse est-elle également vérifiée pour *An. gambiae* de forme moléculaire M ?

V.2.6 CONFIGURATIONS SPATIALES SELON L'INDICE GI* DE GETIS-ORD POUR *Anopheles gambiae* FORME MOLECULAIRE M

Pour *An. gambiae* forme moléculaire M, les statistiques du Z de Getis pour les 300 points d'observation sont les suivants : le minimum est de -1,13 ; le maximum de 3,67 ; la moyenne de -0,002 ; l'écart type est de 1,002 et le cumul de -0,67. Discretisés en trois classes au seuil naturel de Jenks et spatialisés, nous obtenons la Figure 46.

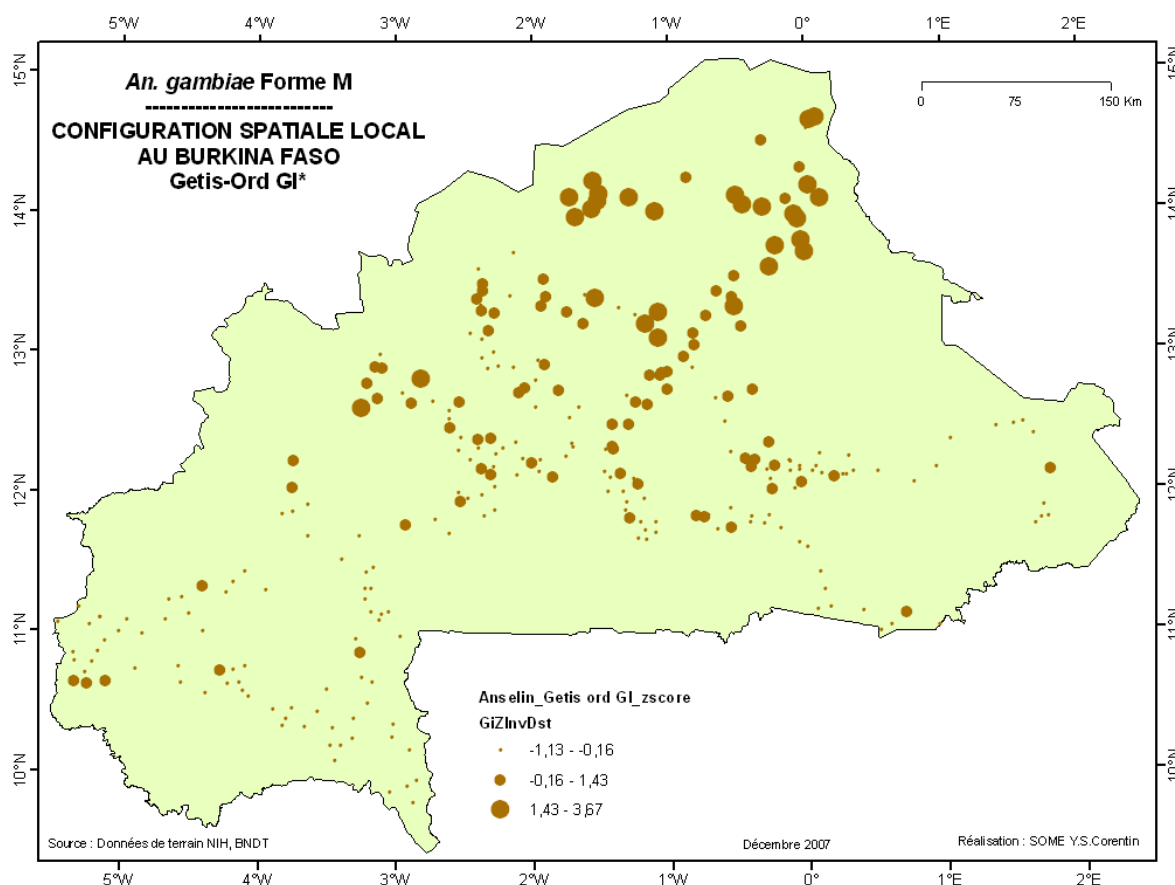


Figure 46 : configuration spatiale d'*Anopheles gambiae* forme moléculaire M selon l'indice Getis et Ord

Deux zones se dégagent à l'observation. Une zone de Z de GI* moyen à élevé, située au dessus de la latitude 12°30' et une zone de Z de GI* faible en dessous de la latitude 12°30'. Il existe dans la partie sud, quelques points isolés à valeurs moyennes. Une structure est tout de même visible : au nord, des niveaux de Z scores relativement identiques et plus élevés dans leur ensemble par rapport au sud. Il y a une orientation nord-est/sud-ouest également dans la disposition des agrégats. Cette orientation est conforme à celle observée avec l'ellipse de déviation standard effectuée sur les fréquences. Cette disposition suit l'organisation du milieu physique. Elle est inversée dans l'ensemble par rapport à celle observée avec *An. gambiae* de forme moléculaire S. L'existence de quelques points de valeurs moyennes dans la partie sud où prédominent des valeurs de Z score faible, suggère un niveau de

complexité certainement plus élevé qu'*An. gambiae* de forme moléculaire S. La détermination de la distribution spatiale d'*An. gambiae* forme moléculaire M devrait présenter plus de difficultés qu'*An. gambiae* forme moléculaire S.

Nous retenons à la suite de cette analyse que l'hypothèse d'une distribution spatiale non aléatoire d'*An. gambiae* forme moléculaire M est vérifiée. En effet, l'ellipse de déviation standard montre une orientation dans la distribution, et des agrégats sont mis en exergue par l'analyse des indices locaux d'association spatiale d'Anselin et Moran et celui de Getis et Ord. Ces agrégats sont également orientés dans la direction mise en exergue par l'ellipse de déviation standard.

Les données entomologiques collectées sont analysées à trois niveaux : le niveau *An. gambiae sl*, le niveau *An. gambiae* forme moléculaire S et le niveau *An. gambiae* forme moléculaire M. La distribution des fréquences observées d'*An. gambiae sl* est assez normale. Celle d'*An. gambiae* forme moléculaire M est centrée au nord-est tandis que celle d'*An. gambiae* forme moléculaire S est plutôt centrée au sud-ouest. Toutes ces distributions sont spatialement indépendantes à l'échelle de collecte des données. Toutefois, elles présentent localement des dépendances spatiales. L'analyse des agrégats et des configurations spatiales révèle une homogénéité spatiale au nord pour *An. gambiae* forme moléculaire M et au sud pour *An. gambiae* forme moléculaire S.

Les statistiques spatiales descriptives, en l'occurrence les ellipses de déviation standard, les barycentres et les centroïdes ont permis de mettre en exergue le caractère orienté et non aléatoire de la distribution spatiale des fréquences du complexe et de ses membres. L'utilisation des indices locaux d'association spatiale, en particulier l'indice d'Anselin et Moran puis l'indice Gi* de Getis et Ord a mis en exergue des agrégats spatiaux orientés dans la même direction que celle des ellipses de déviation standard pour les membres du complexe pris individuellement. Ces indices ont révélé également trois niveaux de structuration spatiale des fréquences :

- *An. gambiae* forme moléculaire S est très bien structuré avec des agrégats marqués et bien repartis dans l'espace ;
- *An. gambiae* forme moléculaire M est également bien structuré mais quelques bruits sont observables dans la zone sud ouest avec l'apparition de quelques agrégats ;
- pour *An. gambiae* s.l., les agrégats sont dispersés.

Ces résultats permettent de dire que l'hypothèse de distribution spatiale non aléatoire des formes moléculaires S et M d'*An. gambiae* s.s. est vérifiée. Une direction et des agrégats ont été également

identifiés. Les tests d'autocorrelation ont mis en évidence l'indépendance spatiale et le fait que l'espace (absolu) est insuffisant pour expliquer la distribution observée.

Si la disposition des objets dans l'espace n'explique pas la configuration spatiale observée du fait de l'indépendance spatiale à l'échelle de collecte, l'existence d'agrégats spatiaux pourrait s'expliquer par un ou plusieurs facteurs mésologiques ou leur combinaison. L'analyse des relations entre la distribution spatiale des fréquences observées et la distribution spatiale des facteurs environnementaux pourrait donner une explication des configurations spatiales observées.

V.3 ANALYSE DES CORRELATIONS ENTRE LA DISTRIBUTION SPATIALE DES VECTEURS ET LES FACTEURS ENVIRONNEMENTAUX

L'espace géométrique, la position relative des points de collecte les uns par rapport aux autres n'ayant pas été suffisantes pour expliquer la structure spatiale observée, la prise en compte des propriétés de l'espace pourrait le faire. Il s'agit donc de considérer plutôt l'espace comme environnement, c'est-à-dire milieu de vie. A ce titre, l'espace offre des conditions écologiques qui s'avèrent plus ou moins contraignantes au développement d'une espèce. Ce sont les facteurs environnementaux. Est-ce que l'offre écologique n'est pas en définitive le déterminant de la structure spatiale observée ? Est-ce que ces facteurs environnementaux ne sont pas finalement les déterminants qui structurent la distribution spatiale des membres d'*An. gambiae* s.s. ? C'est l'objet de cette section.

V.3.1 ANALYSE DES CORRELATIONS ENTRE LA DISTRIBUTION DES VECTEURS ET LES PARAMETRES ENVIRONNEMENTAUX

Si l'espace dans sa dimension géométrique est insuffisant pour expliquer la distribution des fréquences observées, certaines de ses caractéristiques peuvent être plus pertinentes. Il convient donc de vérifier la relation entre les attributs de l'espace en tant que milieu de vie, environnement des formes moléculaires M et S d'*An. gambiae* s.s., et les fréquences observées. L'environnement se positionne dans cette perspective comme facteur susceptible d'expliquer la distribution observée. Il joue donc le rôle de variable explicative.

L'environnement est la résultante de plusieurs facteurs physiques et humains qui jouent de façon différentielle sur la qualité de vie des espèces. Analyser les rapports entre une espèce et l'environnement devient peu confortable si l'on se situe au niveau global, car le concept n'est pas opératoire. Le retour sur l'analyse des relations entre le complexe d'espèces et les différents facteurs

de l'environnement se présente comme une alternative pour identifier l'influence des différents facteurs et évaluer la qualité globale de l'environnement par rapport au complexe d'espèces.

L'environnement se décline à travers 11 paramètres. Ces derniers sont classés en deux groupes de variables : les variables quantitatives et les variables qualitatives. Les variables quantitatives au nombre de 9 représentent 84,61% des variables contre 15,39% pour les variables qualitatives. De plus, les variables à expliquer sont quantitatives. Elles ont été transformées pour être continues par le calcul des abondances relatives (fréquences relatives). En raison de cette forte prédominance des variables quantitatives, l'analyse en composante principale avec variables complémentaires a été utilisée.

V.3.2 CARACTERISATION DES DONNEES ET PARAMETRES DE L'ANALYSE

L'analyse en elle-même est une analyse en composante principale (ACP normée avec une variance en $1/n$), sans axe de rotation. Le coefficient de corrélation est celui de Pearson. Le seuil de signification est de 0,05.

Le tableau soumis à l'analyse compte 13 variables quantitatives et 2 variables qualitatives supplémentaires. Le nombre de variables associées à une valeur propre non triviale est de 13.

La première variable complémentaire, relative à la pédologie, compte 6 modalités que sont : sols hydromorphes, sols isohumiques, sols ferrugineux tropicaux, sols peu évolués et les minéraux bruts ou dégradés, sols à mull, vertisols.

La seconde variable complémentaire, relative à l'occupation des terres, compte 8 modalités : champ, forêt sèche, habitat rural, savane, steppe, territoire agro-forestier, zone humide et zone nue. Les variables quantitatives présentent les caractéristiques statistiques suivantes (**Tableau 9**).

Tableau 9 : statistiques descriptives sur les variables quantitatives

VARIABLE\ STATISTIQUES	Contenu	Moyenne	Ecart - type	Minimum	Maximum
PERCENT_S	Pourcentage d' <i>An. gambiae</i> de forme moléculaire S	44,42	37,23	0	100
PERCENT_M	Pourcentage d' <i>An. gambiae</i> de forme moléculaire M	26,58	28,39	0	100
ETP	Evapo - transpiration potentielle moyenne annuelle	168,52	13,24	143,86	218,51
VENT	Vitesse moyenne annuelle du vent	1,83	0,22	1,17	2,72
TOPO	Topographie/ altitude	320	56	157	556
TEMPER	Température moyenne	28,01	0,42	26,9	28,9
PRESSION	Pression de vapeur moyenne annuelle	18,31	1,26	15,17	21,24
POP	Densité de population	81	93	4,12	1152
PLUIE	Précipitation moyenne annuelle	783	183	327	1133
LOCALITE	Distance par rapport aux localités	1489	2012	0	10510
SOLEIL	Insolation	0,68	0,03	0,61	0,73
HYDRO	Distance par rapport au cours d'eau	492	446	0	3324
NDVI	Végétation / Intensité moyenne de l'activité chlorophyllienne	135,87	22,39	75	173

Conformément à l'échantillon, le nombre d'individus est de 300 pour les variables quantitatives. Les modalités des variables qualitatives dans ce type d'analyse interviennent comme individus. Il y a de ce fait, respectivement 6 et 8 individus supplémentaires provenant des variables qualitatives associées aux sols et à l'occupation des terres.

V.3.3 RESULTATS ET INTERPRETATION DE L'ANALYSE EN COMPOSANTE PRINCIPALE

Le test de sphéricité de Bartlett sur l'ensemble des données donne les valeurs suivantes (Tableau 10) :

Tableau 10 : Les résultats des tests de sphéricité de Bartlett

Khi ² (valeur observée)	4138,618
Khi ² (valeur critique)	99,617
ddl	78
p-value unilatérale	< 0,0001
Seuil de signification Alpha	0,05

Ce test permet de vérifier l'hypothèse nulle selon laquelle toutes les corrélations seraient égales à zéro. Il permet donc de savoir si les données sont factorisables ou pas. Il assure par un test de khi² que les variables sont suffisamment corrélées pour autoriser l'analyse en composante principale. Les résultats de ce test permettent de rejeter l'hypothèse nulle au seuil de signification de 0,05. Cela

permet d'accepter l'hypothèse d'existence de corrélation significative entre les variables. En d'autres termes, les corrélations qui existent entre les variables sont significatives.

V.3.3.1 Analyse des corrélations

Le Tableau 11 permet d'observer les corrélations. Elles sont positives à plus de 0,8 entre le pourcentage d'*An. gambiae* forme moléculaire S, les précipitations, la végétation et les pressions. A l'opposé, elles sont négatives entre les mêmes pourcentages et les variables insolation (-0,82), température (-0,70) et ETP (-0,55).

Tableau 11 : Matrice des corrélations entre les variables

PERCENT_S	1													
PERCENT_M	-0,73	1												
ETP	-0,55	0,65	1											
VENT	-0,33	0,19	0,39	1										
TOPO	0,36	-0,25	-0,22	-0,18	1									
TEMPER	-0,70	0,57	0,68	0,20	-0,56	1								
PRESSION	0,80	-0,71	-0,75	-0,42	0,23	-0,74	1							
POP	-0,20	0,04	-0,00	0,03	-0,09	0,22	-0,14	1						
PLUIE	0,80	-0,69	-0,75	-0,37	0,39	-0,85	0,96	-0,15	1					
LOCALITE	0,05	0,00	0,11	0,05	-0,16	0,092	-0,04	-0,16	-0,08	1				
SOLEIL	-0,82	0,2	0,73	0,46	-0,42	0,84	-0,92	0,21	-0,95	0,01	1			
HYDRO	-0,19	0,04	0,04	0,16	-0,08	0,10	-0,18	0,11	-0,18	-0,07	0,19	1		
NDVI	0,803	-0,75	-0,73	-0,31	0,28	-0,79	0,88	-0,15	0,89	-0,03	-0,86	-0,18	1	
	PERCENT_S	PERCENT_M	ETP	VENT	TOPO	TEMPER	PRESSION	POP	PLUIE	LOCALITE	SOLEIL	HYDRO	NDVI	

En fond jaune, valeurs significatives (hors diagonale) au seuil alpha de 0,050 (test bilatéral)

A l'opposé d'*An. gambiae* forme moléculaire S, *An. gambiae* forme moléculaire M a une forte corrélation positive avec l'ETP (0,65), l'insolation (0,62) et les températures (0, 57). Le niveau de corrélation avec ces variables n'atteint pas celui d'*An. gambiae* de forme moléculaire S. En corrélation négative, ce sont la végétation (-0,75), les pressions (-0,71) et les précipitations (-0,69) qui sont les plus élevées pour *An. gambiae* forme moléculaire M.

Nous remarquons que la corrélation entre l'ETP, l'insolation, les températures et la forme moléculaire M est d'autant plus élevée que la zone est aride. Cela laisse présager une association de la forme moléculaire M à l'aridité. A l'opposé, la forme moléculaire S semble s'associer aux facteurs évoquant plutôt l'humidité comme la pression, la précipitation, la végétation. Il y a donc, en première analyse des corrélations, une opposition entre la forme moléculaire M et la forme moléculaire S. Cette opposition semble se construire le long d'un axe aridité – humidité. Nous constatons cependant que les corrélations entre le pôle de l'humidité et la forme moléculaire S sont plus élevées que celles entre le pôle de l'aridité et la forme moléculaire M.

V.3.3.2 Analyse des valeurs propres

Le **Tableau 12** contient, pour 13 facteurs issus de l'analyse, les valeurs propres, les pourcentages de variances simples et les pourcentages de variances cumulés. La

Figure 47 en fournit une représentation graphique. La moyenne des valeurs propres est égale à 1.

Tableau 12 : les valeurs propres des axes factoriels

	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13
Valeur propre	6,84	1,28	1,07	1,03	0,79	0,72	0,44	0,36	0,14	0,13	0,09	0,03	0,01
% variance	52,63	9,90	8,27	7,93	6,11	5,58	3,45	2,80	1,14	1,03	0,74	0,25	0,09
% cumulé	52,63	62,54	70,82	78,75	84,87	90,46	93,91	96,71	97,86	98,89	99,64	99,90	100,00

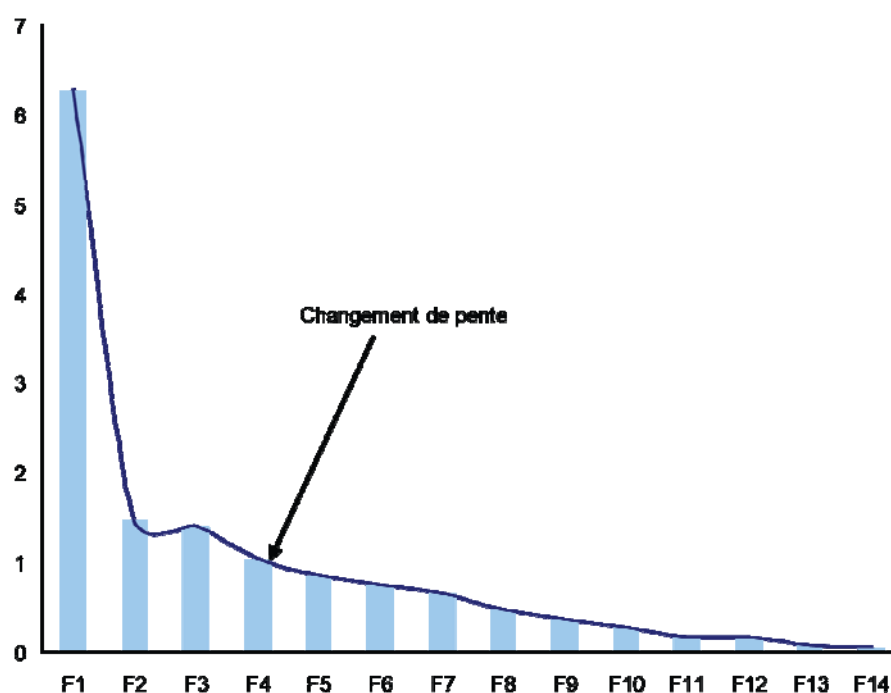


Figure 47 : Histogramme des valeurs propres

Le premier axe représente 52,63 % de l'inertie. Les deux premiers axes représentent à eux seuls 62,54% de l'inertie. En nous référant au critère empirique de Cattell, le critère du Coude, nous retenons comme pertinents les 4 premiers facteurs. Il y a, en effet, une forte rupture de la variation entre l'axe 4 et l'axe 5 comparativement à la pente entre les axes 2, 3 et 4. Le critère de Kayser, qui veut que soient considérés les axes dont les valeurs propres sont supérieures à la moyenne, conduit à retenir les axes 1, 2, 3 et 4. Il confirme également la décision à prendre d'après le critère du coude de Cattell.

V.3.3.3 Analyse de coordonnées des variables

Le Tableau 13 contient les coordonnées des variables sur les quatre premiers axes factoriels retenus au terme de critère de Cattell et de Kayser Lebart et *al.*, (2006). Ces quatre premiers axes comptent 78,75% des variances. Les coordonnées centrées réduites représentent les corrélations entre les variables et les différents axes factoriels.

Tableau 13 : Coordonnées des variables

VARIABLE \ AXE	F1	F2	F3	F4
PERCENT_S	0,870	0,144	-0,041	0,057
PERCENT_M	-0,776	0,074	0,271	-0,174
ETP	-0,803	0,224	0,219	0,057
VENT	-0,438	0,019	-0,027	0,656
TOPO	0,441	-0,157	0,723	0,136
TEMPER	-0,872	0,042	-0,226	-0,227
PRESSION	0,943	0,012	-0,168	-0,065
POP	-0,189	-0,652	-0,347	-0,308
PLUIE	0,968	-0,028	-0,005	0,009
LOCALITE	-0,060	0,732	-0,384	0,132
SOLEIL	-0,961	-0,054	-0,041	0,036
HYDRO	-0,201	-0,467	-0,252	0,612
NDVI	0,929	0,019	-0,130	0,054

A la lecture des valeurs absolues des coordonnées des variables, nous observons que les variables « Pourcentage (abondance relative) » d'*An. gambiae* forme moléculaire M, « Pourcentage (abondance relative) » d'*An. gambiae* forme moléculaire S, ETP, température, pression, précipitation et végétation sont fortement corrélées à l'axe 1 avec des coefficients allant en valeur absolue de 0,776 à 0,968. En dehors de la variable « pourcentage d'*An. gambiae* forme moléculaire M », les coordonnées sont toutes supérieures à 0,8.

A l'axe factoriel 2 sont corrélés les facteurs population (0,652) et localités (0,732). À l'axe factoriel 3 est corrélé à la topographie (l'altitude) avec un coefficient de 0,723 et à l'axe factoriel 4, les vents (0,656) et l'hydrographie (0,612).

La prise en compte du signe dans l'interprétation met en lumière des oppositions entre les variables sur les différents axes. La Figure 48 montre bien ces oppositions sur les axes factoriels 1 et 2 à l'intérieur du cercle de corrélation.

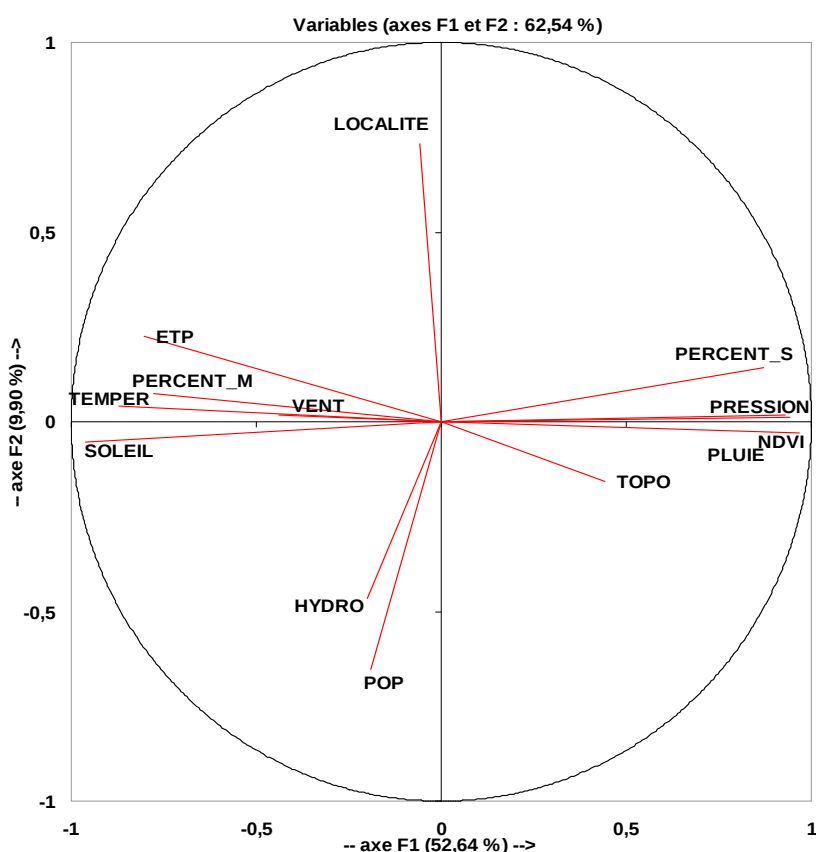


Figure 48 : Représentation des variables sur les plans factoriels 1 et 2

Ainsi, sur l'axe factoriel 1, s'opposent les abondances relatives des 2 formes moléculaires d'*An. gambiae* s.s.. S'y opposent également deux groupes de facteurs environnementaux. Le premier est constitué de la précipitation, de la pression de vapeur et de la végétation. Le second groupe réunit l'ETP, la température, l'insolation et le vent. Le premier groupe de facteurs fait penser aux zones à humidité assez élevée où les températures sont assez fraîches. A l'opposé, se trouvent les facteurs climatiques qui évoquent l'assèchement, l'aridité. Sur l'axe 2 s'opposent les localités et les densités de population.

Sur l'axe 3 représenté sur la Figure 49 avec l'axe 1, nous observons la forte corrélation avec la topographie (altitude).

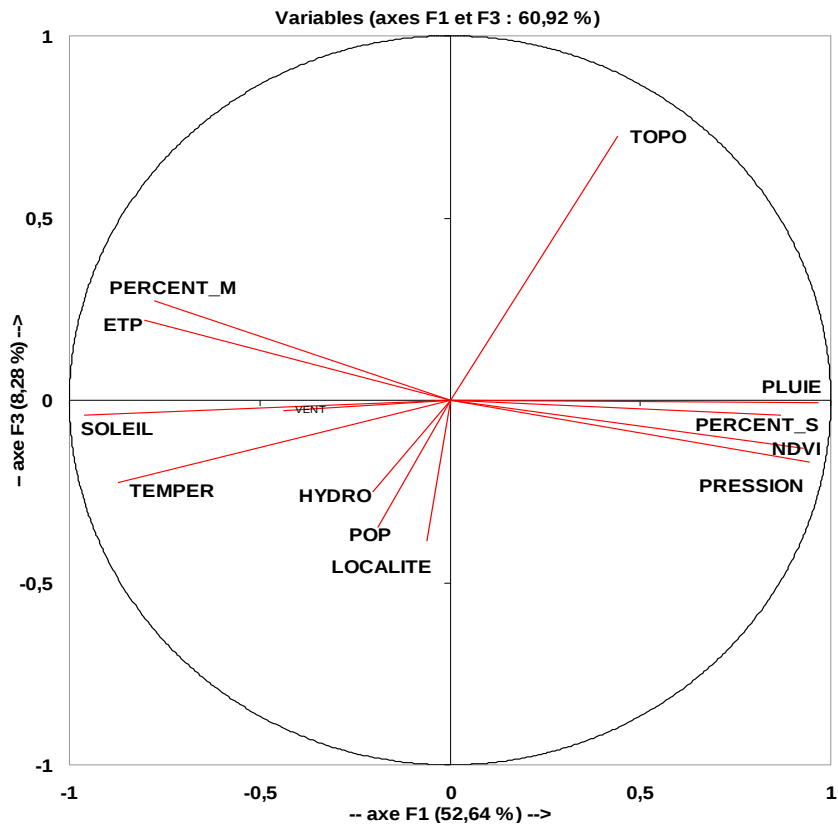


Figure 49 : Représentation des variables sur les plans factoriels 1 et 3Enfin, sur l’axe 4 représenté avec l’axe 1 sur la Figure 50, les vents et l’hydrographie sont très corrélés.

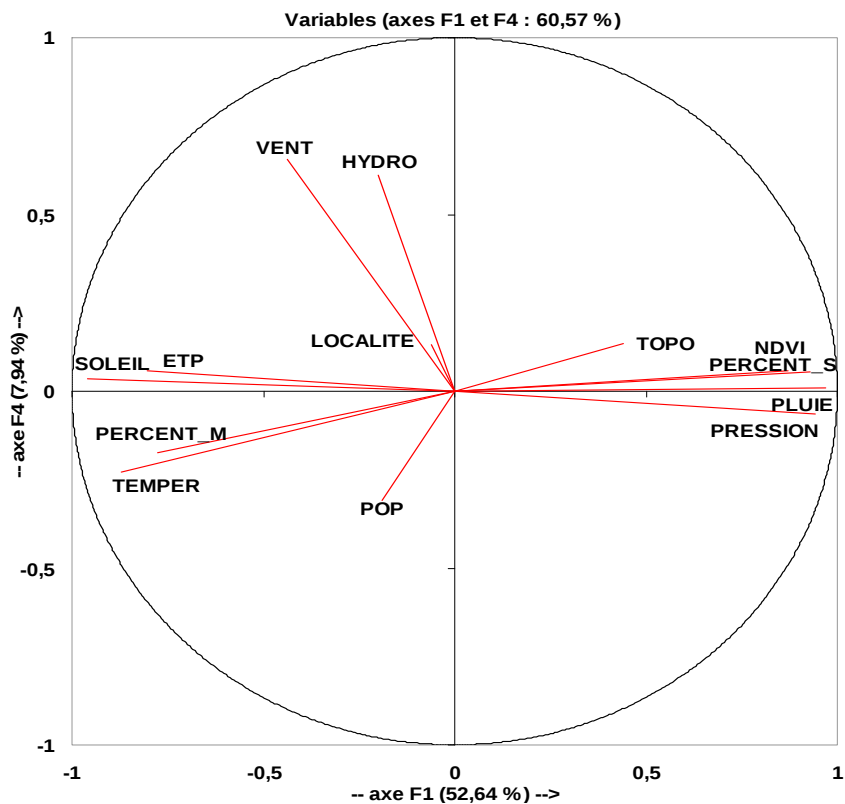


Figure 50 : Représentation des variables sur le plan factoriel 1 et 4

Nous retenons l'opposition entre la forme moléculaire M et la forme moléculaire S d'*An. gambiae* s.s., très remarquable sur l'axe factoriel 1. Cet axe factoriel constitue un axe sur lequel s'opposent également deux pôles de facteurs environnementaux. Le premier pôle regroupe les facteurs NDVI, pluie, pression, topographie (altitude), constitue le pôle humide et est associé à *An. gambiae* forme moléculaire S. Le second pôle regroupe les facteurs ETP, insolation, température, vent et constitue le pôle aride.

L'axe factoriel 2, semble ne pas influencer tellement les deux formes moléculaires. Il oppose les localités à l'hydrographie et aux densités de population. Nous en déduisons que ces trois derniers facteurs environnementaux ne sont pas associés à la distribution des membres du complexe *An. gambiae* s.s..

L'axe factoriel 3 semble avoir la même structure que l'axe factoriel 2, à la seule différence, que la topographie (altitude) n'est plus associée à la répartition des membres du complexe *An. gambiae*. L'axe factoriel 4 exclut les vents comme facteur environnemental associé aux membres du complexe *An. gambiae*.

V.3.3.4 Analyse de la qualité de la représentation des variables par rapport aux différents axes factoriels

La qualité de la représentation des différentes variables par rapport à l'axe factoriel s'exprime à travers les valeurs de cosinus de variables qui sont consignées dans le Tableau 14.

Tableau 14 : Cosinus carrés des variables

VARIABLES \ AXES	F1	F2	F3	F4
PERCENT_S	0,756	0,021	0,002	0,003
PERCENT_M	0,602	0,005	0,074	0,030
ETP	0,645	0,050	0,048	0,003
VENT	0,192	0,000	0,001	0,430
TOPO	0,194	0,025	0,523	0,019
TEMPER	0,761	0,002	0,051	0,052
PRESSION	0,890	0,000	0,028	0,004
POP	0,036	0,426	0,120	0,095
PLUIE	0,938	0,001	0,000	0,000
LOCALITE	0,004	0,536	0,148	0,018
SOLEIL	0,923	0,003	0,002	0,001
HYDRO	0,040	0,218	0,063	0,374
NDVI	0,862	0,000	0,017	0,003

L'analyse de ce tableau révèle que les pourcentages (abondance relative) d'*An. gambiae* forme moléculaire S, la température, la pression, les précipitations, l'insolation, la végétation sont très bien représentés sur l'axe 1, les carrés de leurs cosinus étant supérieurs à 0,6 sur cet axe. Les variables

« pourcentage d'*An. gambiae* forme moléculaire M et ETP » y sont moyennement bien représentées avec des carrés de cosinus compris entre 0,4 et 0,65. Sur l'axe factoriel 2, les localités avec un cosinus carré égale à 0,536 et les densités de population avec un cosinus carré de 0,426 sont moyennement bien représentées. Aucune variable n'y est bien représentée. Cette mauvaise représentation de variables s'observe sur les deux autres axes : l'axe 3 et l'axe 4. Sur ces deux derniers axes, la variable topographie (altitude) a une bonne représentation sur l'axe factoriel 3 et les vents sur l'axe factoriel 4.

V.3.3.5 Analyse de la contribution des variables à la formation des axes factoriels

Les contributions des variables aux axes factoriels sont consignées dans le Tableau 15. La Figure 51 en donne une représentation graphique.

Tableau 15 : Contributions des variables (%)

VARIABLES \ AXES	F1	F2	F3	F4
LOCALITE	0,052	41,646	13,727	1,696
POP	0,524	33,059	11,160	9,221
HYDRO	0,591	16,962	5,880	36,255
VENT	2,804	0,028	0,066	41,645
TOPO	2,837	1,904	48,610	1,797
PERCENT_M	8,801	0,426	6,849	2,926
ETP	9,421	3,905	4,438	0,320
PERCENT_S	11,053	1,605	0,155	0,310
TEMPER	11,117	0,139	4,754	5,007
NDVI	12,600	0,027	1,577	0,281
PRESSION	13,004	0,011	2,624	0,408
SOLEIL	13,493	0,223	0,158	0,125
PLUIE	13,702	0,062	0,003	0,009

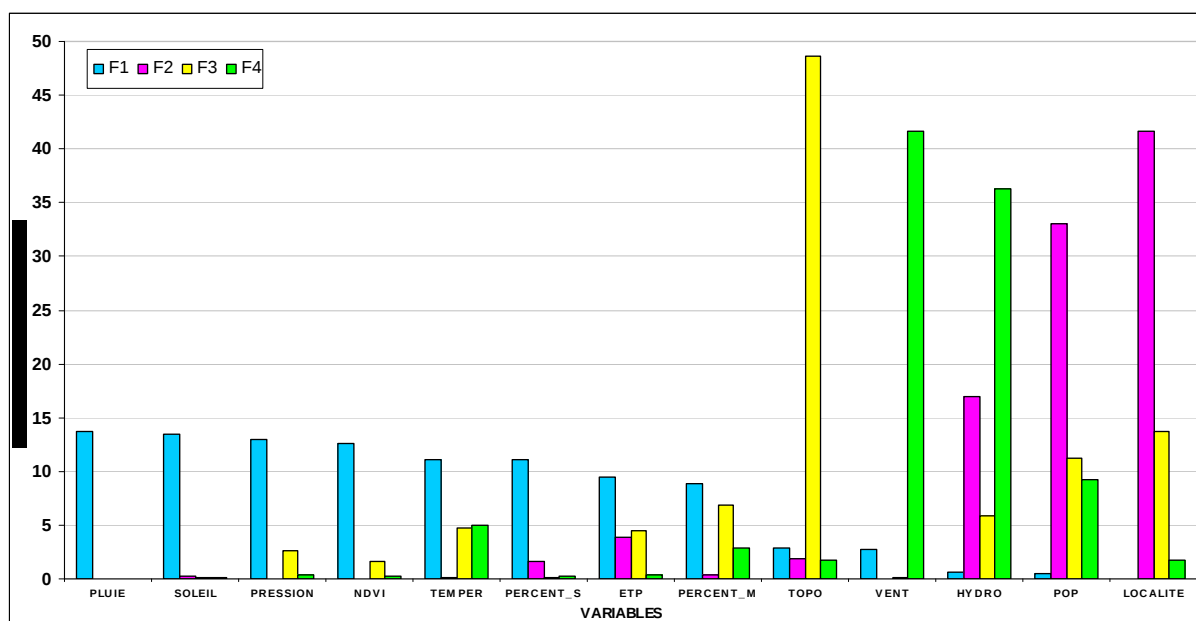


Figure 51 : contribution des variables aux axes factoriels

Les contributions maximales à l'axe factoriel 1 ne dépassent guère 13,5%. De nombreux facteurs contribuent donc à la définition de cet axe. Nous distinguons trois groupes de variables sur l'axe 1 :

- les variables dont les contributions varient entre 15% et 12,5% constituées des précipitations, de l'insolation, la pression et la végétation ;
- les variables dont les contributions varient entre 12,5% et 7,5% avec les variables températures, pourcentage d'*An. gambiae* forme moléculaire S, ETP, pourcentage d'*An. gambiae* forme moléculaire M ;
- les variables avec des contributions très faibles, inférieures à 7,5%, que sont, la topographie (altitude), les vents, l'hydrographie, les densités de population et les localités.

Nous considérons donc que les variables précipitation, insolation, pression et température contribuent au premier axe factoriel et auquel sont également associées les formes moléculaires M et S.

Les autres axes factoriels disposent de deux ou trois variables aux contributions très élevées par rapport aux restes de variables dépassant parfois les 40% des contributions. Ainsi, sur l'axe factoriel 2, les localités représentent plus de 40% des contributions. Celles-ci, associées à l'hydrographie et à la population, représentent plus de 90% des contributions. Pour l'axe factoriel 3, près de la moitié des contributions est détenue par la topographie (altitude). Après la topographie (altitude), les contributions les plus élevées des autres variables n'atteignent pas les 15% et concernent seulement deux variables : localité et population. Sur l'axe factoriel 4, des vents (41,64%) et de l'hydrographie (36,25%) sont dominants. Les variables hydrographie, population, localité et vent ont des

contributions intéressantes sur les axes sur lesquels la contribution des formes moléculaires M et S d'*An. gambiae* s.s. sont très faibles. Nous en déduisons que les formes moléculaires M et S ne sont pas associées à ces axes. Les variables qui contribuent aux axes 2, 3, 4 contribuent donc peu à la détermination de la distribution des formes moléculaires M et S d'*An. gambiae* s.s.. L'occupation des terres a été présentée comme un facteur écologique pouvant influencer la distribution spatiale d'*An. gambiae* s.s.. Aussi, est-il important d'apprécier le positionnement des modalités de cette variable qualitative par rapport aux différents axes factoriels.

V.3.3.6 Contribution des variables supplémentaires aux différents axes factoriels

Les coordonnées des variables supplémentaires sur les 4 premiers axes factoriels sont représentées dans le Tableau 16.

Tableau 16 : Les coordonnées des variables supplémentaires

VARIABLES \ AXES	F1	F2	F3	F4
Champ	0,065	-0,006	-0,117	0,034
Forêt sèche	2,449	-0,007	1,123	0,455
Habitat rural	-0,072	-0,564	0,943	-0,186
Savane	0,644	0,103	-0,172	0,081
Steppe	-3,572	0,755	0,359	-0,276
Territoire agro-forestier	0,755	-0,113	0,055	-0,027
Zone humide	-0,737	0,237	-0,640	-1,424
Zone nue	-3,980	0,334	1,121	-0,363

Les Figure 52, Figure 53 et Figure 54, ci-dessous présentent la distribution des individus supplémentaires (puisque la variable supplémentaire est nominale) sur les 4 axes factoriels.

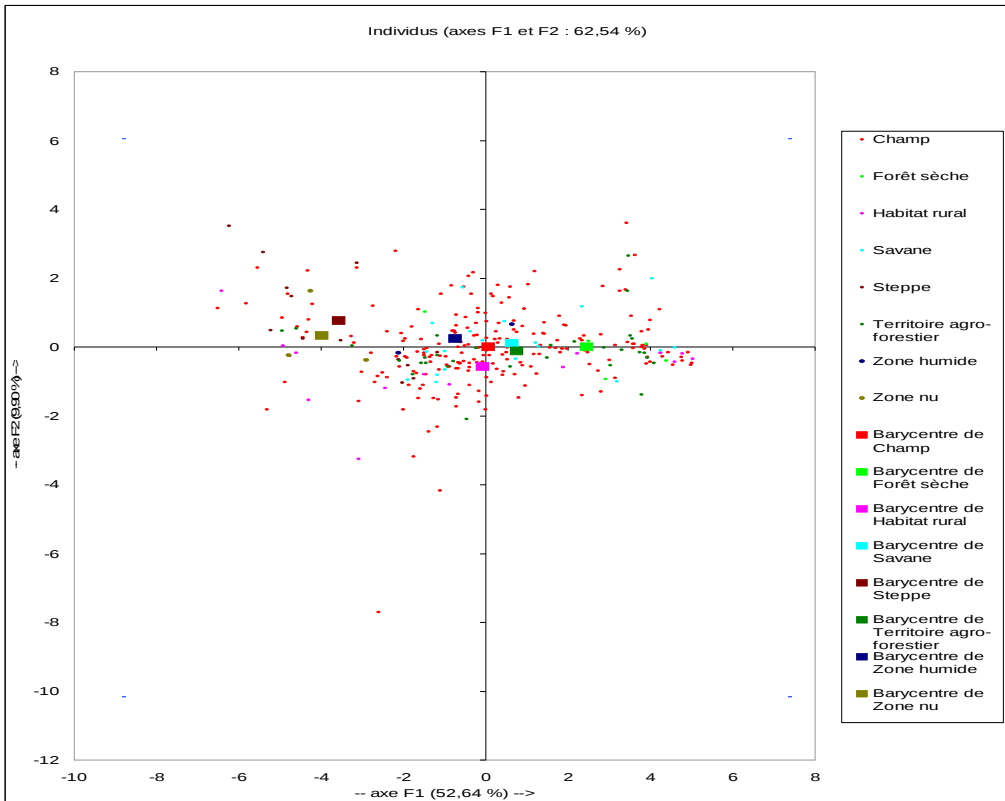


Figure 52 : Positionnement de nominales supplémentaires sur les plans factoriels 1 et 2

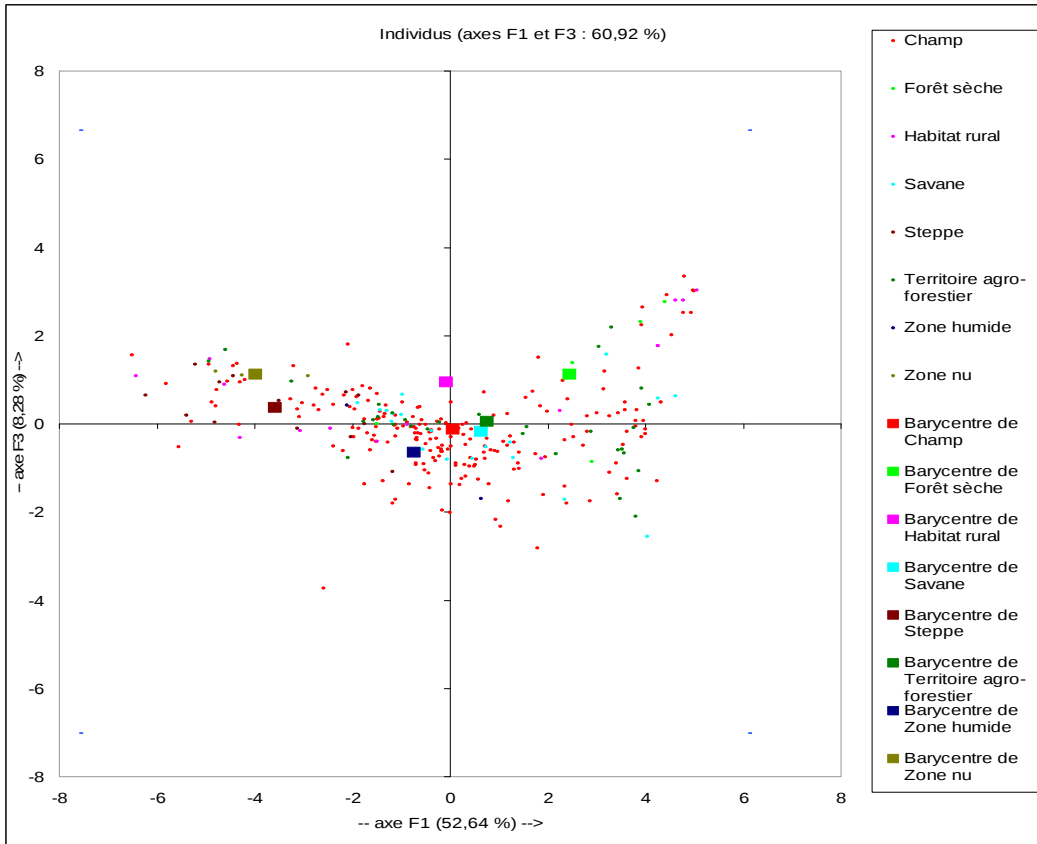


Figure 53 : Positionnement de nominales supplémentaires sur les plans factoriels 1 et 3

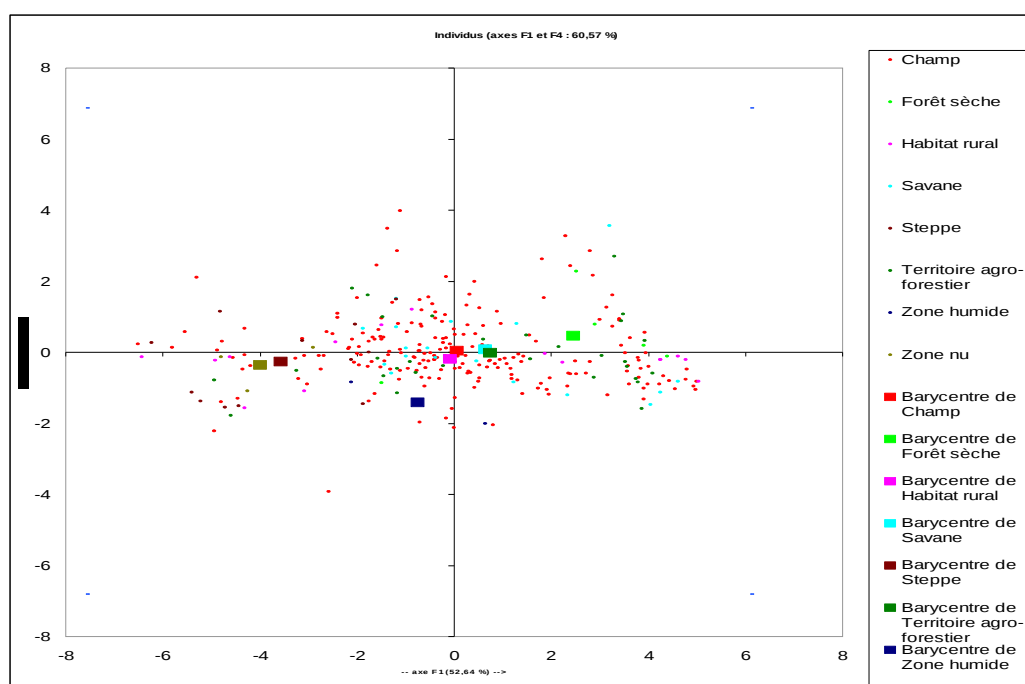


Figure 54 : Positionnement de nominales supplémentaires sur les plans factoriels 1 et 4

La position des barycentres des individus supplémentaires issus de la variable nominale « occupation des terres » le long de l'axe factoriel 1, et l'opposition entre steppe, zone nue et forêt sèche, confirment l'aridité comme élément fondamental de séparation des variables. Elle met aussi et surtout en exergue l'opposition entre *An. gambiae* forme moléculaire M et *An. gambiae* forme moléculaire S.

De cette analyse en composantes principales, nous déduisons :

1. qu'il existe des relations entre les facteurs de l'environnement et la distribution des formes moléculaires M et S d'*An. gambiae* s.s. ;
2. qu'il existe une opposition entre les préférences des formes moléculaires M et S d'*An. gambiae*. La forme moléculaire M est associée aux facteurs environnementaux évoquant l'aridité comme la température, l'ETP, l'insolation et la steppe. Elle préfère visiblement les régions les plus arides, correspondant au nord du Burkina Faso. *An. gambiae* forme moléculaire S quant à lui est associé aux facteurs environnementaux qui évoquent l'humidité comme la précipitation, la pression, les savanes et le NDVI. Il préfère donc les régions humides du pays c'est-à-dire le sud-ouest du Burkina Faso. Le pôle de l'aridité et celui de l'humidité constituent les extrémités d'un axe qui se présente à première vue comme le paramètre déterminant de cette opposition ;
3. que la liaison entre les facteurs environnementaux et *An. gambiae* forme moléculaire S est plus évidente que celle d'avec *An. gambiae* forme moléculaire M ;

4. que les facteurs écologiques « hydrographie » et « densité de population » ne présentent pas un intérêt dans la détermination de la distribution ; leurs contributions sont donc différentes des autres paramètres déjà cités et associés à l'axe factoriel 1, l'axe humidité aridité ;
5. que l'occupation des terres présente des modalités qui pourraient apporter un plus dans la modélisation de la distribution spatiale des vecteurs.

En somme, la deuxième hypothèse est vérifiée. En effet, tous les facteurs écologiques ne contribuent pas de la même façon à la détermination de la structure spatiale observée. Certains facteurs environnementaux peuvent devenir plus facilement une contrainte importante, un facteur limitant au développement et à la propagation de l'espèce.

D'après les résultats de l'analyse, nous pouvons :

- procéder à la réduction du nombre de variables explicatives par exclusion des variables hydrographie et densité de population ;
- procéder à une modélisation séparée des deux membres du complexe ;
- retenir l'analyse des covariances (ANCOVA) comme technique de modélisation. Elle est adaptée à la modélisation incluant des variables qualitatives. Toutefois, elle est remplacée si la contribution de ces variables qualitatives à la modélisation est faible et insuffisante.

CONCLUSION DU CHAPITRE V

Les données entomologiques collectées sont analysées à trois niveaux : le niveau *An. gambiae* *sl*, le niveau *An. gambiae* forme moléculaire S et le niveau *An. gambiae* forme moléculaire M. La distribution des fréquences observées d'*An. gambiae* *sl* est assez normale. Celle d'*An. gambiae* forme moléculaire M est centrée au nord-est tandis que celle d'*An. gambiae* forme moléculaire S est plutôt centrée au sud-ouest. Toutes ces distributions sont spatialement indépendantes à l'échelle de collecte des données. Toutefois, elles présentent localement des dépendances spatiales. L'analyse des agrégats et des configurations spatiales révèle une homogénéité spatiale au nord pour *An. gambiae* forme moléculaire M et au sud pour *An. gambiae* forme moléculaire S.

Les statistiques spatiales descriptives, en l'occurrence les ellipses de déviation standard, les barycentres et les centroïdes ont permis de mettre en exergue le caractère orienté et non aléatoire de la distribution spatiale des fréquences du complexe et de ses membres. L'utilisation des indices locaux d'association spatiale, en l'occurrence l'indice d'Anselin et Moran et G_i^* de Getis et Ord ont permis de montrer l'existence d'agrégats spatiaux orientés dans la même direction que celle des ellipses de déviation standard pour les membres du complexe pris individuellement. Ces indices ont révélé également trois niveaux de structuration spatiale des fréquences :

- *An. gambiae* forme moléculaire S est très bien structuré avec des agrégats marqués et bien repartis dans l'espace ;
- *An. gambiae* forme moléculaire M est également bien structuré mais il y a quelques bruits dans la zone sud-ouest;
- pour *An. gambiae* s.l., les agrégats sont dispersés.

Ces résultats permettent de dire que l'hypothèse de distribution spatiale non aléatoire des formes moléculaires S et M d'*An. gambiae* s.s. est vérifiée. Une direction et des agrégats ont été également identifiés. Les tests d'autocorrelation montre l'indépendance spatiale et le fait que l'espace (absolu) est insuffisant pour expliquer la distribution observée.

Les principaux résultats de l'analyse des relations entre la distribution spatiale des fréquences observées et la distribution spatiale des facteurs environnementaux obtenus sont la mise en exergue des zones de colonisation préférentielle opposées entre la forme moléculaire M et la forme moléculaire S d'*An. gambiae* au Burkina Faso. La forme moléculaire S préfère les régions du sud les plus humides tandis que la forme moléculaire M préfère les régions du nord arides.

Les facteurs les plus déterminants de la distribution spatiale de la forme moléculaire S sont : l'ETP, l'insolation et le NDVI. Les deux premiers sont négativement corrélés avec l'abondance de la forme moléculaire S. L'abondance relative de S est positivement corrélée au NDVI et exprime la préférence de cette forme pour les milieux les plus humides du pays.

Pour la forme moléculaire M, les facteurs les plus déterminants dans la distribution spatiale sont : l'altitude, la pression, l'insolation et la densité de végétation (NDVI). L'altitude, la pression et le NDVI sont tous d'autant plus élevés que le milieu est humide et sont tous en corrélation négative avec l'abondance de la forme moléculaire M. L'insolation qui évoque l'aridité est en liaison positive avec l'abondance de la forme moléculaire M, indiquant la préférence de cette forme pour les milieux arides.

Compte tenu de la lisibilité de ces tendances observées, il est tout à fait possible que cette distribution des formes moléculaires puisse être décrite de façon formelle par une fonction. C'est l'objet du chapitre 6 dans lequel nous procédons à l'expression des abondances relatives des deux formes moléculaires M et S en fonction des facteurs environnementaux.

CHAPITRE VI : MODELISATION DE LA DISTRIBUTION SPATIALE

L'analyse des relations entre l'environnement et la distribution de chaque forme moléculaire d'*An. gambiae* s.s., laisse penser à une possibilité de formalisation mathématique de leur distribution en s'appuyant sur les facteurs environnementaux. De plus, il est apparu nécessaire de considérer séparément chaque membre du complexe, compte tenu des différences observées dans leurs relations avec l'environnement.

Plusieurs techniques statistiques peuvent être utilisées pour formaliser les relations entre l'environnement et les différents membres du complexe. Dans le cadre de cette étude, l'analyse de la covariance (ANCOVA) et la régression multiple sont utilisées pour permettre de prendre en compte à la fois les variables quantitatives et les variables qualitatives.

VI.1 MODELISATION PAR ANALYSE DE COVARIANCE (ANCOVA)

CONFIGURATION DE L'ANCOVA ET DESCRIPTION DES DONNEES

L'analyse est sans contrainte (contrainte : $\alpha_1 = 0$), les données contribuent de façon uniforme à la modélisation. Aucune contrainte particulière n'est ajoutée. Deux tests statistiques sont effectués pour évaluer la contribution des variables au modèle. Il s'agit du test Fisher Type I SS et du Test de Fisher type III SS.

Les données d'entrée dans cette analyse de covariance se présentent comme suit :

Les variables à expliquer sont les pourcentages, d'*An. gambiae* forme moléculaire S (pourcent_S) et les pourcentages d'*An. gambiae* forme moléculaire M (pourcent_M).

Les deux variables à expliquer sont caractérisées dans le Tableau 17.

Tableau 17 : Statistiques descriptives sur les variables à expliquer

Variable	Nbre de valeurs totale	Nbre de valeurs utilisées	Nbre de valeurs ignorées	Somme des poids	Moyenne	Ecart-type
POURCENT_M	255	255	0	255	27,755	28,448
POURCENT_S	255	255	0	255	42,066	37,135

Les variables explicatives sont au nombre de 13, dont 11 quantitatives et 2 qualitatives. Le résumé des

variables explicatives quantitatives⁵⁹ pour les deux variables à expliquer est contenu dans le Tableau 18.

Tableau 18 : Statistiques descriptives sur les variables explicatives

VARIABLE	PARAMETRE REPRESENTE	MOYENNE	ECART-TYPE
ETP	Evapo - transpiration potentielle moyenne annuelle	169	13
VENT	Vitesse moyenne annuelle du vent	1,841	0,221
TOPO	Topographie/ altitude	320	54
TEMPER	Température moyenne	28	0,41
PRESSION	Pression moyenne annuelle	18,257	1,264
POP	Densité de population	79	72
PLUIE	Précipitation moyenne annuelle	776	183
LOCALITE	Distance par rapport aux localités	1543	2083
SOLEIL	Insolation	0,68	0,03
HYDRO	Distance par rapport au cours d'eau	496	452
NDVI	Végétation / Intensité moyenne de l'activité chlorophyllienne	134	22

Quant aux variables explicatives qualitatives, elles sont résumés dans le Tableau 19.

Tableau 19 : Statistiques descriptives sur les variables qualitatives

VARIABLE	PARAMETRE REPRESENT	NOMBRE DE MODALITES	MODALITES	FREQUENCES D'OBSERVATION
OCTER_NOMA	Occupation des terres	8	Habitat rural	15
			Champ	180
			Territoire agro-forestier	27
			Steppe	10
			Zone nue	3
			Savane	17
			Zone humide	1
			Forêt sèche	2
SOL	Formation pédologique	6	Sols ferrugineux tropicaux	93
			Sols Hydromorphes	68
			Sols peu évolués, minéraux bruts ou dégradés	67
			Sols à Mull	14
			Vertisols	12
			Sols Isohumiques	1

RESULTATS ET INTERPRETATION DE L'ANCOVA

⁵⁹ Volontairement, les variables identifiées comme peu déterminantes par l'analyse en composantes principales ont été maintenues comme variables, car les modules de calcul disposent de test d'analyse de leur contribution aux modèles.

L'ANCOVA est effectuée de façon séparée pour chaque forme moléculaire d'*An. gambiae* s.s.. Aussi, la présentation des résultats est-elle faite de manière à refléter cette séparation entre les deux formes moléculaires.

VI.1.2.1 Résultats et interprétation de l'ANCOVA pour *An. gambiae* forme moléculaire S

L'ANCOVA donne les résultats suivants pour *An. gambiae* forme moléculaire S :

Pour *An. gambiae* forme moléculaire S, nous obtenons, un coefficient de corrélation $R = 0,880$, un coefficient de détermination $R^2 = 0,774$, un coefficient de détermination ajusté $R^2_{aj} = 0,751$ et un $SCR = 79193,944$. Nous retenons de ce résultat que 77% de la variance observée dans la distribution d'*An. gambiae* forme moléculaire S s'explique par les 11 variables d'entrées quantitatives.

Le test de Fisher associé à l'ANCOVA dans son ensemble permet d'évaluer d'abord la significativité de la quantité des informations fournies par les 13 variables (quantitative et qualitative) au modèle. Les résultats du test sont contenus dans le Tableau 20.

Tableau 20 : Test de significativité de la quantité d'information apportée par les variables

Source	ddl	Somme des carrés	Carré moyen	F de Fisher	Probabilité > F
Modèle	23	271070,777	11785,686	34,378	< 0,0001
Résidus	231	79193,944	342,831		
Total	254	350264,722			

La probabilité associée au test de Fisher est inférieure à 0,0001. Il y a donc moins de 0,1% de chance de se tromper en disant que les variables explicatives utilisées sont pertinentes pour expliquer la variance de pourcentages d'*An. gambiae* forme moléculaire S.

Les tests précédents permettent de vérifier que les variables prises dans leur globalité apportent une quantité d'informations significatives au modèle. Si cela est avéré au niveau global, en est-il de même vrai pour chacune d'elles ? L'évaluation de la significativité de l'apport de chaque variable au modèle est effectuée avec le test de Fisher type I SS et type III SS et les résultats sont contenus dans les

Tableau 21 et Tableau 22

Tableau 21 : Analyse du modèle (Type I SS) : test de Fisher sur les variables dans l'ordre de sélection

Source	ddl	Somme des carré	Carré moyen	F de Fisher	Probabilité > F
ETP	1	108868,911	108868,911	317,559	< 0,0001
VENT	1	7289,890	7289,890	21,264	< 0,0001
TOPO	1	16962,617	16962,617	49,478	< 0,0001
TEMPER	1	49930,813	49930,813	145,643	< 0,0001
PRESSION	1	58172,652	58172,652	169,683	< 0,0001
POP	1	2697,595	2697,595	7,869	0,005
PLUIE	1	334,621	334,621	0,976	0,324
LOCALITE	1	3414,847	3414,847	9,961	0,002
SOLEIL	1	8582,472	8582,472	25,034	< 0,0001
HYDRO	1	236,964	236,964	0,691	0,407
NDVI	1	10486,994	10486,994	30,589	< 0,0001
OCTER_NOMA	7	803,819	114,831	0,335	0,937
SOL	5	3288,584	657,717	1,918	0,092

Tableau 22 : Analyse du modèle (Type III SS) : test de Fisher sur les variables quel que soit l'ordre de sélection

Source	ddl	Somme des carré	Carré moyen	F de Fisher	Probabilité > F
ETP	1	3377,367	3377,367	9,851	0,002
VENT	1	624,765	624,765	1,822	0,178
TOPO	1	3695,415	3695,415	10,779	0,001
TEMPER	1	182,119	182,119	0,531	0,467
PRESSION	1	2041,856	2041,856	5,956	0,015
POP	1	14,189	14,189	0,041	0,839
PLUIE	1	893,095	893,095	2,605	0,108
LOCALITE	1	1512,114	1512,114	4,411	0,037
SOLEIL	1	3252,624	3252,624	9,488	0,002
HYDRO	1	7,962	7,962	0,023	0,879
NDVI	1	11368,180	11368,180	33,160	< 0,0001
OCTER_NOMA	7	1254,919	179,274	0,523	0,817
SOL	5	3288,584	657,717	1,918	0,092

D'après les résultats de ces tests, les variables qualitatives, (occupation des terres et sols) n'apportent pas une quantité d'information significative au modèle. Pour le premier test, il y a respectivement 93 % et 9,2 % de chance de se tromper en disant que l'occupation des terres et les sols apportent une information

significative au modèle. Pour le second, il y a respectivement 81 % et 9 % de chance de se tromper en disant que ces variables qualitatives apportent une quantité d'informations significative au modèle. Puisque ces variables qualitatives qui ont justifié l'utilisation de l'ANCOVA, n'apportent pas une quantité d'informations significatives au modèle, elles sont retirées du processus de modélisation.

Une régression multiple est utilisée en remplacement de l'ANCOVA pour la modélisation de la distribution des *An. gambiae* forme moléculaire S, puisqu'il s'agit d'une « ANCOVA sans variable qualitative ».

VI.1.2.2 Résultats et interprétation de l'ANCOVA pour *An. gambiae* forme moléculaire M

Pour *An. gambiae* forme moléculaire M, les coefficients d'ajustement sont consignés dans le Tableau 23 ci-dessous :

Tableau 23 : Coefficients d'ajustement du modèle

R (coefficient de corrélation)	0,835
R ² (coefficient de détermination)	0,697
R ² aj. (coefficient de détermination ajusté)	0,666
SCR	62372,707

Au regard de ces coefficients, les variables retenues expliquent 70% de la variabilité observée. Les 30% restant sont liés à d'autres facteurs ou à des erreurs de mesure. La significativité des informations fournies par les variables dans leur ensemble est également bonne comme le confirment les résultats du test de Fisher consignés dans le Tableau 24 ci-dessous.

Tableau 24 : test de Fisher sur la significativité de la quantité de données par rapport au modèle

Source	ddl	Somme des carrés	Carré moyen	F de Fisher	Probabilité > F
Modèle	23	143188,133	6225,571	23,057	< 0,0001
Résidus	231	62372,707	270,012		
Total	254	205560,840			

Ainsi, y a-t-il plus de 99% de chance que les variables apportent une information de quantité significative au modèle, car la probabilité associée au test de Fisher est de 0,0001.

La pertinence des variables prises individuellement par rapport au modèle a fait l'objet d'un test de Fisher de Type I SS et d'un test de Fisher de Type III SS. Les résultats de ces tests sont consignés dans les Tableau 25 et Tableau 26.

Tableau 25 : Analyse du modèle (Type I SS) : test de Fisher sur les variables dans l'ordre de sélection

Source	ddl	Somme des carrés	Carré moyen	F de Fisher	Probabilité > F
--------	-----	------------------	-------------	-------------	-----------------

ETP	1	87770,956	87770,956	325,064	< 0,0001
VENT	1	712,121	712,121	2,637	0,106
TOPO	1	2174,481	2174,481	8,053	0,005
TEMPER	1	3517,021	3517,021	13,025	0,000
PRESSION	1	25029,875	25029,875	92,699	< 0,0001
POP	1	35,273	35,273	0,131	0,718
PLUIE	1	1365,148	1365,148	5,056	0,025
LOCALITE	1	652,732	652,732	2,417	0,121
SOLEIL	1	4146,747	4146,747	15,358	0,000
HYDRO	1	582,817	582,817	2,158	0,143
NDVI	1	13339,163	13339,163	49,402	< 0,0001
OCTER_NOMA	7	2506,583	358,083	1,326	0,239
SOL	5	1355,217	271,043	1,004	0,416

Tableau 26 : Analyse du modèle (Type III SS) : test de Fisher sur les variables quel que soit l'ordre de sélection

Source	ddl	Somme des carrés	Carré moyen	F de Fisher	Probabilité > F
ETP	1	2639,968	2639,968	9,777	0,002
VENT	1	826,599	826,599	3,061	0,082
TOPO	1	6677,828	6677,828	24,732	< 0,0001
TEMPER	1	513,079	513,079	1,900	0,169
PRESSION	1	3408,675	3408,675	12,624	0,000
POP	1	301,586	301,586	1,117	0,292
PLUIE	1	4,016	4,016	0,015	0,903
LOCALITE	1	1596,891	1596,891	5,914	0,016
SOLEIL	1	5345,737	5345,737	19,798	< 0,0001
HYDRO	1	1048,351	1048,351	3,883	0,050
NDVI	1	12245,092	12245,092	45,350	< 0,0001
OCTER_NOMA	7	2374,138	339,163	1,256	0,273
SOL	5	1355,217	271,043	1,004	0,416

La contribution des variables en rouge dans les deux tableaux ci-avant n'est pas significative. L'apport de ces variables n'est donc pas significatif dans la qualité du modèle. Les deux variables qualitatives (l'occupation des terres et sols) font partie des variables à contribution faible, peu significative au modèle. Puisque ce sont elles qui ont justifié l'utilisation de l'ANCOVA comme technique de modélisation, leurs faibles contributions autorisent à utiliser seulement les variables quantitatives. Ainsi, plutôt que l'ANCOVA, une régression multiple est utilisée pour la modélisation d'*An. gambiae* forme moléculaire M.

VI.2 MODELISATION PAR REGRESSION MULTIPLE

VI.2.1 CONFIGURATION DE LA REGRESSION ET DESCRIPTION DES DONNEES

L'ANCOVA a mis en exergue la contribution non significative des variables qualitatives à la modélisation des relations entre les formes moléculaires M et S d'*An. gambiae* s.s.. L'ANCOVA s'en trouve donc réduite à peu d'intérêt, d'où l'utilisation de la régression⁶⁰.

La régression réalisée est de type linéaire à trois variables minimum et quatre variables maximum. Elle est exécutée une première fois avec le pourcentage d'*An. gambiae* forme moléculaire S comme variable dépendante et une seconde fois avec *An. gambiae* forme moléculaire M comme variable dépendante. Sur les 300 individus que constitue l'échantillon, 255 soit 85% ont été tirés de façon aléatoire pour la modélisation. Les 45 autres individus, soit 15% de l'échantillon, ont été réservés pour tester la qualité du modèle avec des données empiriques.

Les variables à expliquer, les variables dépendantes, présentent les caractéristiques suivantes (Tableau 27).

Tableau 27 : Statistiques descriptives sur les variables à expliquer

Variable	Nbr. de valeurs tot	Nbr. de valeurs utilisées	Nbr. de valeurs ignorées	Somme des poids	Moyenne	Ecart-type
POURCENT_S	255	255	0	255	42,066	37,135
POURCENT_M	255	255	0	255	27,755	28,448

Les variables explicatives retenues sont toutes les variables quantitatives de l'ANCOVA. Le Tableau 28 ci-dessous contient la liste des variables explicatives et leurs résumés statistiques.

Tableau 28 : Résumé pour les variables explicatives

Variable	Signifiant de la variable	Moyenne	Ecart-type
ETP	Evapo-transpiration	169	13
VENT	Vitesse du vent	1,84	0,22
TOPO	Topographie / Altitude	320	54
TEMPER	Température moyenne	28	0,41
PRESSION	Pression	18,25	1,26
POP	Densité	79	72
PLUIE	Pluviométrie	776	183

⁶⁰ La régression linéaire semble avoir les meilleurs résultats en terme de détermination de la relation entre l'environnement et la distribution moléculaire M et S d'*An. gambiae* ss.

SOLEIL	Insolation	0,68	0,03
LOCALITE	Distance aux localités	1543	2083
HYDRO	Distance aux cours	496	
NDVI	Végétation/intensité de l'activité chlorophylle	134	22

Chaque régression utilise un minimum de deux variables. Le maximum de variables utilisées par régression est de 3 pour *An. gambiae* forme moléculaire S et de 4 pour *An. gambiae* forme moléculaire M. Le seuil de confiance est de 95% et le critère d'évaluation de la régression est le coefficient de détermination (R^2). Les tests de Fisher de type I SS et type III SS sont associés au traitement pour l'évaluation de la significativité de la quantité de l'information apportée par les différentes variables à la modélisation.

VI.2.2 RESULTATS DE LA REGRESSION

La présentation des résultats de la régression est faite également de façon séparée pour chaque forme moléculaire d'*An.gambiae* s.s. afin de conserver ce caractère individualisé de la modélisation. Chaque forme moléculaire fait donc l'objet d'un sous paragraphe qui présente les paramètres de la modélisation et une analyse de la qualité des variables explicatives à la qualité des modèles.

VI.2.2.1 Modèle de distribution d'*An. gambiae* forme moléculaire S

Les résultats de la régression sont les suivants pour *An. gambiae* forme moléculaire S. D'après le critère retenu pour l'évaluation de la régression, le R^2 , les deux meilleurs modèles pour *An. gambiae* forme moléculaire S sont consignés dans le Tableau 29.

Tableau 29 : Valeur du coefficient de détermination pour les 2 meilleurs modèles

Modèle	Nombre de variables dans le modèle	Nom des Variable	R^2
1	2	SOLEIL, NDVI	0,720
2	3	ETP, SOLEIL, NDVI	0,746

D'après ce Tableau 29, les trois meilleures variables qui permettent d'expliquer la variance observée sont : l'insolation, la végétation et l'évapotranspiration potentielle, en dépit de la colinéarité⁶¹. Le Tableau 30, ci-dessous, donne les coefficients d'ajustement du meilleur modèle, le modèle 2.

Tableau 30 : Coefficients d'ajustement du modèle 2

R (coefficient de corrélation)	0,864
R ² (coefficient de détermination)	0,746
R ² aj. (coefficient de détermination ajusté)	0,743
SCR	89081,950

Le coefficient de détermination atteint jusqu'à 0,746. Cela signifie que 74,6% de la variance sont expliqués par la combinaison linéaire des trois variables que sont l'insolation, l'ETP et la végétation. Le test de Fisher sur l'analyse de la significativité de la quantité d'information apportée par les variables prises dans leur globalité confirme que les variables fournissent suffisamment d'informations à l'élaboration du modèle (Tableau 31).

Tableau 31 : Le test de Fisher sur l'analyse de la significativité de la quantité d'information apportée par les variables à la régression pour *Anopheles gambiae* forme moléculaire S

Source	ddl	Somme des carrés	Carré moyen	F de Fisher	Probabilité > F
Modèle	3	261182,772	87060,924	245,305	< 0,0001
Résidus	251	89081,950	354,908		
Total	254	350264,722			

La probabilité associée au test F de Fisher est de 0,0001 soit moins de 1% de chance de se tromper en considérant que la quantité d'information apportée par l'ETP, l'insolation et la végétation est significative pour le modèle.

Les Tableau 32 et Tableau 33 contiennent les résultats de test de Fisher de type I SS et de type III SS.

Tableau 32 : Test de Fisher de type I SS

Source	ddl	Somme des carrés	Carré moyen	F de Fisher	Pr > F
ETP	1	108868,911	108868,911	306,752	< 0,0001
SOLEIL	1	138071,254	138071,254	389,034	< 0,0001

⁶¹ La régression repose sur des postulats dont l'absence colinéarité entre les variables explicatives. Toutefois, ce postulat est violé mais la qualité du modèle et sa conformité au terrain ont conduit à conserver dans la fonction des variables liées sans altérer la pertinence du modèle. Cette situation est bien connue dans la littérature scientifique avec par exemple Kennedy, P., 1992, A Guide to Econometrics, Blackwell Publishers, 410 p. et Conlisk, J., 1971, When collinearity is desirable: Western Economic Journal, v. 9, p. 393-407.

NDVI	1	14242,607	14242,607	40,130	< 0,0001
-------------	---	-----------	-----------	--------	----------

Tableau 33 : Analyse du modèle (Type III SS)

Source	ddl	Somme des carrés	Carré moyen	F de Fisher	Pr > F
ETP	1	8839,892	8839,892	24,908	< 0,0001
SOLEIL	1	34816,671	34816,671	98,101	< 0,0001
NDVI	1	14242,607	14242,607	40,130	< 0,0001

Les probabilités associées à ces tests sont toutes inférieures à 0,0001. Ce qui permet de dire que les trois variables, dans un ordre ou non, apportent chacune une quantité d'informations significatives au modèle.

Puisque les variables sont pertinentes par rapport au modèle du fait de la significativité de la quantité d'informations apportées, les paramètres du modèle peuvent être présentés (Tableau 34).

Tableau 34 : Paramètres du modèle de distribution d'*An. gambiae* forme moléculaire S

Paramètre	Valeur	Ecart-type	t de Student	Pr > t	Borne inférieure 95 %	Borne supérieure 95 %
Constante	382,484	67,080	5,702	< 0,0001	250,372	514,596
ETP	0,713	0,143	4,991	< 0,0001	0,431	0,994
SOLEIL	-813,672	82,151	-9,905	< 0,0001	-975,465	-651,878
NDVI	0,701	0,111	6,335	< 0,0001	0,483	0,919

Le modèle s'écrit : $S = 382,484 + 0,713 \times ETP - 813,672 \times SOLEIL + 0,701 \times NDVI$

Le pourcentage d'*An. gambiae* forme moléculaire S dans l'intervalle de variation de l'ETP, de l'insolation et de l'activité chlorophyllienne, peut s'obtenir en additionnant la valeur de l'activité chlorophyllienne multipliée par 0,701 à l'insolation, multipliée par -813,672 et à l'évapotranspiration potentielle, multipliée par 0,713 à une constante égale à 382,484. L'ensemble des paramètres environnementaux intervenant dans le modèle sont ceux associés à l'axe factoriel humidité – aridité. Le sud-ouest du Burkina Faso se présente donc comme la zone préférentielle pour *An. gambiae* forme moléculaire S.

VI.2.2.2 Le modèle de distribution d'*An. gambiae* forme moléculaire M

De la régression réalisée entre les pourcentages de la forme moléculaire M d'*An. gambiae* et les données environnementales, les trois modèles se dégagent (le Tableau 35).

Tableau 35 : Valeur du coefficient de détermination pour les 2 meilleurs modèles

Modèle	Nbr. dans le modèle	Variable(s)	R ²
1	2	ETP NDVI	0,594
2	3	PRESSION SOLEIL NDVI	0,617
3	4	TOPO PRESSION SOLEIL NDVI	0,641

Le meilleur modèle est le modèle 3 avec un coefficient de détermination de 0,641, car il a le coefficient de détermination le plus élevé. Les coefficients associés à ce modèle sont consignés dans le

Tableau 36 :

Tableau 36 : les coefficients d'ajustement pour le modèle 3.

R (coefficient de corrélation)	0,800
R ² (coefficient de détermination)	0,641
R ² aj. (coefficient de détermination ajusté)	0,635
SCR	73851,489

Il s'avère, à la lecture de ce tableau, que 64,1% de la variance observée dans les pourcentages d'*An. gambiae* forme moléculaire S s'expliquent par la combinaison linéaire de quatre variables que sont la topographie (altitude), la pression, la végétation et l'insolation. Ces variables apportent une quantité d'informations statistiquement significatives au modèle. En témoigne le test de Fisher réalisé pour évaluer la qualité de l'information apportée par les variables. Les résultats sont consignés dans le Tableau 37.

Tableau 37 : Evaluation de la valeur de l'information apportée par les variables (H0 = Y=Moy(Y))

Source	ddl	Somme des carrés	Carré moyen	F de Fisher	Pr > F
Modèle	4	131709,350	32927,338	111,465	< 0,0001
Résidus	250	73851,489	295,406		
Total	254	205560,840			

Il ressort, à la lecture de ce tableau, que la probabilité associée au test de Fisher est inférieure à 1%. Les variables apportent une quantité d'informations significatives au modèle à plus de 99%.

Cette appréciation générale de la contribution des variables dans leur globalité pourrait cacher la faible contribution d'une des quatre variables intervenant dans le modèle. Les tests de Fisher de Type I SS et de Type III SS sont effectués pour évaluer la contribution de chaque variable, d'abord, dans l'ordre puis de façon quelconque. Les résultats de ces tests sont reportés dans les Tableau 38 et Tableau 39.

Tableau 38 : Analyse du modèle (Type I SS)

Source	ddl	Somme des carrés	Carré moyen	F de Fisher	Pr > F
TOPO	1	13648,055	13648,055	46,201	< 0,0001
PRESSION	1	94782,658	94782,658	320,856	< 0,0001
SOLEIL	1	5452,629	5452,629	18,458	< 0,0001
NDVI	1	17826,008	17826,008	60,344	< 0,0001

Tableau 39 : Analyse du modèle (Type III SS)

Source	ddl	Somme des carrés	Carré moyen	F de Fisher	Pr > F
TOPO	1	4907,930	4907,930	16,614	< 0,0001
PRESSION	1	11375,484	11375,484	38,508	< 0,0001
SOLEIL	1	10053,822	10053,822	34,034	< 0,0001
NDVI	1	17826,008	17826,008	60,344	< 0,0001

Ils nous permettent de conclure, au regard des probabilités associées au test de Fisher, que les variables, prises individuellement et quelque soit l'ordre, apportent une quantité d'information significative au modèle. En effet, le test de probabilité associée à chacune des variables prises dans l'ordre ou non est inférieur à 1%. Nous affirmons à plus de 99%, que les variables altitude, pression, insolation et végétation apportent une quantité d'informations significatives à l'élaboration du modèle.

Fort de cette significativité de l'information utilisée et de la contribution de chacune des variables, les paramètres du modèle sont les suivants (Tableau 40).

Tableau 40 : Paramètres du modèle

Paramètre	Valeur	Ecart type	t de Student	Pr > t	Borne inférieure 95 %	Borne supérieure 95 %
Constante	958,241	128,165	7,477	< 0,0001	705,820	1210,662
TOPO	-0,098	0,024	-4,076	< 0,0001	-0,146	-0,051
PRESSION	-17,389	2,802	-6,205	< 0,0001	-22,908	-11,870

SOLEIL	-687,661	117,874	-5,834	< 0,0001	-919,814	-455,508
NDVI	-0,832	0,107	-7,768	< 0,0001	-1,042	-0,621

L'équation du modèle s'écrit de la façon suivante :

$$M = 958,241 - 0,098 \times \text{TOPO} - 17,389 \times \text{PRESSION} - 687,661 \times \text{SOLEIL} - 0,832 \times \text{NDVI}$$

Le pourcentage d'*An. gambiae* forme moléculaire M peut être déduit en additionnant la constante 958,241 à l'altitude, à la pression, à l'insolation et à la végétation respectivement multipliées par -0,098 ; -17,389 ; -687,661 et -0,832 (dans l'intervalle de variation défini dans le tableau consacré aux paramètres du modèle, au seuil de confiance de 95%). Nous constatons aussi que les paramètres utilisés sont ceux associés à l'axe factoriel 1, l'axe humidité – aridité. La topographie (altitude) est associée à l'axe factoriel 1 mais aussi à l'axe factoriel 4 où elle a la meilleure contribution.

L'augmentation du nombre de variable dans le modèle doublé d'une diminution du coefficient de détermination de plus 10 points par rapport au modèle d'*An. gambiae* forme moléculaire S, montre qu'il a été plus difficile de modéliser *An. gambiae* forme moléculaire S que *An. gambiae* forme moléculaire M. La forme moléculaire M est plus hétérogène que la forme S (Manguin et al., 1999; Touré et al., 1991).

VI.3 EVALUATION DES MODELES

L'analyse de la contribution globale puis individuelle des variables et celle des coefficients permettent d'apprécier la qualité de l'information par rapport au modèle. L'évaluation du modèle aide à apprécier la compatibilité du modèle par rapport à la réalité du terrain. Cette évaluation va se faire de deux manières : l'analyse des résidus et la comparaison des résultats du modèle théorique avec des données empiriques.

VI.3.1 ANALYSE DES RESIDUS

L'évaluation par analyse de résidus est une évaluation statistique. La validité d'un modèle linéaire suppose au plan statistique que les résidus suivent une loi normale. Le seuil de confiance retenu est de 95%. A ce seuil, les individus ayant des valeurs de résidus centrés réduits situées en dehors de l'intervalle]1,96 et -1,96[doivent représenter au plus 5% de l'ensemble de l'échantillon

VI.3.1.1 Analyse des résidus pour *An. gambiae* forme moléculaire S

Pour *An. gambiae* forme moléculaire S, les valeurs de résidus centrés réduits sont représentées sur la Figure 55.

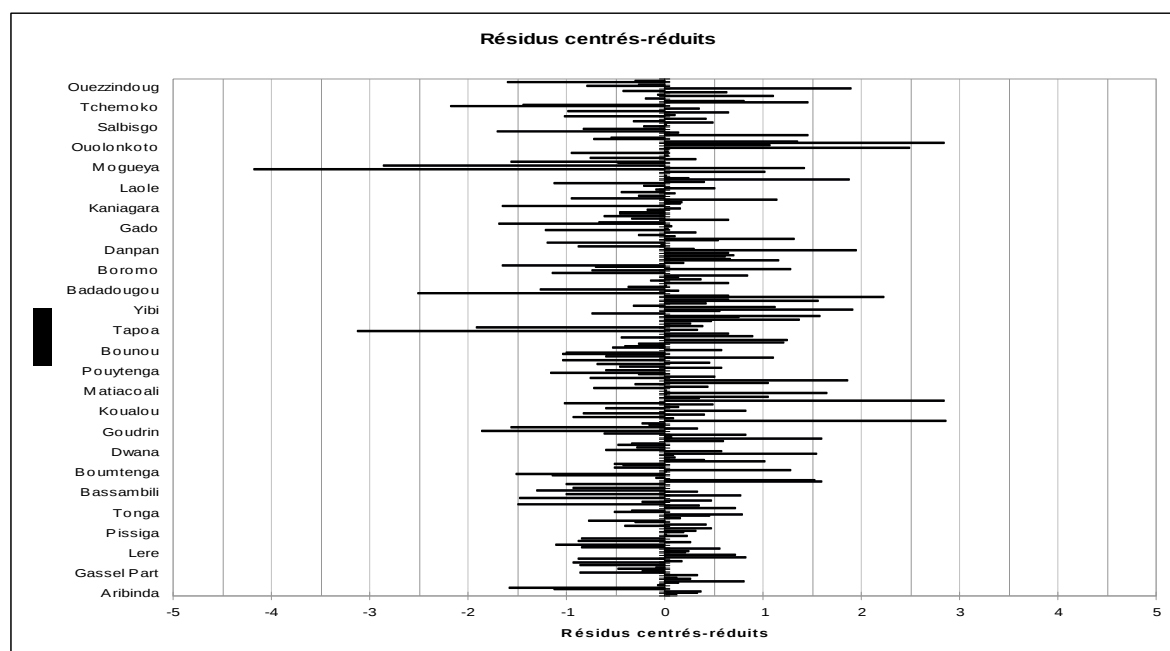


Figure 55 : Histogramme des résidus centrés réduits pour *An. gambiae* forme moléculaire S

Le Tableau 41 contient la liste des localités qui ont des pourcentages d'*An. gambiae* forme moléculaire S anormaux.

Tableau 41 : Liste des localités ayant des résidus anormaux pour *An. gambiae* forme moléculaire S

Individus	Poids	% S empirique	% S (Modèle)	Résidus	Résidus centrés-réduits	Interv. Inf. Moy.	Interv. Sup. Moy.	Interv. Inf. Ind.	Interv. Sup. Ind.
Koassa	1	88,235	34,459	53,776	2,854	32,001	36,918	-2,725	71,644
Lantaogo	1	88,889	35,321	53,568	2,843	32,840	37,801	-1,865	72,506
Tapoa	1	19,231	77,968	-58,737	-3,118	70,792	85,143	40,177	115,758
Zinekwe	1	84,211	42,301	41,910	2,225	38,870	45,731	5,040	79,562
Alana	1	44,444	91,558	-47,114	-2,501	84,254	98,862	53,743	129,373
Mogueya	1	4,348	83,045	-78,697	-4,177	79,093	86,997	45,732	120,358
Monkui	1	0,000	53,876	-53,876	-2,860	50,905	56,846	16,654	91,097
Rabyiri	1	77,778	30,941	46,837	2,486	27,796	34,085	-6,295	68,176
Tomo	1	92,857	39,354	53,503	2,840	35,508	43,200	2,052	76,655
Timba	1	54,167	95,185	-41,018	-2,177	87,977	102,393	57,389	132,981

Sur les 255 individus de l'échantillon, 10 ont des résidus centrés réduits situés en dehors de l'intervalle $]-1,96 ; 1,96[$, soit 4%. Puisque 4% est inférieure à 5%, le modèle est donc une bonne approximation de la distribution et reste valable au seuil de confiance de 95%.

Ce premier niveau de validation du modèle donne la possibilité de production d'une carte théorique de la distribution de la forme moléculaire S d'*An. gambiae* s.s. au Burkina Faso (Figure 56). Nous observons sur cette carte des zones très homogènes de pourcentage de la forme moléculaire S d'*An. gambiae* s.s.. En dehors de quelques îlots très restreints, surtout dans la partie nord du pays, les pourcentages prédits sont disposés en bandes, avec une augmentation quasi régulière du nord vers le sud-ouest. Cette structure rubanée est très semblable à celle que présentent certains éléments du milieu physique comme les températures, les précipitations, la végétation (Figure 6, Figure 8, Figure 9, Figure 10, Figure 11, Figure 13 et Figure 14).

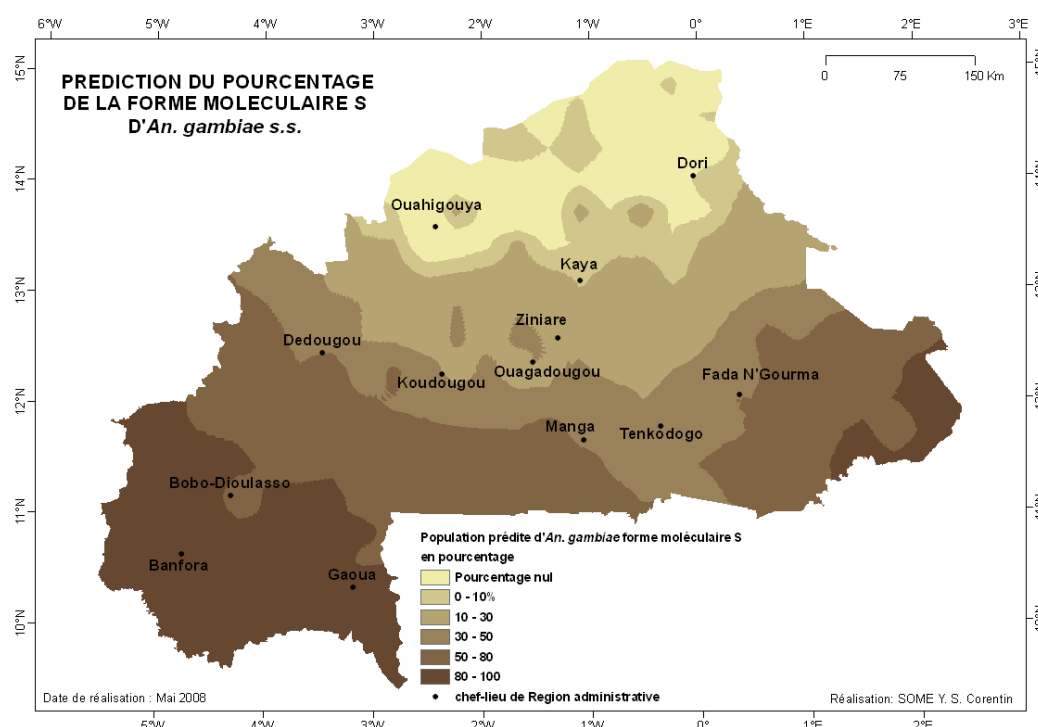


Figure 56 : Prédiction de la distribution spatiale des abondances relatives d'*An. gambiae* forme moléculaire S

Nous pouvons donc dire que l'environnement présente une qualité décroissante pour *An. gambiae* forme moléculaire S au fur à mesure que l'on va du nord vers le sud-ouest. Le nord n'est donc pas favorable à la prolifération de la forme moléculaire S. Toutefois, une mosaïque de peuplements isolés se trouve dans certaines parties du nord (Figure 57).

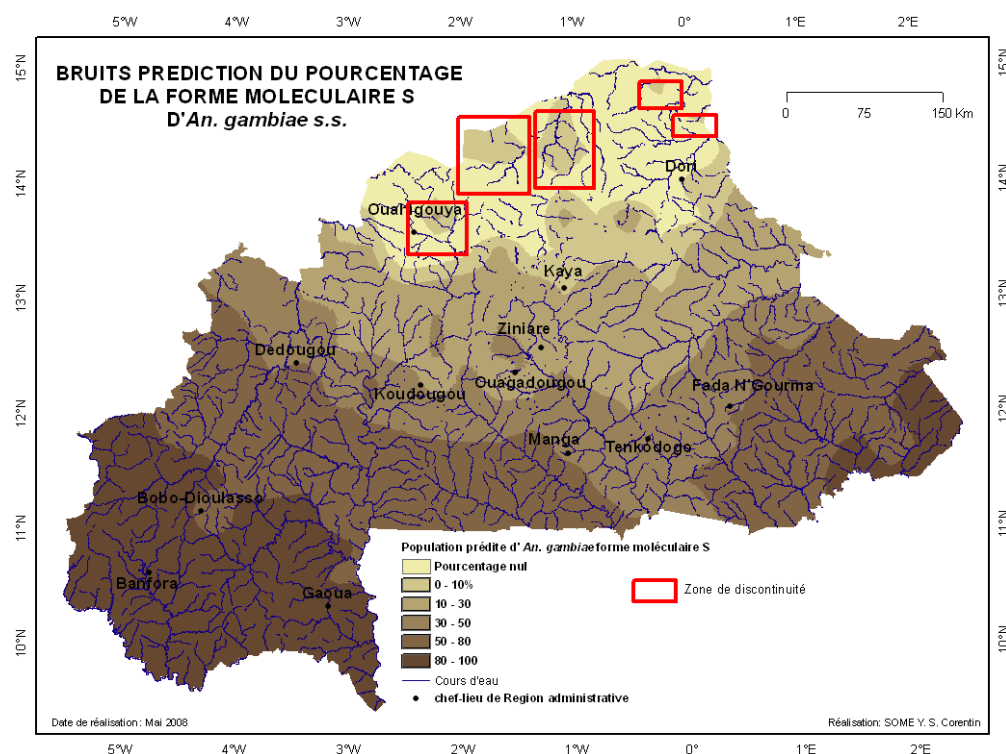


Figure 57 : les zones de discontinuité dans la répartition de la forme moléculaire S et hydrographie.
Ces discontinuités observées correspondent à l'existence de bas-fonds importants. L'humidité est ainsi confirmée comme le facteur déterminant de la distribution d'*An. gambiae* forme moléculaire S. Cette même continuité sera-t-elle observable pour *An. gambiae* forme moléculaire M, si le modèle est validé au plan statistique ?

VI.3.1.2 Analyse des résidus pour *An. gambiae* forme moléculaire M

Pour *An. gambiae* forme moléculaire M, la Figure 58 présente le diagramme des résidus centrés réduits.

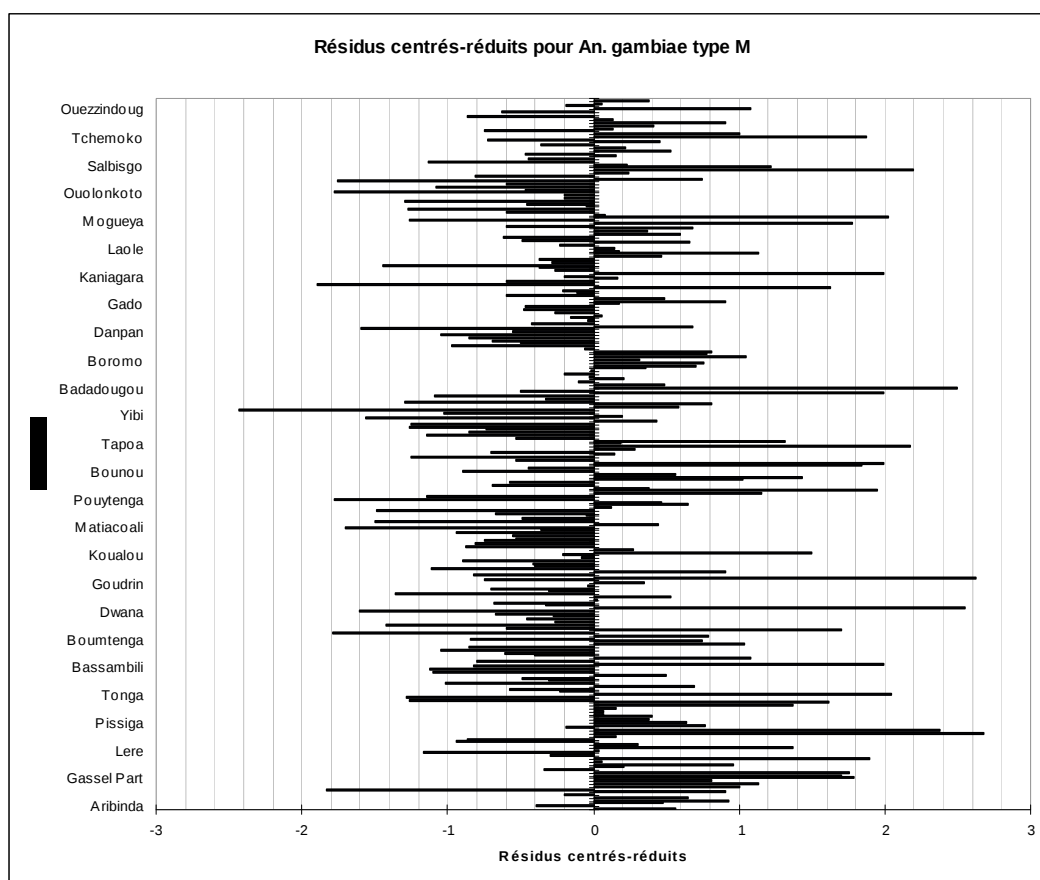


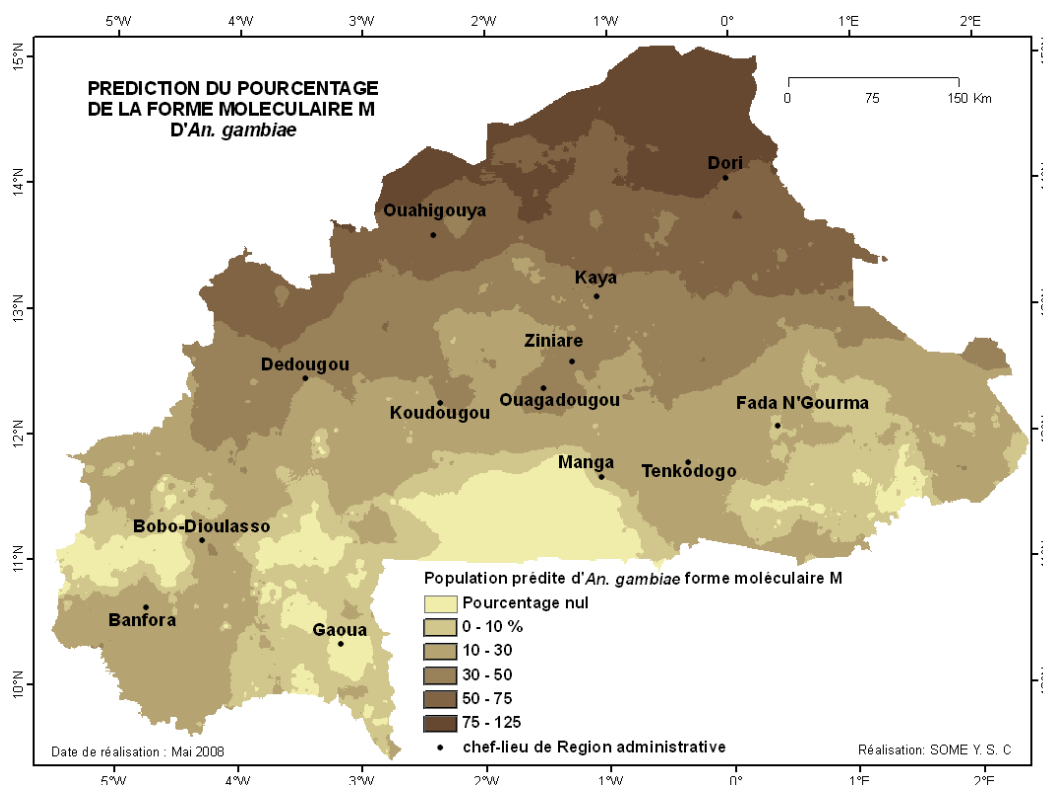
Figure 58 : Histogramme des résidus centrés réduits pour *An. gambiae* forme moléculaire M

Certains individus, sur ce graphique, ont des valeurs situées à l'extérieur de l'intervalle de $[-1,96 ; 1,96]$. Le tableau suivant contient les informations sur les individus de la distribution qui ont des valeurs anormalement élevées ou anormalement basses.

Tableau 42 : Liste des localités ayant des résidus anormaux pour *An. gambiae* forme moléculaire M

Individus	Poids	%_M	%_M (Modèle)	Résidus	Résidus centrés-réduits	Interv. Inf. Moy.	Interv. Sup. Moy.	Interv. Inf. Ind.	Interv. Sup. Ind.
Niounougou	1	90,909	44,819	46,090	2,682	41,698	47,941	10,825	78,813
Ouenbouloug	1	81,250	40,459	40,791	2,373	37,270	43,647	6,459	74,459
Tougouri	1	88,000	52,946	35,054	2,040	49,591	56,300	18,930	86,962
Belmin	1	60,870	26,633	34,237	1,992	22,693	30,572	-7,446	60,712
Fingla	1	63,158	19,286	43,872	2,553	15,219	23,354	-14,808	53,380
Goumaogo	1	44,444	-0,599	45,044	2,621	-5,824	4,626	-34,850	33,652
Tamassogo	1	88,000	53,826	34,174	1,988	50,278	57,374	19,790	87,862
Tapoa	1	53,846	16,480	37,366	2,174	9,674	23,287	-18,048	51,008
You	1	5,263	47,106	-41,843	-2,434	41,362	52,849	12,771	81,440
Alana	1	55,556	21,248	34,308	1,996	14,379	28,117	-13,293	55,788
Badara	1	88,235	45,372	42,863	2,494	40,279	50,465	11,140	79,603
Iboudie	1	66,667	32,399	34,268	1,994	29,362	35,435	-1,588	66,385
Monkui	1	71,429	36,709	34,720	2,020	32,116	41,302	2,548	70,870
Sandimisso	1	43,750	6,104	37,646	2,190	1,363	10,845	-28,077	40,285

Sur les 255 individus que compte la distribution, 14 ont des valeurs hors de l'intervalle de confiance soit 5%. Nous retenons donc que le modèle, dans l'intervalle de confiance retenu, est une bonne approximation de la distribution. Ce résultat en validant statistiquement le modèle permet de produire une carte de distribution de la forme moléculaire M d'*An. gambiae* s.s. (Figure 59).

**Figure 59 : Prédiction de la distribution spatiale des abondances relatives d'*An. gambiae* forme moléculaire M**

La distribution des abondances relatives de la forme moléculaire M d'*An. gambiae* s.s. présente également une forme rubanée mais moins lisse, comparativement à *An. gambiae* forme moléculaire S. Il y a, en effet, beaucoup plus de poches de peuplement dans la zone défavorable du sud-ouest. Au regard de la densité des poches observées, le sud-ouest n'est pas totalement défavorable à la forme moléculaire M d'*An. gambiae*. Une observation plus poussée permet de remarquer que ce sont les zones d'altitudes les plus élevées du pays qui ont des valeurs nulles. Ces poches correspondent également à des réserves (forêts classées). Elles présentent les conditions les moins favorables à l'écologie de la forme moléculaire M, puisque le modèle y prédit des abondances relatives nulles. Au fond, il s'agit des zones les plus humides. Cela suggère une plus grande plasticité écologique de la forme moléculaire M par rapport à la forme moléculaire S. Cela montre également la sensibilité de la forme moléculaire M par rapport à des conditions d'humidité très élevées. En effet, les faciès introduisent par la topographie (altitude) et la végétation (les zones de réserve ont les NDVI les plus élevés) évoquent les conditions d'humidité les plus fortes (Figure 60).

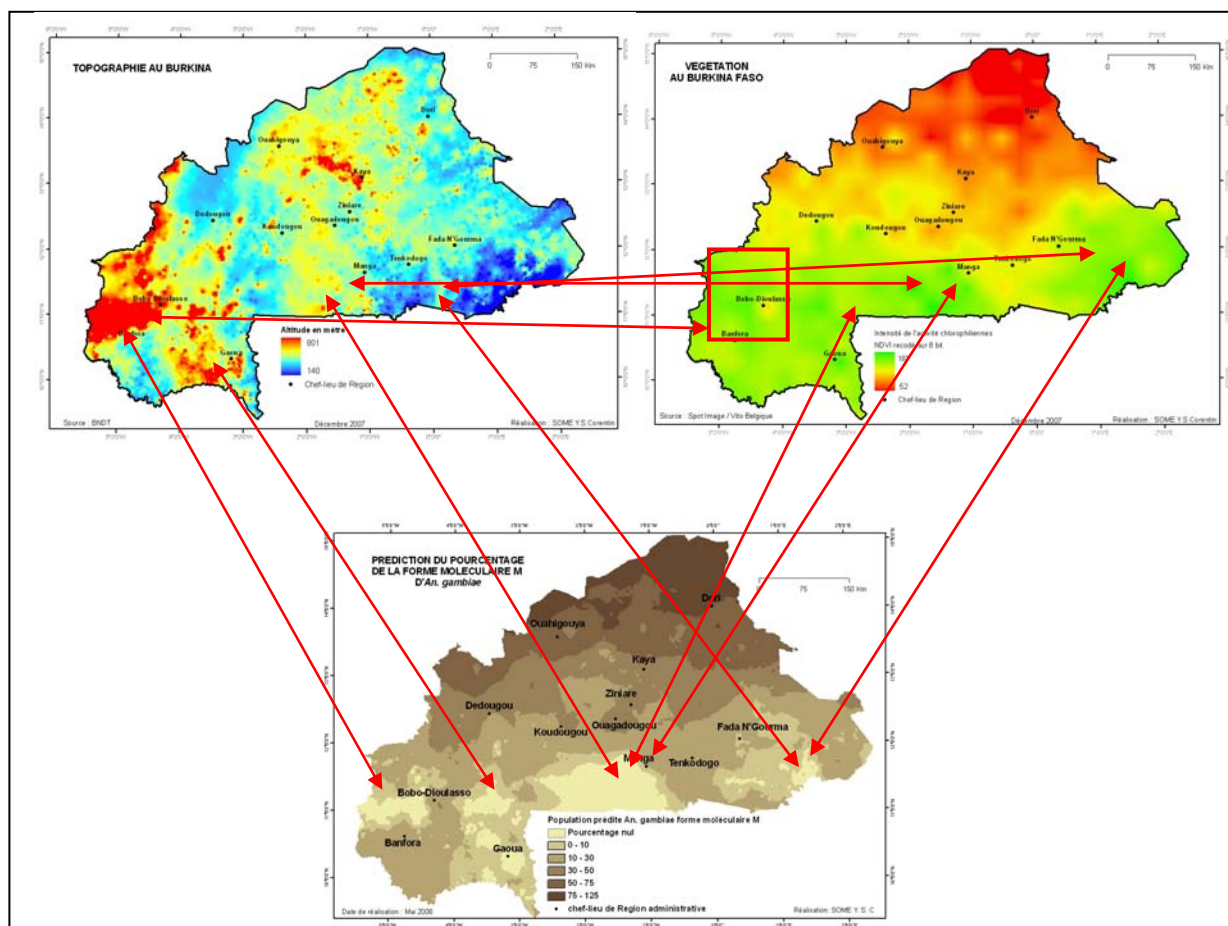


Figure 60 : Discontinuité de la répartition de la forme moléculaire M d'*An. gambiae* s.s., Altitude et NDVI

Cette situation permet de comprendre la prise en compte des altitudes dans la détermination de la distribution. Elle montre également que les facteurs écologiques invoquant l'humidité/aridité demeurent

les déterminants les plus importants dans la répartition des formes moléculaires S et M d'*An. gambiae* s.s.. Toutefois, la forme moléculaire M présente une plus grande capacité d'adaptation à l'humidité que la forme moléculaire S à l'aridité.

L'analyse des résidus a conduit à une validation statistique de fonctions qui modélisent la distribution spatiale des formes moléculaire S et M d'*An. gambiae* s.s.. L'analyse des cartes issues de ces modèles confirme les paramètres associés à l'aridité et à l'humidité comme les facteurs écologiques les plus déterminants dans la distribution spatiale de ces vecteurs du paludisme. Il apparaît également à l'analyse que la forme moléculaire M est plus résistante à la variabilité du milieu que la forme S.

Sur le plan statistique, ces modèles sont valables. Mais cette validité peut-elle être confirmée par les données empiriques, non utilisées dans la modélisation. Comment se comportent les modèles face aux données de terrain ?

VI.3.2 CONFRONTATION DES DONNEES DES MODELES AUX DONNEES EMPIRIQUES

Pour confronter les résultats aux modèles, 45 points ont été tirés au hasard sur les 300 points que comptait le transect. Ces 45 points, soit 15% de l'échantillon, ont été réservés pour confronter les modèles aux données empiriques.

VI.3.2.1 Confrontation du modèle d'*An. gambiae* forme moléculaire S à la vérité terrain

Pour *An. gambiae* forme moléculaire S (Figure 61), le coefficient de corrélation entre la distribution observée et la distribution prédite est de 0,81. Le coefficient de corrélation est positif et très élevé. Ce qui signifie que les valeurs évoluent dans le même sens que les valeurs observées.

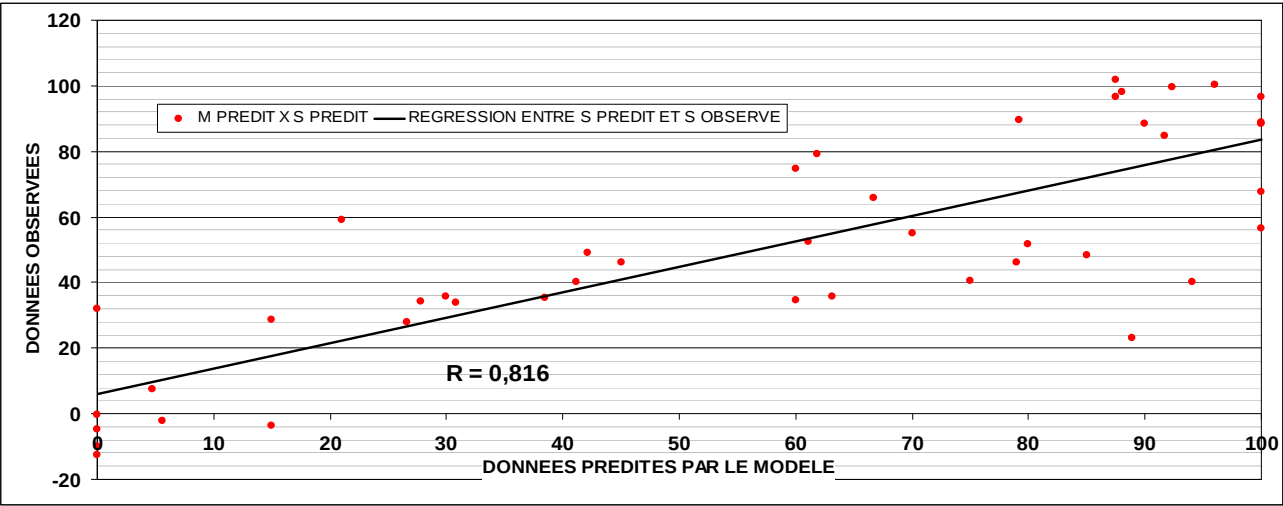


Figure 61 : Corrélation entre pourcentage de S prédit par le modèle et pourcentage S observé.

Après observation du coefficient de corrélation, le Test de Kolmogorov-Smirnov sur deux échantillons / Test bilatéral a été effectué pour vérifier si la distribution empirique était significativement différente de la distribution prédite par le modèle ou pas.

Les données d’entrée dans le test présentent les caractéristiques suivantes (Tableau 43).

Tableau 43 : Caractéristiques des données utilisées pour le test de Kolmogorov-Smirov

Variable	Observation	Obs. avec données manquantes	Obs. sans données manquantes	Minimum	Maximum	Moyenne	Ecart type
POURC S	45	0	45	0,000	100,000	57,784	35,324
S PREDIT	45	0	45	-12,573	101,954	50,874	33,473

Les résultats du test sont les suivants (Tableau 44 et la Figure 62).

Tableau 44 : Test de Kolmogorov-Smirnov pour la forme moléculaire S

D	0,244
P-value	0,115
alpha	0,05

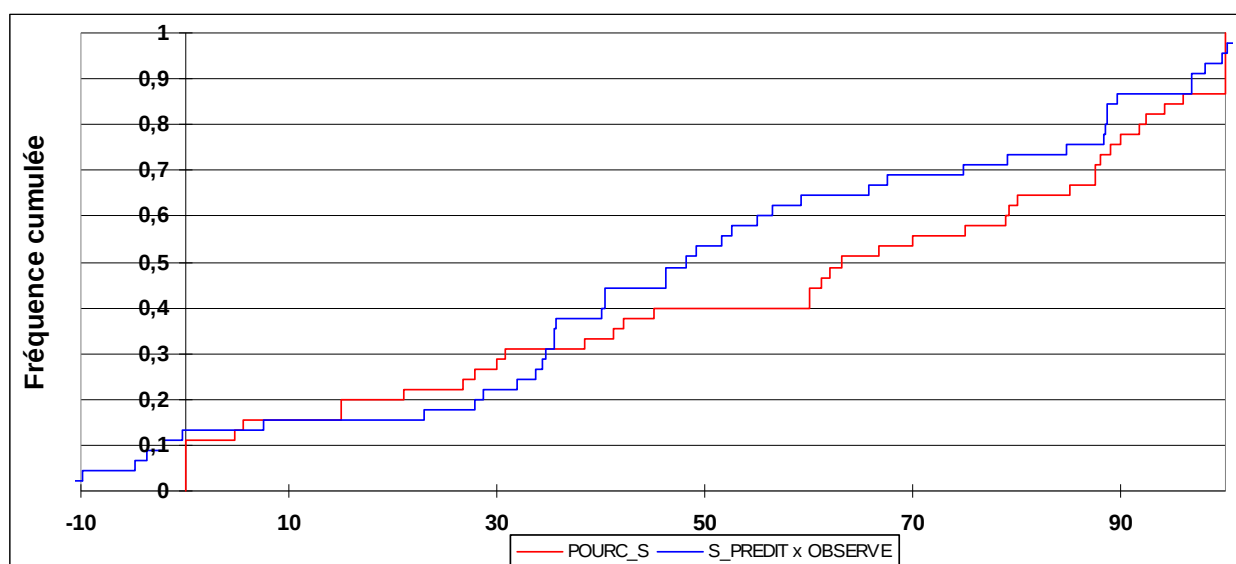


Figure 62 : Comparaison des pourcentages observés avec des pourcentages prédits de la forme moléculaire S

La différence D entre les deux échantillons est 0,244 sur une échelle de 0 à 1. La différence est proche de 0, le niveau de ressemblance maximum, le niveau d'identité parfaite. La p-Value est aussi supérieure au seuil de confiance. Les écarts entre la distribution calculée et la distribution observée ne sont donc pas significatifs. Ce dernier contrôle confirme la validité du modèle pour la forme moléculaire S d'*An. gambiae* s.s..

VI.3.2.2 Confrontation du modèle d'*An. gambiae* forme moléculaire M à la vérité terrain

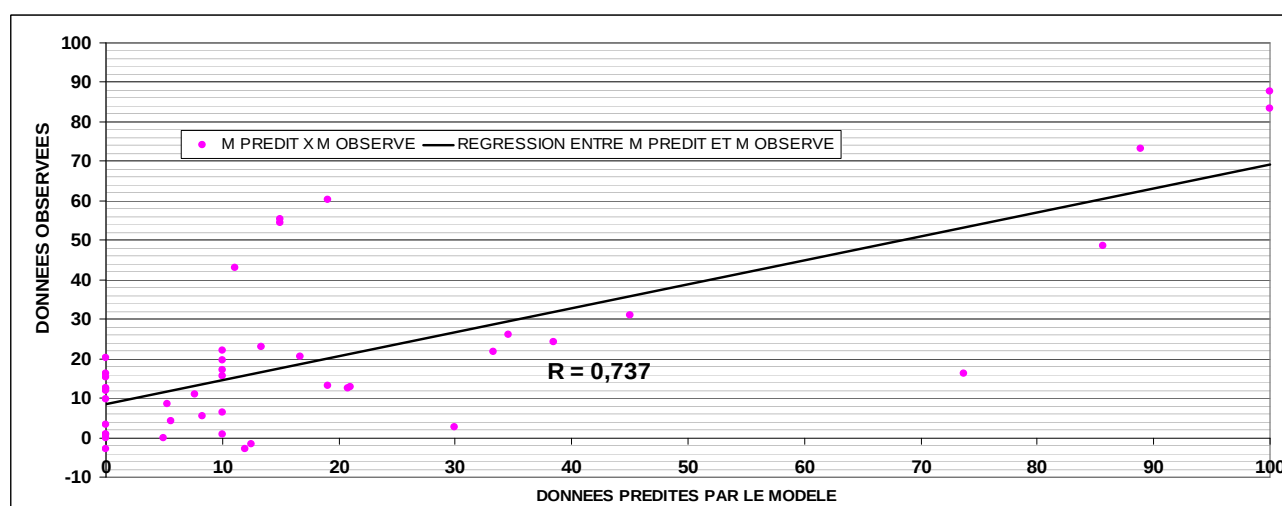


Figure 63 : Corrélation entre pourcentages de M prédits par le modèle et pourcentages M observés

Pour la forme moléculaire M (Figure 63), la corrélation entre les valeurs de pourcentages prédits par le modèle et celles observées sur le terrain est de 0,73.

Ce coefficient est également positif. Les données prédites sont très bien corrélées avec les données observées. Les modèles sont donc de bons prédicateurs des valeurs observées. Le Test de Kolmogorov-Smirnov sur deux échantillons / test bilatéral effectué entre les données calculées avec le modèle et les données empiriques est effectué. L'objectif de ce test est de vérifier s'il y a une différence significative entre elles. Les données utilisées pour ce test présentent les caractéristiques suivantes (Tableau 45).

Tableau 45 : Caractéristiques des données test de Kolmogorov-Smirnov pour la forme moléculaire M

Variable	Observations	Obs. avec données manquantes	Obs. sans données manquantes	Minimum	Maximum	Moyenne	Ecart-type
POURC_M	45	0	45	0	100	19	27
M_PREDIT x OBSERVE	45	0	45	-2,8	87,5	20,5	22,5

Le test de Kolmogorov-Smirnov donne les résultats suivants (Tableau 46 et Figure 64)

Tableau 46 : Test de Kolmogorov-Smirnov pour la forme moléculaire M

D	0,200
p-value	0,295
alpha	0,05

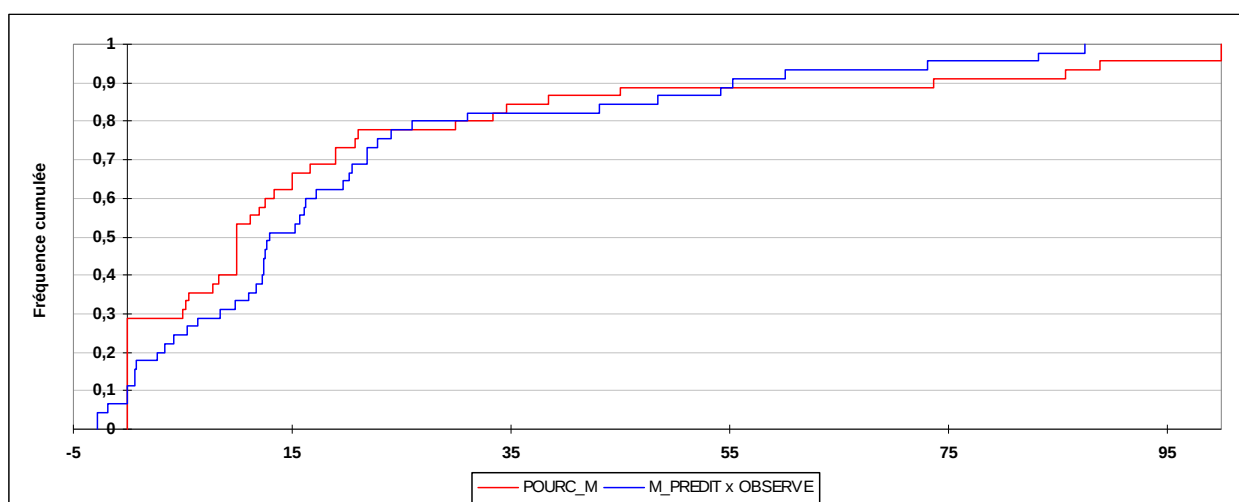


Figure 64 : Comparaison des pourcentages observés avec les fréquences prédites de la forme moléculaire M

La différence D sur une échelle de 0 à 1, entre les deux échantillons est 0,244. Cette différence est faible car plus proche de 0 (valeur au cas d'identité parfaite). La P-value est de 0,295 et est supérieure au seuil de confiance qui est de 0,05. La différence entre les données calculées et les données observées n'est pas significative. Ce test confirme que le modèle est valable pour l'évaluation de la distribution spatiale de la forme moléculaire M d'*An. gambiae* s.s..

Il est bon d'observer (Figure 65) que les valeurs prédites par le modèle de S et celles prédites par le modèle M ont un coefficient de corrélation de -0,84.

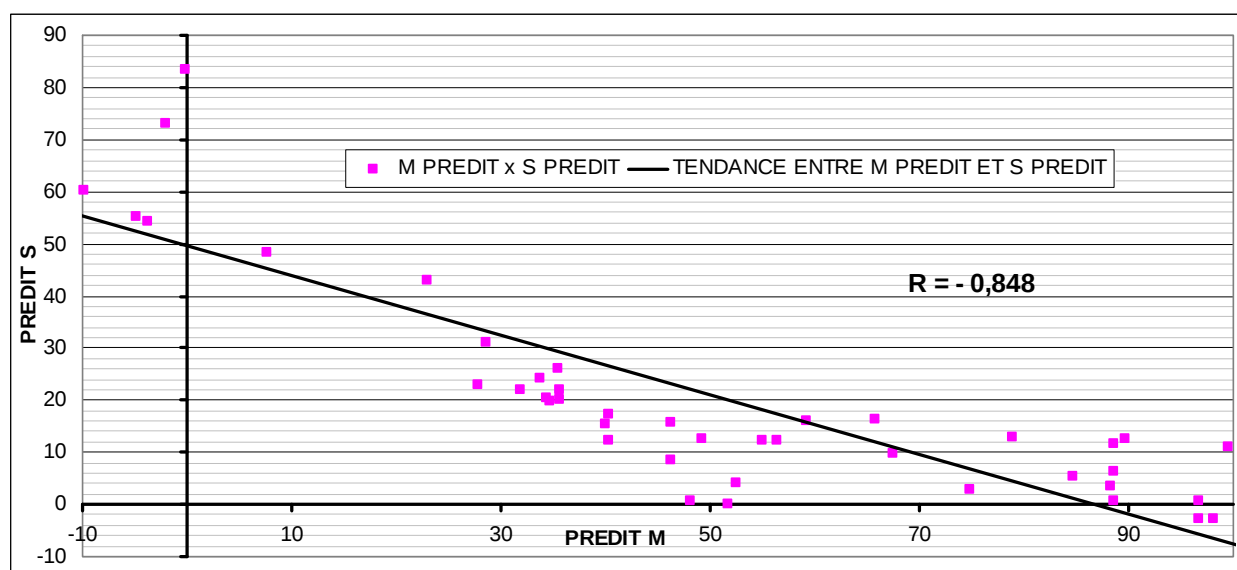


Figure 65 : Corrélation entre pourcentages de S prédits et pourcentage de M prédits

Cela signifie que le modèle produit pour la forme moléculaire S d'*An. gambiae* s.s. est pratiquement l'opposé de celui produit pour la forme moléculaire M d'*An. gambiae* s.s.. Lorsque le pourcentage d'*An. gambiae* forme moléculaire S augmente en un lieu, celui d'*An. gambiae* forme moléculaire M diminue. Nous en déduisons une opposition des préférences écologiques entre les formes moléculaires M et S d'*An. gambiae* s.s.. La forme moléculaire S préfère les milieux humides alors que la forme moléculaire M préfère les milieux plus arides.

Ces derniers résultats confirment l'hypothèse de départ selon laquelle il est possible de prédire la population des deux formes moléculaires M et S d'*An. gambiae* s.s., à partir d'une connaissance des facteurs environnementaux les plus déterminants dans leur distribution spatiale. Ces paramètres sont le NDVI, l'insolation, l'ETP, la pression et la topographie (altitude). Ils confirment également l'opposition entre les deux formes moléculaires d'*An. gambiae* s.s. comme nous l'a montré la forte corrélation négative entre les valeurs prédites par les deux modèles ainsi que la distribution spatiale opposée.

CONCLUSION DU CHAPITRE VI

Ce chapitre révèle l'expression des abondances relatives des formes moléculaires M et S en fonction de quatre et trois facteurs environnementaux. Il a cependant été plus difficile de modéliser la forme

moléculaire M par rapport à la forme moléculaire S. Il a fallu plus des paramètres pour modéliser la distribution spatiale de la forme moléculaire M que pour la forme moléculaire S. Malgré l'augmentation du nombre de variables pour la forme moléculaire M, le modèle de S a un coefficient de détermination plus élevé. Il est probable que la relative difficulté de modélisation de forme moléculaire M d'*An. gambiae* s.s. par rapport à la forme moléculaire S provienne de la forte variabilité qu'on observe en son sein ainsi que sa capacité à coloniser une grande variabilité, qu'il soit en zone sèche ou en zone humide Manguin et *al.*, (1999); Touré et *al.*, (1991). Ces fonctions de répartition de la distribution spatiale ont été inversées et des cartes de distribution spatiale de chacune des formes moléculaires M et S d'*An. gambiae* s.s. ont été produites. L'évaluation de la qualité des modèles à partir de l'analyse de résidus était satisfaisante. La comparaison entre les données prédites par le modèle et les données de terrain, non utilisées pour la modélisation ne présente pas de différence significative. Le modèle est donc validé par confrontation avec les données observées. Ces résultats confirment la troisième hypothèse selon laquelle il est possible, connaissant certains paramètres d'un milieu, d'anticiper sur les fréquences de population de forme moléculaire M et S dans ce milieu.

CONCLUSION GENERALE

L'intérêt de ce travail est de contribuer à la lutte contre l'une des maladies les plus meurtrières de notre temps, en menant des investigations sur l'un des maillons de sa chaîne de transmission : les vecteurs. Maladie parasitaire, maladie à transmission vectorielle, la stratégie de lutte contre le paludisme ne peut se soustraire du contrôle de la faune anophélienne, qui elle dépend, pour une part importante, de la capacité de l'environnement à lui offrir les conditions écologiques favorables à sa survie et à sa reproduction.

Cette étude a eu pour objectifs de décrire l'organisation spatiale des formes moléculaires M et S d'*An. gambiae* au Burkina Faso et d'en identifier les facteurs déterminants. Elle visait également à produire des modèles de distribution spatiale des espèces visées.

Le paradigme de la biogéographie, enrichi des apports indispensables de l'entomologie et des outils et méthodes de l'analyse spatiale, de l'analyse de données statistiques et des systèmes d'information géographique, a permis de structurer la réflexion autour de la distribution spatiale des formes moléculaires M et S d'*An. gambiae* s.s. dans une perspective pluridisciplinaire.

Ce travail de recherche a montré que les formes moléculaires M et S d'*An. gambiae* s.s. sont réparties au Burkina Faso en trois grandes zones. Le nord et le nord-est du pays, correspondant aux régions administratives du Sahel, du Centre-Nord et du Nord, est la zone où prédomine la molécule M. Dans le sud-ouest et le sud du Faso ouvrant les régions du Sud-Ouest, des Cascades et des Hauts-Bassins, la forme moléculaire S est largement dominante. Entre ces deux zones de concentrations s'étend une zone de transition.

Deux groupes de facteurs environnementaux opposés caractérisent les zones de concentration et influencent de façon significative la distribution spatiale des deux formes moléculaires. Le premier groupe de facteurs est lié à l'humidité et se compose de la précipitation, de la végétation et de la pression. Le second groupe évoque plutôt l'aridité et se compose de l'insolation, de la température, de l'évapotranspiration potentielle et du vent. Les deux formes moléculaires présentent des gradients opposés dans leurs abondances relatives le long de l'axe humidité aridité.

A partir de ces facteurs déterminants de l'organisation spatiale des deux formes moléculaires, deux modèles de distribution spatiale (un modèle pour chaque forme moléculaire) ont été réalisés. Ces modèles ont permis, par inversion des fonctions, la réalisation d'une carte de distribution spatiale des abondances relatives des formes moléculaires M et S sur l'ensemble du Burkina Faso.

Au regard de ces résultats, nous pouvons affirmer que l'hypothèse de distribution non aléatoire de formes moléculaires M et S est prouvée. Les facteurs environnementaux sont les déterminants de cette distribution. Ils y contribuent toutefois de façon différentielle, certains facteurs étant moins déterminants que d'autres. Ceci confirme la seconde hypothèse de cette étude. La troisième hypothèse, celle de la possibilité de mise en équation des abondances et des facteurs environnementaux est aussi démontrée. L'opposition entre préférences écologiques des formes moléculaires M et S est également vérifiée, même si la fonction de répartition de l'un n'est pas l'opposé *sensu stricto* de l'autre. Les objectifs fixés au départ de ce travail de recherche sont atteints. Des descriptions cartographiques et statistiques de la distribution spatiale des deux formes moléculaires M et S, sont disponibles. Des modèles de distribution spatiale validés à la fois au plan statistique et au plan empirique ont été mis au point.

Le niveau d'atteinte des objectifs et de validation des hypothèses voudrait que l'on regarde la méthodologie utilisée avec intérêt. Elle met en œuvre un ensemble de processus qui pourraient être répliqués à bon escient sur d'autres sites et pour d'autres vecteurs de maladies. En combinant dans une démarche géographique, l'analyse spatiale, l'analyse de données, les systèmes d'information géographique et l'entomologie, ce travail constitue un point de renouvellement des méthodes géographiques dans l'analyse des problèmes de santé, particulièrement dans leurs aspects environnementaux.

En dépit de ces résultats, ce travail a des limites qu'il convient de relever. La première porte sur la temporalité et la seconde sur la variation des échelles. Pour ce qui concerne la temporalité, une question simple portant sur la validité du modèle en saison sèche peut être posée. En effet, les conditions environnementales de la zone nord du pays pendant la saison pluvieuse pourraient être réalisées en saison sèche dans certaines régions du sud-ouest. Il pourrait être intéressant d'envisager un suivi dans le temps et de prendre en compte la variable temporelle dans la modélisation. Pour ce qui concerne les niveaux géographiques, la mise en évidence d'agrégats spatiaux sur des étendues de superficie réduite mérite que l'on s'interroge sur les conditions particulières qui justifient de telles singularités. En effet les résidus du modèle ont mis en exergue des localités atypiques qui

mériteraient une attention particulière et cela n'est pas envisageable au niveau d'échelle auquel cette étude a été réalisée.

Ce travail, au delà des limites qui viennent d'être évoquées, ouvre des perspectives de recherche aussi bien en entomologie qu'en géographie. Il s'agit d'une réflexion sur les possibilités d'adaptation des méthodes de lutte antivectorielle en fonction des espèces en présence. En d'autres termes, une régionalisation des techniques est envisageable. Cela suppose des études plus approfondies sur l'éthologie des vecteurs, sur la sensibilité des formes moléculaires aux différentes méthodes de lutte, depuis le choix des insecticides pour la pulvérisation intra domiciliaire jusqu'aux moustiquaires imprégnées en passant par les périodes de pulvérisation, la gestion des plans d'eau, etc.

Sans vouloir s'installer dans un mirage d'éradication du paludisme et de contrôle immédiat de la population anophélienne au Burkina Faso, cette étude veut apporter un autre regard sur les questions « environnement et santé » en éclairant, sous un angle géographique, les décideurs et les chercheurs qui s'investissent dans la lutte antivectorielle.

BIBLIOGRAPHIE

- Anyamba, A., Linthicum, K.J., et Tucker, C.J., 2001, Climate-disease connections: Rift Valley Fever in Kenya. : *Cadernos de Saúde Pública*, v. 17, p. 133-140.
- Arnaud, M., et Emery, X., 2000, Estimation et interpolation spatiale, Méthodes déterministes et méthodes géostatistiques, 221 p.
- Aubry, P., 2000, Le traitement des variables régionalisées en écologie apports de la géomatique et de la géostatistique: Thèse de Doctorat, Université Claude Bernard - Lyon 1, p. 480.
- Balenghien, T., 2006, De l'identification des vecteurs du virus West Nile à la modélisation du risque d'infection dans le sud de la France [Thèse de doctorat, thesis], Université Des Sciences Des Technologies et De La Santé De Grenoble, 223 p.
- Banos, A., 2001, A propos de l'analyse spatiale exploratoire des données: Cybergeog : *European Journal of Geography*, v. document 197.
- Barrett, A., et Frank, 2000, Finke's 1792 map of human diseases: the first world disease map?: *Social Science & Medicine*, v. 50, p. 915-921.
- Bavoux, J.-J., 2007, La géographie Objet, méthodes, débats, p. 239.
- Bedard, Y., et Valliere, D., 1995, Qualité des données à Référence Spatiale dans un contexte gouvernemental: Laval Canada, Université de Laval.
- Binka, F.N., Kubaje, A., Adjuik, M., Williams, L.A., Lengeler, C., Maude, G.E., Kajihara, B., Adjamah, J.H., et P.G.Smith, 1996, Impact of permethrin impregnated bednets on child mortality in Kassena-Nankana district, Ghana: a randomized controlled trial: *Trop Med int Health*, v. 1, p. 147-154.
- Biteau, F.C.é., 2006, Surveillance et évaluation du risque de transmission des maladies vectorielles émergentes : apport de la capacité vectorielle Exemple de la fièvre catarrhale du mouton [Epidémiologie thesis], Doctorat de L'Université Montpellier II
- Bonn, F., 1991, Précis de télédétection, Principes et méthodes, Presse Universitaire du Québec, 485 pages p.
- Bouyer, J., 2006, Ecologie des glossines du Mouhoun au Burkina Faso : intérêt pour l'épidémiologie et le contrôle des trypanosomoses africaines [Thèse de Doctorat thesis], Université Montpellier II, 204 p
- Bouzou, I.M., Somé, C.Y.S., Nonguierma, A., et Royer, A., 2001, Evaluation et suivi des plans d'eau de surface par télédétection dans le bassin de Loumbila au Burkina Faso: *Revue Sécheresse* v. 12, p. 31-36.
- Bruce-Chwatt, L.J., et Zulueta, J.D., 1985, *Essential Malariology*: W. Heinemann med. Books Ltd, London, v. chap. 8, p. 166-209.
- Brunet, R., 2000a, Des modèles en géographie ? Sens d'une recherche.: *Bulletin de la Société de Géographie de Liège*, v. n°2, p. 21-30.
- , 2000b, Des modèles en géographie ? Sens d'une recherche, *Bulletin de la Société de Géographie de Liège*, v. Vol.2, p. Pages 21-30.
- Brunet, R., Ferras, R., et Thery, H., 1993, Les mots de la géographie : dictionnaire critique: Montpellier-Paris, DDIPR française, 470 p.
- Carlson, J., Olson, K., Higgs, S., et Beaty, B., 1995, Molecular genetic manipulation of mosquito vectors: *Annual Review of Entomology*, v. 40, p. 359-388.
- Chamussy, J., Charre, J., Durand, M.G., et Le-Berre, M., 1977, Espace, que de brouillons commet-on en son nom: *Brouillon Dupont*, v. n°1.
- Charlwood, J.D., et Edoh, D., 1996, Polymerase chain reaction used to describe larval habitat use by *Anopheles gambiae* complex (Diptera, Culicidae) in the environs of Ifakara, Tanzania: *Journal of Medical Entomology*, v. 33, p. 202-204.
- Charmot, G., 1987, La chimioirésistance de *Plasmodium falciparum*: *Bull. Soc. Path Ex*, v. 80, p. 411-413.
- Chauvet, G., Davidson, G., et Coz, J., 1969, Le complexe *Anopheles gambiae* en Afrique Continentale et à Madagascar: *Cah. ORSTOM, Série Ent. Méd. et Parasitologie*, v. 7, p. 9-12.
- Ciattoni, A., Beucher, S., et Reghezza, M., 2005, La géographie : pourquoi ? Comment ? *Objets et démarches de la géographie d'aujourd'hui*, Hatier.
- Claritas, 2002, Aides en ligne de Vertical Mapper pour Mapinfo.
- Collins, F.H., Mendez, M.A., Rasmussen, M.O., Mehaffey, P.C., Besansky, N.J., et Finnerty, V., 1987, A ribosomal RNA gene probe differentiates member species of the *Anopheles gambiae* complex: *Am. J. Trop. Med. Hyg.*, v. 37(1), p. 37-41.
- Coluzzi, M., 2002, *Plasmodium falciparum* en Afrique subsaharienne. Spéciation récente des vecteurs, transmissibilité, évolution de la pathogénèse contrôle de la malaria et capacité vectorielle. : *Ann. Inst. Pasteur Actualites*, p. 81-99.
- Coluzzi, M., et Sabatini, A., 1967, Cytogenetic observations on sp. A and B of the *Anopheles gambiae* complex: *Parassitologia*, v. 9, p. 73-88.
- , 1968a, Cytogenetic observations on species C, merus and melas of the *Anopheles gambiae* complex.: *Cah. ORSTOM Ent. Méd.*, v. 6: , p. 211-213.

- , 1968b, Cytogenetic observations on species C of the *Anopheles gambiae* complex: *Parassitologia* v. 10, p. 155-165.
- Coluzzi, M., Sabatini, A., Petrarca, V., et Deco, M.A.D., 1979, Chromosomal differentiation and adaptation to human environments in the *Anopheles gambiae* complex. : *Trans. Of the Roy. Soc. Of Trop. Med. and Hyg.*, v. 73:, p. 483-497.
- Conlisk, J., 1971, When collinearity is desirable: *Western Economic Journal*, v. 9, p. 393-407.
- Costantini, C., Ayala, D., Guelbeogo, W.M., Pombi, M., Somé, C.Y., Bassole, I.H., Ose, K., Fotsing, J.-M., Sagnon, N.F., Fontenille, D., Besansky, N.J., et Simard, F., 2009, Living at the edge: biogeographic patterns of habitat segregation conform to speciation by niche expansion in *Anopheles gambiae*: *BioMed Center Ecology*, v. 9:16.
- Costantini, C., Sagnon, N.F., Della, T.A., M., D., J., B., Gibson, G., et M., C., 1998, Odor-mediated host preference of West Africa mosquitoes, with particular preferences to malaria vectors: *Am J Trop Med Hyg*, v. 58, p. 56-63.
- Coz, J., 1973, Contribution à l'étude du complexe *Anopheles gambiae*. Répartition géographique et saisonnière en Afrique de l'ouest: *Cahier ORSTOM, Série Ent. Méd. et Parasitologie*, v. 9, p. 3-31.
- Crampton, J.M., et Hill, S.M., 1993, DNA probes for identification of malaria vectors: *Parassitologia*, v. 35, p. 31-44.
- Curtis, C.F., 1996, Impregnated bed nets, malaria control and child mortality in Africa: *Tropical medicine and international Health*, v. 1, p. 137 - 138.
- Dajoz, R., 2006, *Précis d'écologie*, 631 p.
- Davidson, G., 1964, *Anopheles gambiae*, a complex of species: *Bull. Wld. Hlth. Org.*, p. 625.
- Davidson, G., et Jakson, C., 1962, Incipient speciation in *Anopheles gambiae* Giles: *Bull. Wld. Hlth. Org.*, v. 27, p. 303-305.
- Davies, F.G., Kileulu, E., Linthicum, K.J., et Pegram, R.G., 1992, Patterns of Rift Valley fever activity in Zambia: *Epidemiology and Infections*, v. 108, p. 185-191.
- Devilliers, R., et Jeansoulin, R., 2005, Qualité de l'information géographique : Concepts généraux: *Qualité de l'information géographique*, v. IGAT, p. Pages 37-47.
- Diabaté, A., 1999, Evaluation de la résistance des vecteurs du paludisme vis-à-vis des pyréthnoïdes au Burkina Faso [Doctorat 3ème Cycle thesis]: Ouagadougou, Université de Ouagadougou.
- Droesbeke, J.-J., Michel, L., et Gilbert, S., 2006, Analyse statistiques des données spatiales, 468 p.
- Durand-Dastès, F., 2001, Les concepts de la modélisation en analyse spatiale: *Modèles en analyse spatiale*, v. IGAT, p. Pages 31-57.
- ESRI, 2006, ArcGIS 9, Using ArcGIS Desktop.
- Favia, G., Della, T.A., Bagayoko, M., Lanfrancotti, A., Sagnon, N.F., Touré, Y.T., et Coluzzi, M., 1997, Molecular identification of sympatric chromosomal forms of *Anopheles gambiae* and further evidence of their isolation. : *Insect Molecular Biology* v. 6(4), p. 377-383.
- Favia, G., Lanfrancotti, A., Spanos, L., et Coll, 2001, Molecular characterization of ribosomal DNA (rDNA) polymorphisms discriminating chromosomal forms of *Anopheles gambiae* s.s. : *Insect Mol Biol.*, v. 10, p. 19-23.
- Fontenille, D., Faye, O., Konate, L., Sy, N., et Collins, F., 1993, Comparaison des techniques PCR et cytogénétique pour la détermination des membres du complexe *Anopheles gambiae* au Sénégal. : *Ann.Parasitol. Hum. Comp.* : v. 68(5/6), p. 239-240.
- Fontes, J., Diallo, A., et Compaoré, J.A., 1994, Carte de la Végétation Naturelle et de l'Occupation du Sol - Burkina Faso Échelle 1 : 1.000.000.
- Fontes, J., et Guinko, S., 1995, Carte de végétation et d'occupation du sol du Burkina Faso. Notice explicative., Ministère de la coopération française projet Campus ICIV, IDR, 53 p.
- Fotsing, J.-M., 2003, Qu'apportent les SIG à la recherche scientifique ? : *Colloque International Géomatique et application*, v. Presses Universitaire d'Orléans N°1, p. 13-14.
- Franc, A., 2001, Formalisation mathématique des interactions spatiales, *Modèles en analyse spatiale*, Volume IGAT, p. Pages 283-299.
- Gillies, M.T., et Coetzee, M.C., 1987, A supplement to the Anophelinae of Africa south of the sahara (Afrotropical region): publication of the south African institute for Medical Research, Johannesburg, v. volume 55 p. 143pp.
- Gillies, M.T., et Meillon, B.D., 1968, The Anophelinae of Africa south of the sahara (Ethiopian Zoogeographical region): The south African institute for Medical Research, Johannesburg, , v. 2nd Ed. , p. 343p.
- Green, C., 1972, Cytological maps for the practical identification of females of the three freshwater species of the *Anopheles gambiae* complex.: *Ann. Trop. Med. Parasitol.*, v. 66, p. 143-147.
- Guelbeogo, W.M., 1999, Etude du comportement des larves du complexe *Anopheles gambiae* et son impact sur les interactions inter et intra-spécifiques [Mémoire de DEA. thesis]: Ouagadougou Burkina Faso, Université de Ouagadougou Burkina Faso,.
- Guermond, Y., 2005, *Modélisations en géographie, déterminisme et complexité*, p. 389.
- Guinko, S., 1984, *Végétation de Haute Volta* [Doctorat d'Etat thesis]: Bordeaux, Bordeaux III.

- Guisan, A., et Zimmermann, N.E., 2000, Prédictive habitat distribution models in ecology: Elsevier Ecological modelling, v. 135, p. 147-186.
- Gumuchian, H., et Marois, C., 2000, Initiation à la recherche en géographie, Aménagement, développement territorial, environnement, p. 425.
- Haggett, 1965, Locational Analysis in Human Geography.
- Hamon, J., Adam, J.P., et Grjebine, A., 1956, Observation sur la répartition et le comportement des anophèles de l'Afrique Equatoriale française, du Cameroun et de l'Afrique Occidentale. : Bulletin de l'Organisation mondiale de la santé, v. 15, p. 549-591.
- Hendrickx, G., de-La-Rocque, S., Reid, R., et Wint, W., 2001, Spatial trypanosomosis management: from data-layers to decision making. Trends Parasitology, v. 17, p. 35-41.
- Hendrickx, G., Napala, A., Dao, B., Batawui, D., De-Deken, R., Vermeilen, A., Slingenbergh, J.H.W., et 1989, A systematic approach to area-wide tsetse distribution and abundance maps. : Bulletin of Entomological Research, v. 89, p. 231-244.
- Holt, R.A., Subramanian, G.M., Halpern, A., Granger G. Sutton, Rosane Charlab, Nusskern, D.R., Wincker, P., Clark, A.G., Ribeiro, J.M.C., Wides, R., Salzberg, S.L., Loftus, B., Yandell, M., Majoros, W.H., Rusch, D.B., Lai, Z., Kraft, C.L., Abril, J.F., Anthouard, V., Arensburger, P., Atkinson, P.W., Baden, H., Berardinis, V.d., Baldwin, D., Benes, V., Biedler, J., Blass, C., Bolanos, R., Boscus, D., Barnstead, M., Cai, S., Center, A., Chatuverdi, K., Christophides, G.K., Chrystal, M.A., Clamp, M., Cravchik, A., Curwen, V., Dana, A., Delcher, A., Dew, I., Evans, C.A., Flanigan, M., Grundschober-Freimoser, A., Friedli, L., Gu, Z., Guan, P., Guigo, R., Hillenmeyer, M.E., Hladun, S.L., Hogan, J.R., Hong, Y.S., Hoover, J., Jaillon, O., Ke, Z., Kodira, C., Kokoza, E., Koutsos, A., Letunic, I., Levitsky, A., Liang, Y., Lin, J.-J., Lobo, N.F., Lopez, J.R., Malek, J.A., McIntosh, T.C., Meister, S., Miller, J., Mobarry, C., Mongin, E., Murphy, S.D., O'Brochta, D.A., Pfannkoch, C., Qi, R., Regier, M.A., Remington, K., Shao, H., Sharakhova, M.V., Sitter, C.D., Shetty, J., Smith, T.J., Strong, R., Sun, J., Thomasova, D., Ton, L.Q., Topalis, P., Tu, Z., Unger, M.F., Walenz, B., Wang, A., Wang, J., Wang, M., Wang, X., Woodford, K.J., Wortman, J.R., Wu, M., Yao, A., Zdobnov, E.M., Zhang, H., Zhao, Q., et *al.*, 2002, The Genome Sequence of the Malaria Mosquito *Anopheles gambiae* Science v. 298, p. 129 - 149.
- Hunt, R.H., Coetzee, M., et Fittene, M., 1998, The *Anopheles gambiae* complex a new species from Ethiopia: Trans. Of the Roy. Soc. Of Trop. Med. and Hyg., v. 92, p. 231-235.
- Ilboudo-Sanogo, E., 2001, Influence des rideaux imprégnés d'insecticides sur les paramètres entomologiques de la transmission du paludisme en zone rurale au Burkina Faso. [Doctorat Unique thesis]: Ouagadougou, Université de Ouagadougou
- Jambou M., 1999, Méthodes de base de l'analyse des données, p. 412.
- Joost, S., 2006, The geographical dimension of genetic diversity: a giscience contribution for the conservation, of animal genetic resources: Thèse de Doctorat, École Polytechnique Fédérale de Lausanne, p. 178.
- Kennedy, P., 1992, A Guide to Econometrics, Blackwell Publishers, 410 p.
- Koenraadt, C.J.M., Githeko, A.K., et Takken, W., 2004, The effect of rainfall and evapotranspiration on the temporal dynamics of *Anopheles gambiae* s.s and *Anopheles arabiensis* in a Kenyan village: Acta Tropica, v. 90, p. 141-153.
- Lacey, L.A., Urbina, M.J., et Heizman, C.M., 1984, Sustained-release formulations of *Bacillus sphaericus* and *Bacillus thuringiensis* (H-14) for control of container-breeding *Culex quinquefasciatus*.: Mosquito news, v. 44, p. 26-32.
- Lacoste, A., et Salanon, R., 2006, Elément de biogéographie et d'écologie, Une compréhension de la biosphère par l'analyse des composantes majeures des écosystèmes.
- Langlois, P., 2005, Complexité et système spatiaux: Modélisations en géographie, déterminisme et complexité, v. IGAT, p. Pages 299-319.
- Langlois, P., et Reguer, D., 2005, La place du modèle et de la modélisation en sciences humaines: Modélisations en géographie, déterminisme et complexité, v. IGAT, p. Pages 35-48.
- Le-Bras, M., et Malvy, J.-M.D., 2004, Le complexe pathogène tropical : Regard nouveau sur un concept ancien: Médecine tropicale, v. vol. 64, p. 613-618.
- Le-Moigne, J.-L., 1990, La modélisation des systèmes complexes: Paris, Dunod, 180 p.
- Le-Sueur, D., et Sharp, B.L., 1988, The breeding requirements of three members of the *Anopheles gambiae* Giles complex (Diptera: Culicidae) in the endemic malaria area of Natal, : south Africa. Bull. Ent. Res., v. 78, p. 549-560.
- Lebart, L., Piron, M., et Morineau, A., 2006, Statistique exploratoire multidimensionnelle. Visualisation et inférence en fouilles de données, 464 p.
- Linthicum, K.J., Anyamba, A., Tucker, C.J., Kelley, P.W., Myers, M.F., et Peters, C.J., 1999, Climate and satellite indicators to forecast Rift Valley Fever epidemics in Kenya: Science, v. 285, p. 397-400.
- Linthicum, K.J., Bailey, C.L., Davies, F.G., et Tucker, C.J., 1987, Detection of Rift Valley fever viral activity in Kenya by satellite remote sensing imagery. : Science, v. 235, p. 1656-1659.
- Linthicum, K.J., Bailey, C.L., Tucker, C.J., Mitchell, K.D., Logan, T.M., Davies, F.G., Kamau, C.W., Thande, P.C., et Wagatoh, J.N., 1990, Application of polar-orbiting, meteorological satellite data to detect flooding of Rift

- Valley Fever virus vector mosquito habitats in Kenya. : Medical and Veterinary Entomology, . v. 4, p. 433-438.
- Manguin, S., Fontenille, D., F.Chandre, Lochouarn, L., Mouchet, J., Kengne, P., et Guillet, P., 1999, Interactions génétiques entre population vectrices parasitaires et humains: l'exemple du paludisme à *Plasmodium falciparum*, Journée SPE du 13 octobre 1999 : Génétique et maladies infectieuses dans l'environnement tropical.: Institut Pasteur à Paris, p. 7.
- Mathian, H., et Piron, M., 2001, Echelles géographiques et méthodes statistiques multidimensionnelle: Modèles en analyse spatiale, v. IGAT, p. Pages 61-101.
- MC-Crae, A.W.R., 1984, Ovoposition by African malaria vector mosquitoes. II Effects of site tone, water type and conspecific immatures on target selection by freshwater *Anopheles gambiae* Giles *sensu lato*: Annals of Tropical Medicine and Parasitology, v. 78, p. 307-318.
- MEE-Burkina, M.d.I.E.d.I.E., 2002, Rapport sur l'état de l'environnement au Burkina Faso (REEB).
- Minakawa, N., Muteru, C.M., Githure, J.I., Beier, J., et Yan, G., 1999, Spatial distribution and habitat characterization of Anopheline mosquito larvae in Western Kenya. : American Tropical Medicine Hygiene, v. 61, p. 1010-1016.
- Mouchet, J., et Carnevale, P., 1991, Les vecteurs et la transmission: Paludisme M. Danis et J. Mouchet coord, v. ELLIPSES Ed, p. 35-59.
- Mouchet, J., Carnevale, P., Coosemans, M., Julvez, J., Manguin, S., Richard-Lenoble, D., et Sircoulon, J., 2004, Biodiversité du paludisme dans le monde: Paris, Pages 428 p.
- Muirhead-Thomson, R., 1951, Studies on salt-water and fresh-water *Anopheles gambiae* on the East African coast: Bull. Entomol. : , v. Res. 41, p. 487-502.
- Mushinzimana, E., Munga, S., Minakawa, N., Li, L., Feng, C.-c., Bian, L., Kitron, U., Schmidt, C., Beck, L., Zhou, G., Githeko, A.K., et Yan, G., 2006, Landscape determinants and remote sensing of anopheline mosquito larval habitats in the western Kenya highlands: Malaria Journal, v. 5:13.
- Nébié, I., 2002, Influence à long terme des rideaux imprégnés d'insecticide sur l'immunité anti-palustre chez des enfants vivant en zone rurale au Burkina Faso [Doctorat unique thesis], Université de Ouagadougou.
- O.M.S., 2006, Prévenir les maladies par des environnements salubres - vers une évaluation de la charge des maladies environnementales: Genève, OMS.
- O.R.S.T.O.M., C.O.d.D., 1969-1973, Carte pédologique de reconnaissance de la République de Haute-Volta. . Échelle: 1 : 500 000.
- Paterson, H., 1964, Direct evidence for the specific distinctness of Forms A, B and C of the *Anopheles gambiae* complex.: Riv. Malar, v. 43, p. 191-196.
- Petrarca, V., Petrangeli, G., Rossi, P., et Sabatinelli, G., 1986, Etude Chromosomique d'*Anopheles gambiae* et *Anopheles arabiensis* à Ouagadougou (Burkina Faso) et dans quelques villages voisins: Parassitologia, v. 28, p. 41-61.
- Pin-Diop, R., 2006, Spatialisation du risque de transmission de la Fièvre de la Vallée du Rift en milieu agropastoral sahélien du Sénégal septentrional [Géographie - Espace, Développement et Santé thesis], 238p.
- PNLP, 2007, Plan stratégique de lutte contre le paludisme 2006-2010: Ministère de la Santé Burkina Faso, p. 80.
- Pumain, D., et Saint-julien Thérèse., 1997, L'analyse Spatiale. 1. La localisation dans l'espace: Paris, Page 158 p.
- Quinones, M.L., Lines, J.D., Thomson, M.C., Jawara, M., Morris, J., et Greenwood, B.M., 1997, *Anopheles gambiae* gonotrophic cycle duration, biting and exiting behaviour unaffected by permethrin impregnated bednets in the Gambia, : Medical and Veterinary Entomology, , v. 11, p. 71-78.
- Rajagopalan, P.K., Das, K.N., Panicker, R., D., R.R., L.S., S., et Lines, J.D., 1991, Environmental and water management for mosquito control: Control of disease vectors in community, v. Ed C.F. Curtis Wolfe.
- Ratana, S., Lee, W.J., Ugsang, D.M., et Linthicum, K.J., 2005, Identification and characterization of larval and adult anopheline mosquito habitats in the Republic of Korea: potential use of remotely sensed data to estimate mosquito distributions: International Journal of Health Geographics, v. 4:17.
- Robert, V., Awono-Ambene, H.P., et Thioulouse, J., 1998, Ecology of larval mosquitoes, with special reference to *Anopheles arabiensis* (Diptera: Culicidae) in market-garden wells in urban Dakar, Senegal. : J Med Entomol, v. 35(6), p. 948-55.
- Robert, V., et Carnevale, P., 1984, Les vecteurs des paludismes en Afrique subsaharienne: Etudes Médicales, v. 2, p. 79-90.
- Robert, V., Petrarca, V., Carnevale, P., Ovazza, L., et Coluzzi, M., 1989, Analyse cytogénétique du complexe *Anopheles gambiae* dans la région de Bobo Dioulasso (Burkina Faso). Annales de Parasitologie humaine Comparée, v. 64, p. 290-311.
- Robinson, T.P., Rogers, D.J., et Williams, B.G., 1997, Mapping tsetse habitat suitability in the common fly belt of Southern Africa using multivariate analysis of climate and remotely sensed vegetation data: Medical and Veterinary Entomology,, v. 11, p. 235-245.
- Rodhain, F., et Perez, C., 1985, Précis d'entomologie médicale et vétérinaire: Paris France, éditeur Maloine S.A., 485 p.
- Rodolphe, D., et Robert, J., 2005, Qualité de l'information géographique, p. 302.
- Rogers, D.J., et Randolph, S.E., 1993, Distribution of tsetse and ticks in Africa: past, present, future.: Parasitology today, v. 9, p. 266-271.

- Rogers, D.J., et Williams, B.G., 1993, Monitoring trypanosomiasis in space and time *Parasitology*, v. 106, p. S77-S92.
- Sabatier, P., Babouchkine, J.-M., Morvan, J., Polidori, L., Lacaux, J.-P., Braak, L., Dubois, M.A., et Chaude, H., 2004, S2.Dengue. Surveillance spatiale et de la dengue. Conception d'outils d'acquisition, de cheminement, d'analyse et de prévision épidémiologique: ITBM-RBM v. 25 (2004) p. 292-296.
- Sagnon, N.F., 1999, Identification moléculaire dans l'Etude de la distribution spatiale et de la structure de la population préimaginaire des taxa du complexe *Anopheles gambiae* vecteurs du paludisme au Burkina Faso [PhD thesis]: Rome, Université de Rome «la Sapienza»
- Sagnon, N.F., Costantini, C., Torre, A.d., et M., C., 2000, Random distribution of larval populations of sympatric members of *Anopheles gambiae* complex in Burkina Faso: *Parassitologia*, v. 42, p. 157.
- Sanders, L., 2001, Modèles en analyse spatiale, p. 329.
- Sangaré, D., 2000, Dynamique des populations d'*Anopheles gambiae* s.l., d'*Anopheles fenestus* et *Plasmodium falciparum* dans le système de transmission par relais du paludisme à Donéguébougou Arrondissement de Kati, Université du Mali.
- Sirima SB, Konaté A, Tiono AB, Convelbo, Cousens S, Pagnoni F, 2003, early treatment of childhood fevers with pre-packaged antimalarial drug in the home reduces severe malaria morbidity in Burkina Faso, *Tropical medicine and International Health*, 2003,8(2), pp 133-9,
- Scott, J.A., Brogdon, W.G., et Collins, F.H., 1993, Identification of single specimens of *Anopheles gambiae* complex by the polymerase chain reaction. . *Am. J. Trop. Med. Hyg.*, v. 49, p. 520-529.
- Service, M.W., 1993, Mosquitoes(culicidae): Lane R.P. et Crosskey R.W.(eds), *Medical insects and Arachnids*. Chapman and Hall, London., p. 51-77.
- Service, M.W., Joshi, G.P., et Prodhan, G.D., 1978, A Survey of *Anopheles gambiae* (species A) and *An. arabiensis* (Species B) of the *An. gambiae* Giles complex in the Kisumu area of Kenya following insecticidal spraying with OMS-43 (fenitrothion): *Annales of Tropical Medicine and Parasitology*, v. Vol. 72, No.4, p. 377-386.
- Snow, R.W., et Marsh, K., 1995, Will reducing *Plasmodium falciparum* transmission alter malaria mortality among African children? : *Parasitology Today*, v. 11, p. 188-190.
- Somé, Y.S.C., 2000, Suivi et Evaluation des plans d'eau de surface par télédétection dans le bassin de Loumbila au Burkina Faso [DESS thesis]: Niamey Niger, Centre Régional d'Enseignement Spécialisé en Agriculture (CRESA) ; Université Abdou Moumouni.
- , 2006, De la pratique à la conception, pour une utilisation des logiciels comme moyens de formation en SIG et en télédétection: *Cahiers du CERLSHS*, v. Edition Spéciale, p. 81-96.
- Souris, M., 2006, Contraintes d'intégrité spatiales *Traité IGAT*, Hermès, v. Fondements des bases de données spatiales, sous la direction de Sylvie Servigne et Thérèse Libourel., p. 99-123.
- Sysbase, 1994, Aide en ligne de Power AMC Designor.
- Thériault, M., 1995, Système d'Information géographique, concept fondamentaux, notes de cours LATG: Québec, Université de Laval.
- Thomas, E.P., Keating, J., Swalm, C., Mbogo, C.M., Githeko, A.K., Regens, J.L., Githure, J.I., Andrews, L., et Beier, J.C., 2003, Linking field-based ecological data with remotely sensed data using a geographic information system in two malaria endemic urban areas of Kenya: *Malaria Journal*, v. 2:44.
- Touré, Y.T., 1988, Les indicateurs des situations épidémiologiques: Conférence internationale sur les stratégies de lutte contre les paludismes OCCGE - Centre Muraz Bobo-Dioulasso, p. 31-35.
- Touré, Y.T., Petrarca, V., M., C., Coulibaly, S.F., Sankaré, O., Maïga, H., et Sow, M., 1991, The gambiae complex in West Africa: *Science in Africa*, v. Achievement and Prospectus AAAS, p. 55 - 75.
- Touré, Y.T., Petrarca, V., Traoré, S.F., Coulibaly, A., Maïga, H.M., Sankaré, O., Sow, M., et Dideco, 1998, polymorphism of chromosomally recognized taxa of the *Anopheles gambiae* complex in Mali,: *West Africa Parassitologia* v. 40, p. 477-511.
- Tran, A., 2004, Télédétection et Épidémiologie : Modélisation de la dynamique de populations d'insectes et application au contrôle de maladies à transmission vectorielle [Thèse de Doctorat, thesis], Université Louis Pasteur Strasbourg 1.
- Tran, A., Bateau-Coroller, F., Guis, H., et Roger, F., 2005, Modélisation des maladies vectorielles: *Epidémiol et santé anim*, v. 47, p. 35-51.
- Trape, J.F., et Roger, C., 1996, Combating malaria morbidity by reducing transmission. : *Parasitology today*, v. 12, p. 236-240.
- Vallée, J., Souris, M., Fournet, F., Bochaton, A., Mobillion, V., Peyronnie, K., et Salem, G., 2007, Sampling in health geography: reconciling geographical objectives and probabilistic methods. An example of a health survey in Vientiane (Lao PDR): *Emerging Themes in Epidemiology*, v. 4:6.
- Wernsdorfer, W.H., 1991, The development and Spread of drug resistant malaria: *Parasitology Today*, v. 7, p. 297-303.
- White, G.B., 1974, *Anopheles gambiae* complex and disease transmission in Africa: *Transactions of the Royal Society of Tropical Medicine and hygiene*, v. 68, p. 278-301.
- White, G.B., et Rosen, P., 1973, Comparative studies on sibling species of the *Anopheles gambiae* Giles complex (Diptera., Culicidae). II. Ecology of species A and B in savanna around Kaduna, Nigeria, during transition from wet to dry season. : *Bull. Ent. Res.*, v. 62, p. 613-625.

- Ye-Ebiyo, Y., Pollack, R., A.Kiszewski, et A., S., 2003, Enhancement of development of larval *Anopheles arabiensis* by proximity and when crowded Am. J. Trop. Med. Hyg., v. 68 (6), p. 748 - 52.
- YÉ, Y., 2005, Incorporating environmental factors in modelling malaria transmission in under five children in Rural Burkina Faso: Doctor scientiarum humanarum, Ruprecht-Karls-University Heidelberg., p. 170.
- Zaninetti, J.-M., 2005, Statistique spatiale. Méthodes et applications géomatiques, 320 p.

SITES INTERNET CITES DANS LE TEXTE⁶²

http://www.fnh.org/francais/doc/en_ligne/sante/dossier1_intro.htm

http://www.fnh.org/francais/doc/en_ligne/sante/dossier1_art1.htm

<http://www.who.int/mediacentre/news/releases/2006/pr32/fr/index.html>

<http://www.aidh.org/sante/sante-environn.htm>

http://www.scienceshumaines.com/-0ales-sciences-humaines-et-la-sante-0a_fr_13795.html

<http://www.csiss.org/classics/content/106>

http://fr.wikipedia.org/wiki/John_Snow

<http://cat.inist.fr/?aModele=afficheN&cpsid=16594466>

<http://www.cybergeog.eu/index4056.html>

<http://www.techno-science.net/?onglet=glossaire&definition=2569>

http://agora.qc.ca/mot.nsf/Dossiers/Justus_von_Liebig

http://fr.wikipedia.org/wiki/Justus_von_Liebig

<http://web.worldbank.org/WBSITE/EXTERNAL/ACCUEILEXTN/PAYSEXTN/AFRICAINFRENCHEXT/BURKINAFASOINFRENCHEXTN/0,,contentMDK:21744044~pagePK:1497618~piPK:217854~theSitePK:463455,00.html>

<http://209.85.129.132/search?q=cache:T1SAeMEEo0YJ:www.rollbackmalaria.org/countryaction/nsp/burkina.pdf+PNLP+Burkina+p%C3%A9riode+de+transmission&cd=7&hl=fr&ct=clnk&gl=fr&client=firefox-a>

<http://www.traveldocor.co.uk/malaria.htm>

<http://www.primature.gov.bf/burkina/infopratique/fsites.htm>

http://www.inforoute-communale.gov.bf/col_cadre1.htm

<http://www.insd.bf/>

http://www.insd.bf/donnees/donnees_structurelles/T0317.htm

<http://www2.unil.ch/biomapper/>

http://openmodeller.sourceforge.net/index.php?option=com_frontpage&Itemid=1

<http://www.mbr-pwrc.usgs.gov/software/comdyn.html>

<http://www.biomedcentral.com/1472-6785/9/16>

http://www.th.ird.fr/publications/publications_auteurs_souris.html

⁶² Tous les liens ont été contrôlés actif pour la dernière fois le 1^{er} août 2009

<http://www.ete-online.com/content/4/1/6>

<http://france.elsevier.com/direct/RBMRET/>

ANNEXES

ANNEXE 1 : METHODES DE COLLECTE DES MOUSTIQUES SUR LE TERRAIN

CAPTURE PAR PULVERISATION INTRA-DOMICILIAIRE D'INSECTICIDES

Elle consiste à pulvériser une solution d'insecticide à faible rémanence (d-phenothrin et pyrèthre) vers les parois intérieures et le toit des habitations afin de collecter les moustiques au repos.

Après avoir évacué les vivres des habitations humaines, des draps sont étalés à l'intérieur des maisons afin de recouvrir tout le plancher. Une solution d'insecticide est ensuite pulvérisée simultanément par deux agents, l'un à l'intérieur et l'autre à l'extérieur sur les issues, par lesquelles les moustiques peuvent s'échapper. Après une dizaine de minutes d'attente, les moustiques sous l'effet « knock-down » des insecticides sont dispersés sur les draps. Les draps sont ensuite retirés des maisons avec précaution afin de ne pas endommager les moustiques morts. Ces derniers sont alors récupérés et placés dans des boîtes à pétri portant des indications sur le lieu, la date de collecte, le numéro de la concession et de la maison (Photo 1 Photo 1& 2). Les habitations sont aérées quelques minutes après le retrait des draps afin d'éliminer les effluves résiduels de l'insecticide.



Photo 1 : Séance de capture par pulvérisation intradomiciliaire : Etape de nébulisation



Photo 2 : Séance de capture par pulvérisation intradomiciliaire : Etape de ramassage des moustiques

ANNEXE 2 : IDENTIFICATION DES FORMES MOLECULAIRES

PROTOCOLE D'IDENTIFICATION DES FORMES MOLECULAIRES

CENTRE NATIONAL DE RECHERCHE ET DE FORMATION SUR LE PALUDISME

PRINCIPE DE L'IDENTIFICATION DES MEMBRES DU COMPLEXE *AN. GAMBIAE* PAR LA PCR (Polymerized Chain Reaction)

Le séquençage complet du génome d'*An. gambiae* a débuté en 2001 et est achevé en octobre 2002. Ce génome contient environ 14000 gènes. Seul un petit nombre de ces gènes ont été identifiés. L'application de la PCR à l'identification des membres du complexe *An. gambiae* est basée sur l'analyse des mutations nucléotidiques au niveau de ces gènes. Grâce à la technique de la PCR-FRLP (Restriction Fragment Length Polymorphisms), la structure de l'ADN ribosomal d'*An. gambiae* a été déterminée. Cette structure contient des mutations nucléotidiques permettant de distinguer les membres du complexe *An. gambiae*. Une illustration des mutations de l'ADNr des membres du complexe *An. gambiae* est présentée à la Figure 66.

	561	V	600
<i>An. gambiae</i> forme S	GTGCCCCTTCCTCGATGGCGCAACGAACCATCTTGGTCTG		
<i>An. gambiae</i> forme M		AT	
<i>An. arabiensis</i>		T.....T.....T.....	
<i>An. melas</i>		TT..TTC	
<i>An. merus</i>		T..TTT	
<i>An. quadriannulatus</i>		AT...T	

Figure 66 : Structure de l'ADN ribosomal des membres du complexe *An. gambiae* (Scott et al, 1993).

L'ADN ribosomal de certains membres du complexe *An. gambiae* comporte également des sites de restrictions reconnus par des enzymes, tel que *HhaI* et *CfoI*. Ces dernières qui sont des enzymes de restriction capables de cliver l'ADNr d'*An. gambiae* s.s.. Cet ADNr est organisé en tandem. Chaque unité répétitive est formée dans l'ordre de l'extrémité 5' vers 3', d'un espace transcrit externe (ETS), des régions codant pour les acides ribonucléiques ribosomaux (ARNr) 18S et 28S qui sont séparés par un espace transcrit interne (ITS) contenant la région codante pour l'ARNr 5,8S (Figure 67). Ces motifs sont séparés entre eux par des espaces inter géniques (IGS).

La PCR permet une identification des membres du complexe *An. gambiae*, mais aussi des formes moléculaires d'*An. gambiae* s.s.. Les amorces utilisés pour l'identification des différentes espèces d'*An. gambiae* s.s. sont :

- l'amorce Universelle (UN) commune à toutes les espèces, dont la séquence est (5'-GTG TGC CCC TTC CTC GAT GT-3').
- l'amorce *gambiae* (GA) spécifique à *An. gambiae* s.s. dont la séquence est (5'-CTG GTT TGG TCG GCA CGT TT-3').

- Enfin l'amorce *arabiensis* (**AR**) spécifique à *An. arabiensis*, qui a pour séquence (5'-AAG TGT CCT TCT CCA TCC TA-3')

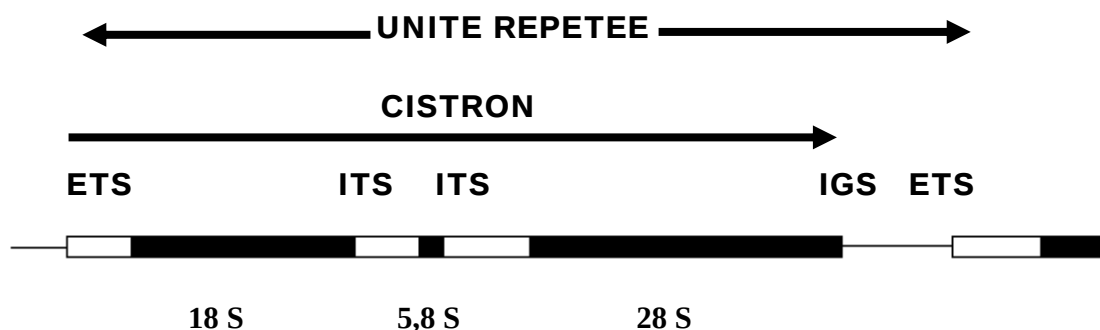


Figure 67 : Schéma d'une unité répétée en tandem

MODE OPERATOIRE

Ces techniques d'identification moléculaire nécessitent trois grandes étapes : l'extraction de l'ADN, l'amplification de l'ADN et l'électrophorèse.

EXTRACTION DE L'ADN

L'ADN de la tête, des pattes et des ailes a été extrait. Le matériel biologique est broyé dans 50µl de solution tampon de cassage (grinding buffer) dont la composition est la suivante 5M de NaCl, 0,5M d'EDTA, 1M de Tris-HCl, 10% de SDS à pH 7,5. L'homogénat obtenu est incubé à 65°C au bain marie pendant 30mn. Puis 7µl d'acétate de potassium 8M est ajouté pour précipiter les protéines. Après agitation la solution est incubée à -20°C pendant 30mn. La température basse accélère la précipitation des protéines. Le mélange est ensuite centrifugé à 14000 rpm pendant 15mn. Le surnageant obtenu est alors transvasé dans un tube contenant 100µl d'éthanol absolue (100%) préalablement préparé (portant le numéro d'identification du moustique). L'alcool solubilise les lipides et facilite la précipitation des acides nucléiques. Le mélange obtenu est incubé à -20°C pendant 30mn, puis centrifugé à 14000 rpm pendant 15mn. Le surnageant est vidé sur du papier buvard et 100µl d'éthanol (70%) sont ajoutés au culot. Cette solution est centrifugée à 14000rpm pendant 5mn. Le surnageant est encore vidé sur du papier buvard et les tubes sont séchés dans un déclicateur sous vide pendant 10 à 15mn. Le culot obtenu est dissout dans 300µl d'eau distillée stérile et incubé à 65°C pendant 10mn. La solution finale obtenue est conservée à 4°C jusqu'à l'amplification de l'ADN.

L'AMPLIFICATION DE L'ADN

Pour l'identification des espèces, *An. gambiae* s.s. et *An. arabiensis*, l'ADN extrait est amplifié dans 9.2µl d'une solution mixte composé d'un tampon PCR (10X), du chlorure de magnésium $MgCl_2$ (25mM), de desoxyribonucléotides triphosphate dNTPs (10mM), d'une amorce universelle UN, d'une amorce GA spécifique à l'espèce *An. gambiae* s.s. (2pM), d'une amorce AR spécifique à l'espèce *An. arabiensis* (6,2pM) et 0,35 unité de Taq polymérase. A cette solution mixte 0,8 µl d'ADN à amplifier est ajouté et le mélange est incubé dans un thermocycleur de type PTC100, PTC200 ou Mastercycler. Le programme d'amplification est le suivant : 2mn à 95°C pour activer la DNAPolymérase suivies de 30 cycles dont chacun à consisté en 30 secondes de dénaturation à 95°C, 30secondes à 50°C pour l'appariement des amorces et enfin 30 secondes d'élongation à 72°C. Les produits de l'amplification sont maintenus à 72°C pendant 7mn pour parfaire les différentes phases d'élongation (Figure 68).

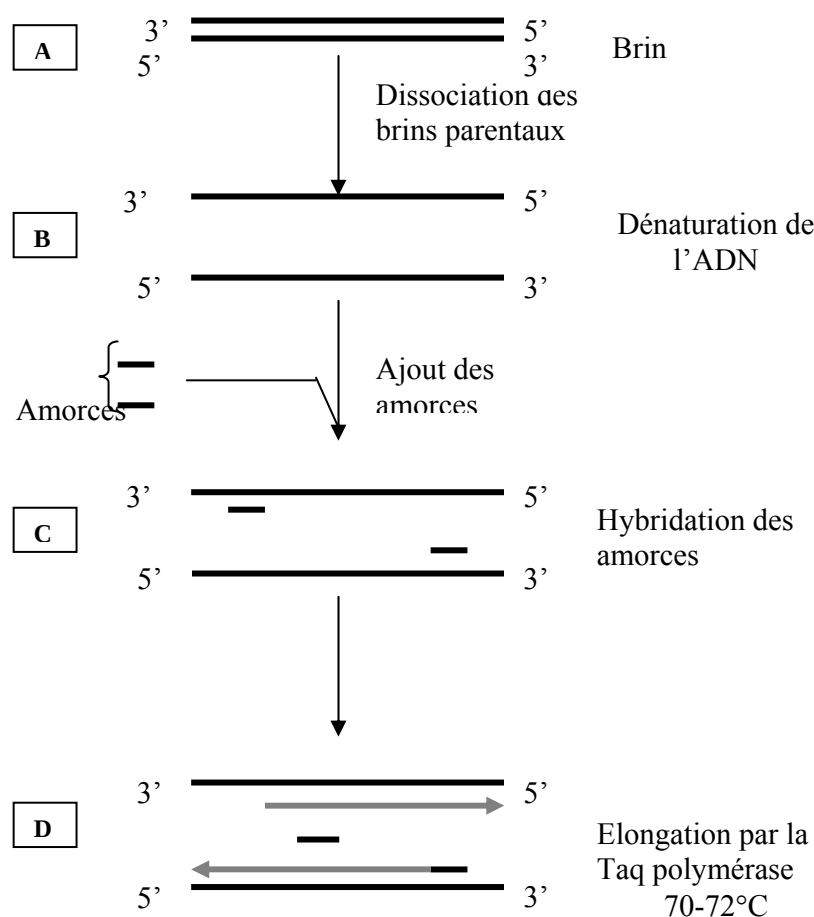


Figure 68 : Schéma d'amplification d'une séquence d'ADN

LA RESTRICTION

Pour l'identification des formes moléculaires d'*An. gambiae* s.s, une enzyme de restriction a été ajouté aux produits d'amplification. Cet enzyme est appelé *CfoI* et est utilisé selon le protocole établi par Fanello *et al*,(2002). Pour cela, 1µl d'une solution mixte contenant un tampon buffer (b4), de l'eau distillée stérile et de l'enzyme (*CfoI*) a été ajouté à chaque produit d'amplification. Les tubes ont été ensuite incubés à 37°C pendant 3heures.

L'ELECTROPHORESE

Les produits obtenus ont été soumis à une électrophorèse sur gel d'agarose à 1,5% contenant du bromide d'éthidium à 3%.

Pour la préparation du gel, 1,5g d'agarose est dissout dans 100ml de tampon TEB 0,5M. le mélange est chauffé jusqu'à ébullition dans un four à micro-onde. Le gel est refroidi à environ 45°C et 3µl de bromide d'éthidium est ajouté. Le bromide d'éthidium se fixe sur les ponts hydrogènes des acides nucléiques, ce qui rend les bandes d'ADN fluorescentes. Le mélange est ensuite coulé dans un bac à électrophorèse contenant des peignes et laissé à la température ambiante pour la solidification.

Après la solidification du gel, les peignes sont retirés. Le gel est déposé dans un bac rempli de TBE 0.5M et 12µl de mélange (2µl de tampon de migration plus 10µl de l'ADN amplifié) est logé dans chaque puits. Le tampon de migration est appelé loading buffer. Les puits extrêmes (le premier et le dernier puits) sont chargés avec un marqueur de poids moléculaire qui est la référence. Le gel est alors soumis à une tension électrique de 80mV pendant 45mn.

La lecture des bandes d'ADN se fait par révélation aux rayons UV à 460nm. Pour cela le gel est placé sur un appareil qui émet des rayons ultraviolets. Grâce à la fluorescence du bromide d'éthidium, des bandes d'ADN apparaissent sur le gel. Une photographie du gel est prise et transféré dans l'ordinateur pour interprétation.

L'INTERPRETATION DES BANDES D'ADN

Les espèces du complexe *An. gambiaes.l.* sont caractérisées par des bandes d'ADN spécifiques qui permettent de les identifier. Ainsi l'espèce *An. gambiae s.s.* a une bande de 390 paires de bases (pb) et *An arabiensis* 315pb.

Pour l'identification des formes moléculaires d'*An. gambiae s.s.*, la bande de 390pb est clivée pour donner deux bandes (110pb et 280pb) qui caractérisent la forme moléculaire (S). La forme moléculaire (M) se caractérise par la bande de 390pb, elle n'est pas digérée par l'enzyme de restriction.

LISTE DES FIGURES

Figure 1 : cycle du plasmodium, http://www.traveldocor.co.uk/malaria.htm	44
Figure 2 : Cycle de développement de l'Anophèle In Yé (2005)	47
Figure 3 : Situation du Burkina Faso	52
Figure 4 : Répartition des ethnies au Burkina Faso	53
Figure 5 : Densités des populations par provinces	54
Figure 6 : Les grands ensembles topographiques du Burkina Faso	56
Figure 7 : géomorphologie du Burkina Faso	57
Figure 8 : Profils de distribution des températures et précipitations au Burkina Faso	59
Figure 9 : Diagrammes ombrothermiques de quelques stations de références au Burkina Faso	60
Figure 10 : Climogramme de quelques stations de référence au Burkina Faso	61
Figure 11 : Débuts et fins de saison pluvieuse au Burkina Faso	61
Figure 12 : Les principaux cours d'eau et leur bassins versants au Burkina Faso	63
Figure 13 : Domaines phytogéographiques du Burkina Faso	64
Figure 14 : Formations végétales du Burkina Faso	66
Figure 15 : Diagramme d'Ishikawa des facteurs environnementaux influençant les fréquences d' <i>Anopheles gambiae</i> s.s., réalisation SOME Y. S. Corentin	79
Figure 16 : Transect, circuits et points de collecte des données entomologiques	85
Figure 17 : Modèle conceptuel de donnée	99
Figure 18 : Modèle conceptuel de traitement	103
Figure 19 : Variogramme et fonction de régression pour l'interpolation des précipitations	109
Figure 20 : Répartition des précipitations au Burkina Faso	109
Figure 21 : Fonction de régression de l'interpolation des températures moyennes annuelles	110
Figure 22 : Répartition des températures moyennes annuelles au Burkina Faso	111
Figure 23 : Variogramme et droite de régression de l'interpolation de l'évapotranspiration potentielle moyenne annuelle	112
Figure 24 : Répartition des l'évapotranspiration potentielle annuelle au Burkina Faso	112
Figure 25 : Droite de régression des pressions de vapeur moyennes annuelles au Burkina Faso	113
Figure 26 : Répartition des pressions de vapeurs moyennes annuelles au Burkina Faso	114
Figure 27 : Droite de régression de l'insolation	115
Figure 28 : Répartition de l'insolation moyenne annuelle au Burkina Faso	115
Figure 29 : Droite de régression de la vitesse des vents	116
Figure 30 : Répartition des vitesses moyennes annuelles des vents au Burkina Faso	116
Figure 31 : Droite de régression des altitudes	117

Figure 32 : Répartition des altitudes au Burkina Faso	117
Figure 33 : Répartition des densités de population au Burkina Faso en 2005	119
Figure 34 : Répartition des densités végétales au Burkina Faso	120
Figure 35 : Répartition des distances en tout point par rapport à l'hydrographie	121
Figure 36 : Répartition des distances par rapport aux localités	121
Figure 37 : Répartition des captures d' <i>Anopheles gambiae</i> s.l. au Burkina Faso de juillet à octobre 2005	124
Figure 38 : Structure spatiale de distribution d' <i>An. gambiae</i> s.s.	125
Figure 39: Structure spatiale de distribution d' <i>Anopheles gambiae</i> forme moléculaire S	128
Figure 40 : Structure spatiale de distribution d' <i>An. gambiae</i> forme moléculaire M	129
Figure 41 : configuration spatiale d' <i>Anopheles gambiae</i> s.l.	132
Figure 42 : configuration spatiale d' <i>Anopheles gambiae</i> de forme moléculaire S	133
Figure 43 : configuration spatiale d' <i>Anopheles gambiae</i> forme moléculaire M	134
Figure 44 : configuration spatiale d' <i>Anopheles gambiae</i> s.l. selon l'indice Getis et Ord	136
Figure 45 : Configuration spatiale d' <i>Anopheles gambiae</i> forme moléculaire S selon l'indice Getis et Ord	137
Figure 46 : configuration spatiale d' <i>Anopheles gambiae</i> forme moléculaire M selon l'indice Getis et Ord	138
Figure 47 : Histogramme des valeurs propres	144
Figure 48 : Représentation des variables sur les plans factoriels 1 et 2	146
Figure 49 : Représentation des variables sur les plans factoriels 1 et 3	147
Figure 50 : Représentation des variables sur le plan factoriel 1 et 4	147
Figure 51 : contribution des variables aux axes factoriels	150
Figure 52 : Positionnement de nominales supplémentaires sur les plans factoriels 1 et 2	152
Figure 53 : Positionnement de nominales supplémentaires sur les plans factoriels 1 et 3	152
Figure 54 : Positionnement de nominales supplémentaire sur les plans factoriels 1 et 4	153
Figure 55 : Histogramme des résidus centrés réduits pour <i>An. gambiae</i> forme moléculaire S	169
Figure 56 : Prédiction de la distribution spatiale des abondances relatives d' <i>An. gambiae</i> forme moléculaire S	170
Figure 57 : les zones de discontinuité dans la répartition de la forme moléculaire S et hydrographie.	171
Figure 58 : Histogramme des résidus centrés réduits pour <i>An. gambiae</i> forme moléculaire M	172
Figure 59 : Prédiction de la distribution spatiale des abondances relatives d' <i>An. gambiae</i> forme moléculaire M	173

Figure 60 : Discontinuité de la répartition de la forme moléculaire M d' <i>An. gambiae</i> s.s., Altitude et NDVI	174
Figure 61 : Corrélation entre pourcentage de S prédit par le modèle et pourcentage S observé.	176
Figure 62 : Comparaison des pourcentages observés avec les pourcentage prédits de la forme moléculaire S	177
Figure 63 : Corrélation entre pourcentages de M prédits par le modèle et pourcentages M observés	177
Figure 64 : Comparaison des pourcentages observés avec les fréquences prédites de la forme moléculaire M	178
Figure 65 : Corrélation entre pourcentages de S prédits et pourcentage de M prédits	179
Figure 66 : Structure de l'ADN ribosomal des membres du complexe <i>An. gambiae</i> (Scott et al, 1993).	194
Figure 67 : Schéma d'une unité répétée en tandem	195

LISTE DES TABLEAUX

Tableau 1 : Les sols au Burkina Faso	58
Tableau 2 : Les bassins hydrographique du Burkina Faso	62
Tableau 3 : les variables, leurs unités de référence et leurs définitions	82
Tableau 4 : Evaluation de la qualité des données collectées	89
Tableau 5 : Liste des informations	99
Tableau 6 : Liste des entités	100
Tableau 7 : Liste des associations	100
Tableau 8 : Propriété de l'entité variable	101
Tableau 9 : statistiques descriptives sur les variables quantitatives	142
Tableau 10 : Les résultats des tests de sphéricité de Bartlett	142
Tableau 11 : Matrice des corrélations entre les variables	143
Tableau 12 : les valeurs propres des axes factoriels	144
Tableau 13 : Coordonnées des variables	145
Tableau 14 : Cosinus carrés des variables	148
Tableau 15 : Contributions des variables (%)	149
Tableau 16 : Les coordonnées des variables supplémentaires	151
Tableau 17 : Statistiques descriptives sur les variables à expliquer	156
Tableau 18 : Statistiques descriptives sur les variables explicatives	157
Tableau 19 : Statistiques descriptives sur les variables qualitatives	157
Tableau 20 : Test de significativité de la quantité d'information apportée par les variables	158
Tableau 21 : Analyse du modèle (Type I SS) : test de Fisher sur les variables dans l'ordre de sélection	159
Tableau 22 : Analyse du modèle (Type III SS) : test de Fisher sur les variables quel que soit l'ordre de sélection	159
Tableau 23 : Coefficients d'ajustement du modèle	160
Tableau 24 : test de Fisher sur la significativité de la quantité de données par rapport au modèle	160
Tableau 25 : Analyse du modèle (Type I SS) : test de Fisher sur les variables dans l'ordre de sélection	160
Tableau 26 : Analyse du modèle (Type III SS) : test de Fisher sur les variables quel que soit l'ordre de sélection	161
Tableau 27 : Statistiques descriptives sur les variables à expliquer	162
Tableau 28 : Résumé pour les variables explicatives	162

Tableau 29 : Valeur du coefficient de détermination pour les 2 meilleurs modèles _____	163
Tableau 30 : Coefficients d'ajustement du modèle 2 _____	164
Tableau 31 : Le test de Fisher sur l'analyse de la significativité de la quantité d'information apportée par les variables à la régression pour <i>Anopheles gambiae</i> forme moléculaire S _____	164
Tableau 32 : Test de Fisher de type I SS _____	164
Tableau 33 : Analyse du modèle (Type III SS) _____	165
Tableau 34 : Paramètres du modèle de distribution d' <i>An. gambiae</i> forme moléculaire S _____	165
Tableau 35 : Valeur du coefficient de détermination pour les 2 meilleurs modèles _____	166
Tableau 36 : les coefficients d'ajustement pour le modèle 3. _____	166
Tableau 37 : Evaluation de la valeur de l'information apportée par les variables ($H_0 = Y = \text{Moy}(Y)$) _____	166
Tableau 38 : Analyse du modèle (Type I SS) _____	167
Tableau 39 : Analyse du modèle (Type III SS) _____	167
Tableau 40 : Paramètres du modèle _____	167
Tableau 41 : Liste des localités ayant des résidus anormaux pour <i>An. gambiae</i> forme moléculaire S _____	169
Tableau 42 : Liste des localités ayant des résidus anormaux pour <i>An. gambiae</i> forme moléculaire M _____	173
Tableau 43 : Caractéristiques des données utilisées pour le test de Kolmogorov-Smirnov _____	176
Tableau 44 : Test de Kolmogorov-Smirnov pour la forme moléculaire S _____	176
Tableau 45 : Caractéristiques des données test de Kolmogorov-Smirnov pour la forme moléculaire M _____	178
Tableau 46 : Test de Kolmogorov-Smirnov pour la forme moléculaire M _____	178

TABLE DES MATIERES

<i>Liste des abréviations</i>	3
<i>Dédicace</i>	5
<i>REMERCIEMENTS</i>	6
<i>INTRODUCTION GENERALE</i>	9
<i>PREMIERE PARTIE :CADRAGE THEMATIQUE, DEMARCHE SCIENTIFIQUE ET METHODE</i>	15
<i>Chapitre I : Etat de l’art, définition du champ conceptuel et positionnement scientifique</i>	17
I.1. Géographie et étude des relations entre environnement et sante	17
I.1.1 Géographie et lutte contre les maladies	18
I.1.1.1 L’objet de la géographie : l’espace	18
I.1.1.2 Relation entre espace et lutte contre les maladies	19
I.1.2 Contribution de la géographie à la lutte contre les maladies	21
I.1.2.1 L’espace dans l’analyse géographique	22
I.1.2.2 Analyse spatiale et analyse des problèmes de santé	23
I.1.2.3 Techniques géographiques et étude des problèmes de santé	24
I.2. définition du champ conceptuel et positionnement scientifique	27
I.2.1 Modélisation : definition et type	28
I.2.2 L’Analyse spatiale	30
I.2.3 Le Système d’Information Géographique (SIG).	31
I.2.4 Paradigme biogéographique : une matrice conceptuelle pour la construction du projet de modélisation	31
I.2.4.1 Objectifs de recherche	32
I.2.4.2 Paradigme biogéographique et conceptualisation de la modélisation de la distribution spatiale	33
I.2.4.3 Les hypothèses de la recherche	36
I.2.4.4 Les indicateurs retenus pour la vérification des hypothèses de recherche	37
<i>Chapitre II : Bio-écologie d’An. gambiae au Burkina Faso : une relation à formaliser</i>	40
II.1 Le paludisme une maladie de santé publique dont l’anophèle est le vecteur	40
II.1.1 Le paludisme dans le monde et au Burkina Faso	40
II.1.2 La lutte contre le paludisme	41
II.1.2.1 La lutte contre le Parasite	42
II.1.2.2 La lutte contre les vecteurs	42

II.1.2.3	Les perspectives de la lutte contre le paludisme	43
II.1.3	Le cycle de transmission du paludisme	43
II.2	Bio-écologie d' <i>Anopheles gambiae</i>	44
II.2.1	La position zoologique	44
II.2.2	Présentation du complexe <i>Anopheles gambiae</i>	45
II.2.3	Le cycle de développement des Anophèles	46
II.2.3.1	La phase de vie aquatique	47
II.2.3.2	La phase de vie aérienne	48
II.2.4	L'écologie des adultes	48
II.2.5	L'écologie des larves (stades pré-imaginaux) : Caractéristiques des habitats larvaires	49
II.2.6	L'environnement du complexe <i>An. gambiae</i>	50
II.3	Présentation de la zone d'étude	52
II.3.1	La population	53
II.3.2	Le relief	55
II.3.3	Les sols	57
II.3.4	Le climat	58
II.3.5	L'hydrographie	62
II.3.6	La végétation	64
Chapitre III : Approche méthodologique de l'étude		68
III.1	Aperçu des méthodes de modélisation de la distribution spatiale des espèces vivantes utilisant l'analyse spatiale et les SIG	68
III.1.1	L'impact des usages des modèles sur la modélisation	69
III.1.1.1	Les usages des modèles pour l'élaboration des politiques et définition de stratégies	69
III.1.1.2	L'impact des usages des modèles pour l'action	70
III.1.2	Les modèles classés d'après les méthodes	71
III.1.2.1	Les modèles rhétoriques	72
III.1.2.2	Les modèles mathématiques	72
III.1.2.3	Les modèles iconiques	75
III.2	La démarche, méthodes et matériels	77
III.2.1	Démarche d'ensemble de la recherche	77
III.2.2	Univers des variables et entités / objets géographiques	78

III.2.3	L'inventaire des variables	80
III.2.4	Entites / objets géographiques	83
III.2.5	Les méthodes de collecte de données	84
III. 2.5.1	Le transect	84
III. 2.5.2	La revue de littérature scientifique	85
III.2.6	Évaluation de la qualité des données	86
III.2.6.1	Méthode d'évaluation utilisée	86
III.2.6.2	Terrain nominal	87
III.2.6.3	Appréciation de la qualité des données	88
III.2.7	Les méthodes de traitement des données	90
III.2.7.1	les traitements de type spatial	90
III.2.7.2	Les traitements de type temporel	92
III.2.8	Matériels / logiciels utilisés	93
<i>Conclusion de la première partie</i>		95
<i>DEUXIEME PARTIE : STRUCTURATION DES DONNEES, ELABORATION ET EVALUATION DES MODELES</i>		96
<i>Chapitre IV : Organisation et structuration de la base de données</i>		98
IV.1.1	L'entité « catégorie »	100
IV.1.2	L'entité « Sites »	100
IV.1.3	L'entité « Période »	101
IV.1.4	L'entité « Variables »	101
IV.2.1	Le processus de collecte et de traitement des données entomologiques	104
IV.2.2	Le processus de collecte et de traitement des données environnementales	105
IV.2.3	Le processus de modélisation et de réalisation de carte de distribution prédictive	106
IV.3.1	Interpolation des données	107
IV.3.1.1	Les précipitations : des données sources aux données interpolées	108
IV.3.1.2	Les températures : des données sources aux données interpolées	110
IV.3.1.3	L'Évapotranspiration potentielle : des données sources aux données interpolées	111
IV.3.1.4	La pression de vapeur : des données sources aux données interpolées	113
IV.3.1.5	L'insolation : des données sources aux données interpolées	114
IV.3.1.6	La vitesse du vent : des données sources aux données interpolées	115
IV.3.1.7	L'altitude: des données sources aux données interpolées	116

IV.3.2	Estimation des données	117
IV.3.3	Les données classifiées	122
<i>Chapitre V : Distribution spatiale des membres du complexe An. gambiae : structures globales et configurations locales</i>		123
V.1.1	les statistiques générales sur les données entomologiques collectées.	124
V.1.2	la structure spatiale d' <i>Anopheles gambiae</i> s.l.	125
V.1.3	La structure spatiale d' <i>Anopheles gambiae</i> forme moléculaire S	127
V.1.4	La structure spatiale d' <i>Anopheles gambiae</i> forme moléculaire M	129
V.2.1	Analyse des configurations spatiales locales d' <i>Anopheles gambiae</i> s.l.	131
V.2.2	Analyse des configurations spatiales locales d' <i>Anopheles gambiae</i> forme moléculaire S	132
V.2.3	Analyse des configurations spatiales locales d' <i>Anopheles gambiae</i> forme moléculaire M	134
V.2.4	Configurations spatiales selon l'indice GI* de Getis-Ord pour <i>Anopheles gambiae</i> s.l.	135
V.2.5	Configurations spatiales selon l'indice GI* de Getis-Ord pour <i>Anopheles gambiae</i> forme moléculaire S	136
V.2.6	Configurations spatiales selon l'indice GI* de Getis-Ord pour <i>Anopheles gambiae</i> forme moléculaire M	138
V.3.1	Analyse des corrélations entre la distribution des vecteurs et les paramètres environnementaux	140
V.3.2	Caractérisation des données et paramètres de l'analyse	141
V.3.3	Résultats et interprétation de l'analyse en composante Principale	142
V.3.3.1	Analyse des corrélations	143
V.3.3.2	Analyse des valeurs propres	144
V.3.3.3	Analyse de coordonnées des variables	145
V.3.3.4	Analyse de la qualité de la représentation des variables par rapport aux différents axes factoriels	148
V.3.3.5	Analyse de la contribution des variables à la formation des axes factoriels	149
V.3.3.6	Contribution des variables supplémentaires aux différents axes factoriels	151
<i>Chapitre VI : Modélisation de la distribution spatiale</i>		156
VI.1	Modélisation par analyse de covariance (ANCOVA)	156
	Configuration de l'ANCOVA et description des données	156
	Résultats et interprétation de l'ANCOVA	157

VI.2	Modélisation par regression multiple	162
VI.2.1	Configuration de la régression et description des données	162
VI.2.2	Résultats de la régression	163
VI.2.2.1	Modèle de distribution d' <i>An. gambiae</i> forme moléculaire S	163
VI.2.2.2	Le modèle de distribution d' <i>An. gambiae</i> forme moléculaire M	165
VI.3	Evaluation des modèles	168
VI.3.1	Analyse des residus	168
VI.3.1.1	Analyse des résidus pour <i>An. gambiae</i> forme moléculaire S	168
VI.3.1.2	Analyse des résidus pour <i>An. gambiae</i> forme moléculaire M	172
VI.3.2	Confrontation des données des modèles aux données empiriques	175
VI.3.2.1	Confrontation du modèle d' <i>An. gambiae</i> forme moléculaire S à la vérité terrain	175
VI.3.2.2	Confrontation du modèle d' <i>An. gambiae</i> forme moléculaire M à la vérité terrain	177
Conclusion du chapitre VI		179
CONCLUSION GENERALE		181
Bibliographie		184
ANNEXES		191
ANNEXE 1 : Méthodes de collecte des moustiques sur le terrain		192
ANNEXE 2 : Identification des formes moleculaires		194
Liste des figures		199
Liste des tableaux		202

RESUME

MODELISATION DE LA DISTRIBUTION SPATIALE DES FORMES MOLECULAIRES M ET S D'*Anopheles gambiae* sensu stricto AU BURKINA FASO AVEC LES SIG ET L'ANALYSE SPATIALE

La lutte anti-vectorielle est une composante importante de la lutte contre les maladies à transmission vectorielle. La connaissance des populations vectorielles, tant dans leur composition spécifique que dans leur répartition spatio-temporelle, est fondamentale pour la conception des stratégies de lutte contre ce type de maladie.

Cette thèse a modélisé la distribution spatiale des formes moléculaires M et S d'*Anopheles gambiae* s.s., des vecteurs majeurs du paludisme au Burkina Faso. La modélisation a été faite à partir de l'analyse d'une série d'observations portant à la fois sur les vecteurs et l'environnement. Elle procède d'une combinaison de concepts et méthodes de biogéographie avec des techniques et outils d'analyse spatiale, d'analyse des données et des systèmes d'informations géographiques.

Nous retenons de ce travail de recherche que l'abondance de la forme moléculaire S d'*Anopheles gambiae* s.s., diminue au fur et à mesure que l'on évolue des régions humides du sud et du sud-ouest vers celles les plus arides du nord et du nord-est. Le NDVI, l'ETP, et l'insolation sont les facteurs les plus déterminants de sa distribution spatiale. Par contre, l'abondance de la forme moléculaire M d'*Anopheles gambiae* s.s. augmente des régions humides du sud et du sud-ouest vers celles plus arides du nord et du nord-est. L'altitude, la pression, l'insolation, la densité de végétation sont les facteurs les plus déterminants de cette répartition spatiale. De ces résultats, ont été dérivés deux modèles qui ont servi à l'élaboration des cartes de distribution des formes moléculaires M et S d'*Anopheles gambiae* s.s.

Enfin, cette thèse révèle le rôle de l'approche géographique dans la réflexion sur les questions de santé et sa méthodologie pourrait être testée sur d'autres sites et pour d'autres vecteurs de maladies.

Elle peut s'enrichir d'une analyse multi échelle et d'une modélisation de la variabilité temporelle.

Mots Clés : *Anopheles gambiae*, Modélisation, Analyse spatiale, SIG, Burkina Faso

ABSTRACT

MODELLING THE SPATIAL DISTRIBUTION OF THE MOLECULAR FORM M AND S OF *Anopheles gambiae* s.s. IN BURKINA FASO WITH GIS AND SPATIAL ANALYST

The anti-vectorial fight is an important constituent for vectorial borne diseases control. The knowledge of the vectors populations as well as their specific composition and spatiotemporal distribution is fundamental for the conception of the strategies of fight against this type of disease.

This thesis modeled the spatial distribution of the molecular forms M and S of *Anopheles gambiae* s.s., major vectors of malaria in Burkina Faso. The modeling was made from the analysis of a series of observations concerning both the vectors and the environment.

It proceeds of a combination of concepts and methods of biogeography with techniques and tools of spatial analysis, data analysis and the geographical information systems.

We retain of this research work that the abundance of the molecular form S of *Anopheles gambiae* s.s., decreases as we move from wet regions of the south and the southwest to those the driest of the north and the northeast. The NDVI, the ETP, and the sunshine are the most determining factors of its spatial distribution.

In contrast, the abundance of the molecular form M of *Anopheles gambiae* s.s. increases from the wet regions of the South and the southwest to those drier of the north and the northeast. The temperatures, the pressure, the sunshine and the NDVI are the most determining factors of this spatial distribution.

Of these results, were diverted two models which we use to elaborate the Maps of distribution of the molecular forms M and S of *Anopheles gambiae* s.s.

This thesis highlights also the role of the geographical approach in the reflection on health issues and its methodology could be tested on the other sites and for the other vectors of diseases. This methodology could be improved by multi-scale analysis and of temporal variability modeling.

Key Words: *Anopheles gambiae*, Modeling, spatial analyst, GIS, Burkina Faso



Centre d'Etudes pour le Développement des Territoires et l'Environnement
École Doctorale Sciences de l'Homme et de la Société
UFR Lettres, Langues et Sciences Humaines
10 Rue de Tours, BP 46 527
45065 Orléans CEDEX 2

