



Master 2 Informatique  
Spécialité Géomatique

---

# **Conceptualisation et développement d'outils pour cartographier des indicateurs relatifs à divers risques environnementaux**

**Cas d'application : Le paludisme**

---

Gilles ENTRINGER

Stage réalisé du 06.02.2012 au 06.08.2012 à l'IRD sous la direction de  
Bertrand GUERRERO.

Tuteur pédagogique : Thérèse LIBOUREL

Date de soutenance : 25.09.2012

Membres du Jury : Thérèse LIBOUREL,

# Remerciements

Je tiens tout d'abord à remercier mon maître de stage Bertrand Guerrero qui m'a permis de réaliser mon stage dans les meilleures conditions possibles. Je le remercie pour ses conseils, le temps qu'il m'a consacré et pour ses commentaires sur mon travail.

Je remercie également mes cotuteurs Nadine Dessay, Maud Loireau et Vincent Herbreteau pour leur disponibilité, leurs conseils et leur aide tout au long de mon stage.

Merci aussi à Christelle Pierkot, Jean-Christophe Desconnets, Samuel Andres et tous les membres de l'équipe SIC de l'UMR Espace-Dev pour leurs conseils techniques et pour leur patience.

Je tiens aussi à remercier Mireille Fargette et tous les membres de l'équipe AIMS de l'UMR Espace-Dev qui sont intervenus tout au long de l'élaboration de mon travail de stage.

Je voudrais remercier tout particulièrement mon tuteur pédagogique Thérèse Libourel de m'avoir aidé à trouver ce stage et à découvrir le monde de la recherche et pour ses conseils et pour le temps qu'elle m'a consacré.

Enfin, je tiens à remercier le personnel administratif de l'IRD, de l'UM2 et du LIRMM pour le temps qu'ils m'ont consacré.

# Sommaire

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Contexte et problématique</b>	<b>3</b>
2.1	Principes généraux . . . . .	3
2.2	Le logiciel SIEL . . . . .	3
2.2.1	Architecture SIEL . . . . .	4
2.2.2	La chaîne de traitements du SIEL . . . . .	4
2.3	Problématiques du stage . . . . .	4
<b>3</b>	<b>Contexte thématique</b>	<b>6</b>
3.1	Interaction environnement-santé . . . . .	6
3.1.1	Définitions . . . . .	6
3.1.2	Facteurs de risque de transmission du paludisme . . . . .	8
3.2	Approches logicielles existantes . . . . .	13
3.2.1	Cas d'utilisation concret dans le contexte environnement-santé : . . . .	14
3.2.2	Conclusion . . . . .	15
<b>4</b>	<b>Méthodologie</b>	<b>16</b>
4.1	Analyse . . . . .	16
4.1.1	Données disponibles . . . . .	18
4.1.2	Modèle simplifié . . . . .	18
4.2	Raisonnement . . . . .	20
4.2.1	Description des données et des traitements . . . . .	20
4.2.2	Les catégories de données et de traitements . . . . .	23
4.2.3	Conceptualisation de la chaîne de traitements . . . . .	24
4.3	Conception . . . . .	27
4.3.1	Architecture informatique de la chaîne de traitements et du logiciel ouvert	28
4.4	Opérationnalisation . . . . .	29
4.4.1	Fonctionnement chaîne de traitements "fermée" . . . . .	29
4.4.2	Fonctionnement du logiciel "ouvert" . . . . .	35
<b>5</b>	<b>Résultats</b>	<b>39</b>
5.1	Résultat . . . . .	39
5.2	Validation des résultats . . . . .	39
5.3	Difficultés rencontrés . . . . .	39
5.3.1	Java . . . . .	39
5.3.2	PostgreSQL/PostGIS . . . . .	40
5.3.3	R . . . . .	40
5.4	Perspectives . . . . .	41
5.5	Conclusion . . . . .	42

<b>6</b>	<b>Discussion</b>	<b>43</b>
<b>7</b>	<b>Glossaire et définitions</b>	<b>44</b>
<b>8</b>	<b>Annexes</b>	<b>45</b>
8.0.1	Définitions, librairies et bibliothèques utilisées . . . . .	45
8.0.2	Les traitements effectués . . . . .	48
8.0.3	Configuration prérequis pour l'utilisation des outils . . . . .	56

# Chapitre 1

## Introduction

Le stage s'inscrit dans la formation du Master 2 Géomatique et s'est déroulé au sein de deux équipes de l'UMR Espace-Dev <sup>1</sup> : l'équipe SIC et l'équipe AIMS.

Les objectifs de l'UMR Espace-Dev sont multiples. L'UMR s'inscrit dans une perspective de développement durable des territoires et propose des méthodologies de spatialisation des dynamiques de l'environnement. L'UMR développe et exploite également un réseau de stations de réception d'images satellites d'observation de la terre.

L'UMR se regroupe en trois équipes de recherche :

- Equipe OSE (Observation spatiale de l'environnement), spécialisée dans la télédétection et les images satellitaires
- Equipe AIMS (Approche intégrée des milieux et des sociétés) qui est spécialisée dans le domaine de l'environnement, de la télédétection et dans l'élaboration de dynamiques socio-environnementales ou d'indicateurs et de modèles des interactions milieux/sociétés.
- Equipe SIC (Systèmes d'information et de connaissances) qui a pour objectif l'acquisition, la gestion, la représentation et le partage des données et des connaissances. D'autres objectifs de l'équipe sont la modélisation de dynamiques spatio-temporelles, la visualisation, la cartographie sémantique et l'aide à la décision

Tout scientifique qui travaille à établir des cartes de risque fonde ses recherches sur un raisonnement qui consiste à enchaîner des procédures, traitements ou algorithmes.

Depuis plusieurs années au sein de l'UMR Espace-Dev (équipes SIC et AIMS) est développé le SIEL (Système d'Information sur l'Environnement l'Echelle Locale). Le SIEL est un logiciel d'aide à la décision dans la gestion de l'environnement. Il permet notamment d'évaluer le risque de dégradation de la végétation en milieu aride et de produire des **indices environnementaux spatialisés** sous la forme de cartes. L'objectif du SIEL est de mettre à la disposition des scientifiques des outils informatiques pour automatiser au mieux leur démarche.

L'objectif du présent stage est d'élargir les potentialités du SIEL. En définissant dans premier temps un nouveau contexte, celui dénommé environnement-santé et plus précisément celui relatif au paludisme. Le développement d'un outil autorisant la mise en oeuvre de chaînes de traitements dédiées à l'évaluation de risques environnementaux est donc un objectifs poursuivis. Basé sur une plateforme Open Source, l'outil devrait permettre à long terme un fonctionnement plus ouvert, plus facile et plus adapté à diverses problématiques. Cet outil sera aussi éprouvé et

---

1. <http://www.espace.ird.fr/index.php>

validé dans le contexte environnement-santé pour ce stage.

Le mémoire s'organise de la façon suivante :

- Une première partie sera dédiée à la présentation du contexte du stage. Je présenterai les problématiques liées à la construction des indicateurs, à la définition des facteurs de risque et au développement d'une chaîne de traitements. Dans cette partie sera également présenté plus en détail le SIEL.
- Dans la deuxième partie, le contexte thématique du stage sera expliqué. A partir d'une recherche bibliographique les principes généraux liés à la problématique environnement-santé ainsi que des approches logicielles existantes seront présentées.
- Une troisième partie présentera la méthodologie de mon travail. Un modèle conceptuel du cycle du paludisme sera présenté. A partir de ce modèle, un raisonnement sur les données et les traitements permet de définir les données et les traitements nécessaires pour l'élaboration et le développement de la chaîne de traitements respectivement du logiciel ouvert. Les différentes étapes de l'élaboration de la chaîne de traitements seront également proposées en commençant avec la conceptualisation des traitements et des données utilisés pour arriver à la présentation de l'implémentation et de l'architecture de la chaîne de traitements automatisée et du logiciel ouvert.
- Dans la dernière partie du mémoire seront présentés les résultats, les difficultés rencontrées et les perspectives de mon travail.

Par la suite, les termes en **gras** seront définis dans le glossaire en fin du mémoire.

# Chapitre 2

## Contexte et problématique

### 2.1 Principes généraux

A l'heure actuelle, le logiciel SIEL permet d'effectuer un certain nombre de traitements bien définis dans le contexte de la désertification et risque de dégradation. Le contexte environnement-santé est également une des préoccupations de l'UMR Espace-Dev et la réalisation de cet outil nous a paru intéressante. Afin de permettre la réutilisabilité, il faut penser les chaînes de traitements dans ce nouveau contexte.

A l'heure actuelle, le SIEL est un plugin ArcGIS qui fonctionne avec des séries de traitements et d'opérations ciblées sur le type d'indices à produire. L'objectif à long terme est de retravailler ce logiciel pour qu'il devienne plus modulaire et générique, de préférence dans un contexte "Open Source". Nous avons donc décidé d'ouvrir l'outil et à repenser son architecture dans un contexte Open Source.

Dans ce contexte, le développement d'une chaîne de traitements indépendante avait comme objectif de servir comme base pour le futur développement d'une nouvelle architecture informatique. Le développement d'un **logiciel ouvert** a permis de franchir une étape supplémentaire dans le développement informatique du SIEL. De plus, l'intégration de la création automatisée d'indicateurs dans le domaine environnement-santé est une première approche dans l'objectif d'intégrer de nouvelles problématiques dans le SIEL comme par exemple la déforestation.

Nous allons maintenant présenter brièvement le SIEL, ses principes, concepts et son architecture informatique. Par la suite les problématiques du présent stage seront expliquées en détail.

### 2.2 Le logiciel SIEL

Le SIEL est un modèle environnemental instrumenté (Loireau *et al.*, 2007) qui spatialise les pratiques d'exploitation des ressources sur un territoire, modélise les paysages et évalue les risques de dégradation des ressources. (Loireau Maud, 2012)

Le SIEL est un outil conçu pour répondre aux besoins des scientifiques et décideurs concernés par la lutte contre la dégradation et la gestion des ressources naturelles. Le SIEL est capable d'anticiper les risques et de les suivre, ce qui est important notamment dans le cadre de la mise en place d'observatoires. Cet outil donne également la capacité aux scientifiques et aux décideurs de mesurer l'impact des opérations déjà effectuées sur un territoire et d'optimiser les

actions futures notamment en terme de gestion des ressources.

Le logiciel a été conceptualisé à partir de 1993 par Maud Loireau dans le cadre de sa thèse. En même temps un premier prototype a été développé. En 2000, dans le cadre de la lutte contre la désertification, la démarche scientifique et le modèle général du SIEL ont été adoptés par le programme ROSELT (Réseau d'observatoires de Surveillance Ecologique à Long Terme) de l'OSS (Observatoire du Sahara et du Sahel). Ceci a permis de développer un premier logiciel SIEL qui couvrait l'ensemble de la chaîne de traitements proposée. Une première version du logiciel traitant la ressource « végétation » a pu être déployée dans le réseau ROSELT/OSS dès la fin de 2003. A partir de 2006, les différents acteurs ont poursuivi les développements du logiciel. Ceci a débouché sur un logiciel opérationnel et prêt à être diffusé, une documentation (guide d'utilisateur, aide en ligne), une documentation technique (guide développeur, dictionnaire des données), une base de données exemple et des supports de formations. Actuellement, l'objectif principal reste le développement informatique du logiciel et le responsable du projet informatique est depuis 2010 Bertrand Guerrero.

### **2.2.1 Architecture SIEL**

Le SIEL se présente sous forme d'un plugin ArcGIS et se base sur deux modules propriétaires :

- Module données (Logiciel Microsoft Access, langage de programmation : Visual Basic for Application (VBA)) pour la gestion et le stockage des tables attributaires. Gestion des **SGBD** (format .mdb).
- Module SIG (Logiciel : ArcGIS Desktop avec l'extension Spatial Analyst, langage de programmation : Visual Basic 6.0 et langage ESRI ArcObjects) pour la gestion et le stockage des couches géographiques et les géotraitements. Même format que pour le module "données" pour la gestion de la base de données.

Le système d'exploitation supporté est Windows XP, les logiciels supportés sont "Microsoft Office 2003" et "Microsoft Office 2007" pour le module données et "ArcGIS 9.1" et "ArcGIS 9.3" pour le module SIG. Les deux modules sont donc basés sur des outils propriétaires. Le SIEL a été développé en Visual Basic, langage informatique qui n'est plus supporté par les nouvelles versions d'ArcGIS ce qui cause un certain nombre de restriction par rapport à son utilisation.

### **2.2.2 La chaîne de traitements du SIEL**

Les deux modules (Données et SIG), en faisant différents traitements, créent des indices environnementaux spatialisés qui permettent la cartographie des risques de dégradation des ressources. Les données utilisés et les résultats sont stockés dans une base de données.

Schéma chaîne SIEL (à revoir Bertrand)

## **2.3 Problématiques du stage**

Étant donné la pluridisciplinarité du sujet de stage, les problématiques se regroupent en 2 catégories :

- Problématiques environnement-santé : Paludisme et les facteurs de risque de transmission



- Problématiques informatiques : Conceptualisation des chaînes de traitements dans le contexte environnement-santé, choix de l'architecture et implémentation des outils (chaîne de traitements et logiciel ouvert).

Concernant l'environnement-santé, la problématique principale est la compréhension du phénomène, la définition des facteurs de risque et la formalisation des éléments nécessaires à la mise en œuvre de la chaîne de traitements.

Au niveau informatique les points à traiter relèvent du choix de l'architecture informatique et de la réalisation d'un prototype de chaîne de traitements. Les principes de réutilisation et de généricité nous amènent à réfléchir sur une architecture "boîte blanche", d'ouvrir la boîte de la chaîne de traitements et de déstructurer la chaîne en boîtes élémentaires réutilisables. Ce logiciel ouvert permet d'exécuter chaque traitement de la chaîne indépendamment ainsi que de les combiner selon les besoins de l'utilisateur.

En résumé une problématique générale du stage se dégage :

**Conceptualisation et développement d'outils pour cartographier des indicateurs relatifs à divers risques environnementaux.**

# Chapitre 3

## Contexte thématique

Sans être exhaustifs, nous relatons ici, les principes relatifs à la problématique environnement-santé appliquée au paludisme et les logiciels et chaînes de traitements existants traitants des problématiques similaires au sujet d'étude.

### 3.1 Interaction environnement-santé

Je me suis focalisé dans un premier temps sur la définition des facteurs de risque de transmission dans le contexte des interaction environnement-santé et plus précisément de la maladie du paludisme. Dans une première partie les termes pertinents pour ce travail sont définis. Par la suite sont expliqués de façon détaillée les facteurs de risques de transmission identifiés à partir d'une recherche bibliographique.

#### 3.1.1 Définitions

##### Approche environnement-santé

L'approche environnement-santé s'intéresse à l'influence de la qualité de l'environnement physique, chimique et biologique sur la santé des hommes et des animaux d'un point de vue spatial et dynamique. Il s'agit donc d'une science dont les frontières sont extrêmement difficiles à délimiter tant les domaines couverts sont potentiellement vastes et susceptibles d'interférer les uns avec les autres (Santé, 2006).

##### Paludisme

Le paludisme, appelé également malaria, est une parasitose due à un protozoaire transmis par la piqure de la femelle d'un moustique (**anophèle**), provoquant des fièvres intermittentes. Le paludisme est la maladie vectorielle la plus commune dans le monde avec une estimation de 216 millions de cas et 665000 décès en 2010 (Organization, 2012) principalement dans les régions tropicales et en Afrique sub-saharienne. Ces chiffres seraient bien inférieurs à l'étendue réelle de cette maladie, dont la mortalité est tout de même observée à la baisse grâce à l'action positive des programmes de contrôle (Murray et al, 2012).

Le médecin français Alphonse Laveran a découvert la cause de la maladie en 1880 à Constantine (Algérie). Le moustique anophèle se reproduit dans les zones marécageuses. Le parasite qui possède plusieurs hôtes intermédiaires, dans l'état endémique, infecte les cellules hépatiques de la victime puis circule dans le sang. Au cours de son cycle de vie, le parasite à l'intérieur de l'organisme humain, fait un certain nombre de transformations qui lui permettent d'échapper au système immunitaire humain. Au final, lorsqu'un moustique non-infecté pique une personne contaminée, le parasite est également transmis de l'homme au moustique.

En France et dans les autres pays développés, le paludisme a disparu depuis les années 1960. Malgré tout, depuis les années quatre-vingt les nombreux voyages effectués dans les pays où il est endémique, notamment dans les régions tropicales et subtropicales ont causé une réapparition de la maladie dans les pays développés.

Actuellement, le paludisme est responsable de plus de 300 millions de cas de maladie aiguë et d'au moins un million de décès dans le monde. Quatre-vingt-dix pour cent des décès dus au paludisme surviennent en Afrique, au Sud du Sahara, principalement chez les jeunes enfants. De nos jours, aucun vaccin efficace n'a encore pu être développé et les scientifiques doutent de plus en plus qu'ils trouveront un jour la solution miracle contre cette maladie (Faire reculer le paludisme, 2010).

## **Vulnérabilité**

La vulnérabilité dans le domaine de l'environnement-santé est la probabilité qu'une personne soit affectée (dans notre cas d'étude de la source d'une maladie) en fonction de sa susceptibilité aux effets de l'aléa et du niveau d'exposition. La précarité sociale et la vulnérabilité médicale sont étroitement associées et s'additionnent souvent (Picheral, 2001).

## **Aléa**

Pour définir l'aléa pour le risque sanitaire, nous nous appuyons sur les travaux de (C. Aschan-Leygonie, 2009). Ces auteurs définissent l'aléa comme une menace d'origine naturelle ou humaine sur un système et distinguent deux types d'aléas. Les perturbations sont des événements ponctuels, repérables dans le temps, dont l'ampleur dépasse la variabilité habituelle du phénomène. Le « stress » est un autre type d'aléa qui exerce une pression continue sur le système, mais dont la variabilité est limitée.

## **Risque**

Le risque peut être défini comme la combinaison entre l'aléa et la vulnérabilité. Le risque est la probabilité, aléatoire ou non d'un événement qui menace la santé ou met en danger la vie d'un individu ou d'une population. Le risque dépend de la capacité d'une population ou d'un système de faire face à des menaces (aléas). On peut distinguer des risques de nature différente : risques génétiques, risques naturels, risques anthropiques ou risques technologiques.

En France et dans les pays de développement le paludisme a disparu parce que les responsables ont été capables de diminuer au minimum la vulnérabilité de la population et leur exposition au risque (en éradiquant les moustiques par exemple).

D'autres notions associées au risque doivent être prises en compte dans l'analyse d'un risque sanitaire. La perception d'un risque donnée varie d'une personne à une autre, d'une société à une autre et conduit à des comportements protecteurs différents. Aussi l'enjeu du risque varie selon personnes : être malade peut avoir des conséquences sociales et économiques graves pour un foyer, tout comme pour une société.

## **Facteur de risque**

Un facteur de risque est la caractéristiques individuelle ou collective, endogène ou exogène, augmentant de façon statistiquement significative la probabilité d'apparition et de développement d'une maladie. Un facteur de risque n'est donc pas une "cause". Il est également important de différencier facteur de risque et marqueur de risque (par exemple âge, sexe, groupe sanguin etc) car tout facteur de risque peut être contrôlé ou supprimé.

Dans le cadre de cette étude, les facteurs de risques sont l'ensemble des éléments qui augmentent la probabilité que la maladie (le paludisme) se développe sur un territoire. Un facteur de risque concerne donc aussi bien les aléas que la vulnérabilité. L'ampleur d'un risque, sa fréquence, sa durée, son aire d'extension et de diffusion éventuelle (espace à risque) sont fonction de facteurs de risque.

## **Indicateur**

Selon la norme ISO 8402, un indicateur est une "information choisie, associée à un phénomène, destinée à en observer périodiquement les évolutions au regard d'objectifs périodiquement définis". Un indicateur est donc une variable qui décrit un élément d'une situation ou une évolution d'un point de vue quantitatif. Un indicateur est un outil d'aide à la décision et n'a d'intérêt que par les choix qu'il aide à faire dans ce cadre (Santé, 2002).

## **Carte de risque**

Une carte de risque permet de visualiser les zones de risque principales face à un "risque" spécifique. Une carte de risque est généralement obtenue en combinaison des cartes de vulnérabilité (par exemple densité de population) et des cartes d'aléas (par exemple proximité d'une surface d'eau ou de végétation dans le cas du paludisme). Ces cartes permettent, dans le futur, d'aménager le territoire de façon plus adaptée au risque et de savoir par exemple, quels quartiers d'une ville sont particulièrement exposés à un risque. Une carte de risque permet de présenter de façon synthétique les risques. De telles cartes doivent être prudemment interprétées et utilisées et ne pas être distribuées sans explications. La façon de présenter et de discrétiser les données dans une carte peut conduire à de mauvaises interprétations.

### **3.1.2 Facteurs de risque de transmission du paludisme**

En me basant sur une recherche bibliographique, j'ai élaboré une liste des différents facteurs causant le développement et la transmission du paludisme respectivement qui sont susceptibles d'agrandir la vulnérabilité des habitants sur un territoire. Les facteurs peuvent être les mêmes que pour d'autres maladies climato-dépendantes. Les facteurs sont regroupés en 3 catégories : Les facteurs liés à l'environnement, les facteurs biologiques et les facteurs humains. Les facteurs jouent un rôle plus ou moins important dans le développement du paludisme. Pour chaque catégorie un tableau présentera les différents facteurs. Chaque facteur sera par la suite expliqué en détail afin d'illustrer les caractéristiques spécifiques à chaque facteur de risque.

#### **Facteurs environnementaux**

Les facteurs environnementaux sont en lien direct avec par exemple les conditions météorologiques ou avec la morphologie des sols sur un territoire. Toutes ces conditions vont influencer le risque que le paludisme puisse être transmis sur un territoire.

Numéro	Facteur
1	Précipitation
2	Température
3	Distance au point d'eau le plus proche
4	Caractéristiques du point d'eau
5	Altitude
6	Humidité
7	Humidité du sol
8	Température mois précédant les précipitations
9	Végétation

## Explications

**Précipitation :** Les précipitations sont parmi les facteurs de risque les plus importants du paludisme puisqu'elles conditionnent l'écologie des moustiques, vecteurs des parasites. En examinant des territoires avec des cas de paludisme et des territoires sans cas de paludisme, il faut qu'il pleuve pendant 3 à 5 mois au moins 80 mm par mois pour que le paludisme puisse se développer (Adjuik *et al.*, 1998). Les précipitations créent également des points d'eau temporaires, souvent difficilement repérables et idéaux pour le développement des moustiques. Ces points d'eau temporaires sont généralement de bonne qualité (non-pollués), ce qui est un autre facteur important.

**Température :** Le pourcentage de survie des moustiques sur un territoire est étroitement lié à la température (Ermert *et al.*, 2011). Les moustiques disparaissent à partir d'une température de moins de 5°C et ne supportent pas des températures supérieures à 40°C. D'autres expériences montrent également que les moustiques ne peuvent pas survivre les 56 jours de leur cycle de reproduction normal si les températures sont inférieures à 16°C et/ou supérieures à 32°C. La température idéale pour le cycle de reproduction du moustique est de 22°C. Dans ce cas-là, le cycle est fini après seulement 22 jours. (Adjuik *et al.*, 1998)

**Distance point eau** Les moustiques anophèles peuvent se déplacer d'un maximum de 7 kilomètres par rapport à un point d'eau (Ermert *et al.*, 2011). Les hommes habitant à plus de 7 kilomètres d'un point d'eau (de bonne qualité) ne sont donc, en théorie, exposés à aucun risque de paludisme. Néanmoins il est plutôt rare, notamment dans les pays en développement et donc dans les zones de risque majeures du paludisme que les gens habitent aussi éloigné d'un point d'eau. En plus, des expériences ont démontré que les moustiques infectés du paludisme peuvent être amenés par des voitures (par exemple) dans des zones non-vulnérables auparavant.

**Caractéristiques du point eau** La surface et surtout la turbidité de l'eau jouent un rôle non négligeable dans le développement du paludisme. Lorsque le courant est trop fort, les œufs des moustiques sont lavés, les moustiques ne peuvent donc pas se reproduire. Des essais en laboratoire (Minakawa *et al.*, 1999) ont montré que les moustiques anophèles pondent plus d'œufs dans les points d'eau se situant sur des sols inondés ou humides que dans de l'eau "sans sol". En plus, les moustiques anophèles n'aiment pas, en général, les habitats ombragés tels que des réservoirs d'eau sans substrats de sol. En effet, le sol fournit des éléments nutritifs qui favorisent l'accumulation de bactéries qui sont la source de nourriture pour les larves.

**Altitude** L'altitude maximale retenue généralement est de 2000m. Ce facteur est bien évidemment étroitement lié au facteur de la température car la température diminue de 0.7°C tous les 100 mètres.

**Humidité de l'air** L'humidité relative de l'air a un impact important sur la présence et la persistance des sites de reproduction des moustiques. Le taux de survie des moustiques est également influencé par ce facteur.

Le facteur peut être extrait à partir d'autres facteurs météorologiques comme les précipitations et la température mais doit être utilisé avec précaution. L'humidité est fortement influencée par la température de l'air et peut donc significativement changer pendant un seul jour. A noter que ce facteur dépend également de l'altitude. L'humidité idéale pour les moustiques est de 60 % (Protopopoff *et al.*, 2009).

**Humidité du sol** L'humidité du sol dépend directement des températures, de la végétation ou de la présence de points d'eau. Ce facteur est particulièrement intéressant car il peut être extrait à partir des images satellites. (Machault *et al.*, 2011)

**Végétation :** Certains types de végétation servent comme habitat pour les moustiques adultes. (Minakawa *et al.*, 1999) (à revoir)

### **Facteurs biologiques = marqueur de risque**

Les facteurs de risque de transmission biologiques sont liés par exemple à l'organisme humain ou à l'organisme du moustique. L'état de l'organisme humain, sous certaines conditions, peut augmenter le risque de l'apparition du paludisme sur un territoire. Certains de ces facteurs comme l'âge sont également des marqueurs de risque et ne peuvent pas être influencés respectivement modifiés.

Numéro	Facteur
1	Etat de santé de la personne
2	Densité du vecteur
3	Transmission Homme-Moustique
4	Immunité
5	Age

**État de santé de la personne** L'état de santé de la personne joue un rôle majeur dans le développement du paludisme. Les personnes les plus en risque sont les enfants de moins de trois ans et les femmes enceintes et donc les personnes avec les systèmes immunitaires les plus fragiles. Les adultes ou les adolescents présentent généralement des systèmes immunitaires suffisamment puissants pour combattre le paludisme. (Protopopoff *et al.*, 2009)

**Densité du vecteur** Pour combattre le paludisme, il n'est pas nécessaire d'éradiquer complètement les moustiques sur le territoire mais il suffit de réduire la densité vectorielle (Gaudart *et al.*, 2005). En général, la densité de l'anophèle diminue avec l'éloignement du gîte, avec la densité du tissu urbain et de la périphérie vers le centre (Gaudart *et al.*, 2005). En analysant les densités vectorielles sur différents territoires, il est possible de déterminer les végétations ou écosystèmes les plus favorables pour le développement des larves.

**Transmission Homme-Moustique** En cas de piqûre, un humain porteur du parasite du paludisme transmet la maladie aux moustiques non-infectés par avant. Ceci est un autre facteur de risque à ne pas négliger, le taux de transmission est de 20% (30% pour une transmission moustique-homme).

**Immunité** La capacité d'un organisme humain pour combattre l'infection du paludisme dépend de son immunité (Protopopoff *et al.*, 2009). A partir de 2 à 3 ans, l'organisme humain développe et augmente l'immunité indépendamment du nombre de piqûres. Dans des régions avec peu de piqûres, le nombre de cas cliniques et d'infection est le même pour tous les groupes d'âge. Il faut donc que l'homme soit piqué régulièrement afin de développer une certaine immunité contre l'infection du paludisme.

**Age** Les enfants de moins de 3 ans sont piqués plus souvent, il semble que la proportion entre piqûres et un être humain soit liée à la taille du corps de l'hôte. (?)

### **Facteurs humains**

Les facteurs humains sont les facteurs liés à la présence humaine sur un territoire qui augmentent la vulnérabilité des habitants du territoire par rapport à la transmission du paludisme.

Numéro	Facteur
1	Urbanisation
2	Agriculture
3	Qualité système santé
4	Croissance démographique et Facteurs socio-économiques

**Urbanisation** L'urbanisation généralement diminue le risque du paludisme. Les points d'eau sont pollués, il y a peu de végétation etc. Ceci dépend malgré tout de la taille de l'urbanisation. Forcément, dans une grande ville le risque d'être piqué est moins grand car il y plus de "cibles" pour les moustiques, en même temps de nouveaux facteurs de risques peuvent apparaître comme par exemple les lumières et télévisions étant une source d'attraction pour les moustiques.

**Agriculture / Irrigation** Les activités agricoles influencent directement le risque du paludisme. Par exemple, en irriguant régulièrement les terres agricoles, ces surfaces (=surface d'eau temporaires et de bonne qualité) deviennent des endroits idéaux pour les moustiques pour pondre des œufs. La déforestation joue également sur le risque du paludisme. Des études récentes ont démontré que des habitations terrestres (des hommes) se situant en altitude, construits après une déforestation étaient des sites de reproduction préférés par les moustiques.(Krefis *et al.*, 2011)

**Qualité système de santé** La qualité du système de santé ou l'accès au système de santé influe directement sur le risque du paludisme (Protopopoff *et al.*, 2009). Des traitements préventifs permettent de réduire le taux de morbidité des femmes enceintes et des enfants, qui généralement ont le système immunitaire le plus fragile.

**Croissance démographique / Facteurs socio-économiques** Le statut socio-économique d'un individu est également d'importance. Les personnes les plus prospères sont capables de mieux se protéger contre les moustiques que les personnes très pauvres. Aussi, le niveau d'éducation peut influencer le risque de paludisme, Sachant que souvent les personnes n'utilisent pas les filets anti-moustiques mis à leur disposition car ils ne comprennent pas vraiment le risque d'être piqués et le risque de paludisme, le niveau d'éducation des habitants est également à prendre en compte.

Le type d'habitat de l'homme joue également un rôle important. En fonction du type d'habitat, le moustique peut entrer plus ou moins facilement dans l'habitat et piquer l'humain. Certains types d'habitations (en fonction du type de construction) peuvent même servir comme lieu d'habitat aux moustiques.

### **Alea et Vulnérabilité selon les facteurs**

Les différents facteurs peuvent être regroupés en deux catégories différentes : les facteurs liés à l'aléa et les facteurs liés à une vulnérabilité. Certains facteurs peuvent créer un aléa et une vulnérabilité en même temps. Par exemple, en irriguant des terres, les agriculteurs créent involontairement un aléa (en créant des points d'eau temporaires) et aggravent en même temps la vulnérabilité des habitants du territoire. La figure 3.1 récapitule quels facteurs peuvent être à la base d'un aléa et lesquels peuvent aggraver la vulnérabilité.



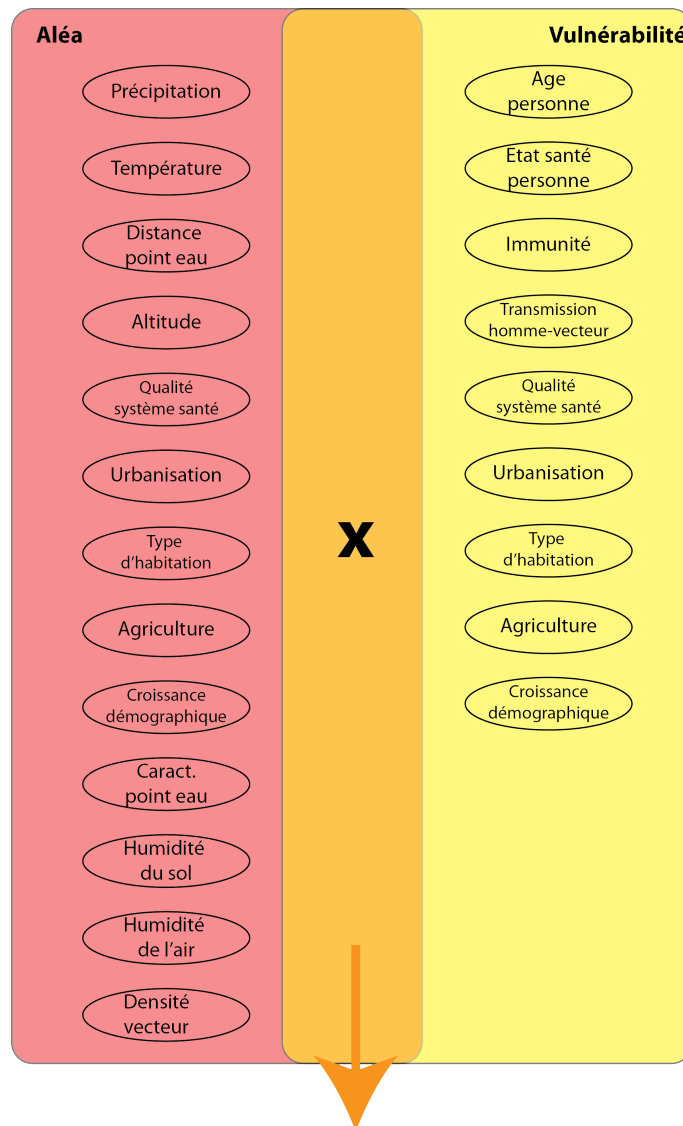


FIGURE 3.1 – Alea et Vulnérabilité selon les facteurs

## 3.2 Approches logicielles existantes

Nous nous sommes simultanément intéressés aux logiciels et aux chaînes de traitements existants qui traitent les mêmes problématiques et / ou des problématiques similaires. Je présenterai dans cette partie du mémoire les approches logicielles que j'ai pu découvrir en faisant des recherches. Ces logiciels et outils ne feront pas partie du prototype de la chaîne de traitements.

Nous avons focalisé nos recherches sur des outils qui se basent sur le principe d'analyse par rapport à la variable de certaines variables (par exemple présence de points d'eau sur un territoire). Un autre principe d'analyse aurait été de définissant le risque à travers l'absence de certaines variables (par exemple pas de végétation à proximité des lieux d'habitation des population).

## **Repast Symphony**

Repast Symphony est une plate-forme de modélisation basée sur Java. Repast Symphony est un plugin Eclipse et permet de développer des programmes avec de nombreuses interactions. Repast Symphony a déjà été utilisé dans de nombreux domaines, comme en sciences sociale ou en sciences humaines.

## **Maxent**

Maxent est un logiciel gratuit qui permet de prédire la distribution potentielle (modélisation de l'habitat) des espèces animales ou végétales en se basant sur la distribution ponctuelle et certains facteurs environnementaux des espèces.

L'outil utilise comme variables d'entrée des données géoréférencées des animaux respectivement de la végétation à modéliser et des données appropriées aux variables environnementales (par exemple pluviométrie, température, topographie etc.) au format ASCII (format ESRI). La modélisation est basée sur la méthode de l'entropie maximum. L'outil livre en tant que résultat une carte indiquant l'apparition potentielle ainsi que la d'autres résultats statistiques. La visualisation des résultats se fait à l'aide d'un système d'information géographique. L'outil a été développé en Java et est disponible pour tous les systèmes d'opération.

## **OpenModeller**

OpenModeller vise à fournir un environnement multiplateforme permettant la réalisation de l'ensemble du processus de "niche environnementale". Le logiciel facilite la lecture de l'occurrence des espèces et des données environnementales, la sélection des couches de données environnementales, la création d'un modèle de niche fondamentale et la projection du modèle dans un scénario de l'environnement. Un certain nombre d'algorithmes sont également fournis sous forme de plugins. Le projet est Open Source.

### **3.2.1 Cas d'utilisation concret dans le contexte environnement-santé :**

#### **Le projet SimMasto**

Le projet SimMasto s'inscrit dans le cadre d'un projet de recherche sur la dynamique des populations de rongeurs. Il vise à développer une plate-forme générique de simulation des rongeurs dans leur environnement. Les données utilisées sont des cartes géographiques numériques, des cartes raster, des grilles théoriques etc. La chaîne de traitements qui a été développée comprend les éléments permettant le traitement des images d'entrée (géoréférencement, détourage, changement de résolution, clipping, etc.) ainsi que tous les éléments de programmation nécessaires.

La chaîne de traitements est composée de plusieurs "modules", comme par exemple :

- Géoréférencement d'une image de type raster
- Vectorisation d'une image de type raster (utilisation de la librairie Grass)
- Modification du système de coordonnées

La chaîne de traitements fait par la suite appel à Repast Symphony pour mettre en place une simulation. Dans Repast Symphony est implanté un SIG, c'est-à-dire les différents fichiers de formes géoréférencés créés par avant ainsi qu'un fichier raster (une grille). D'autres outils / bibliothèques utilisés dans cette chaîne de traitements sont par exemples PostgreSQL / PostGIS ou Eclipse.

### **3.2.2 Conclusion**

Dans le contexte choisi de l'environnement-santé, il est intéressant de conceptualiser le cycle de paludisme. Le cycle du paludisme est bien connu de nos jours, mais étant très complexe, un modèle conceptuel permet également à des non-experts du domaine de mieux comprendre ce cycle.

En ce qui concerne le volet informatique du travail, il n'existe à ce jour aucune chaîne de traitement automatisée permettant de cartographier facilement les zones de risque du paludisme. Cette partie du travail réalisé consiste donc une partie innovante, notamment à cause de l'approche "Open Source".

Au contraire, de nombreux travaux et études ont démontré l'apport potentiel des outils informatiques à la lutte contre paludisme.

Les différentes définitions sont clairement établies, par contre sur plan réalisation informatiques de chaînes de traitements, aucun outil semble proposer la solution attendue.

Il est donc particulièrement intéressant de proposer des prototypes d'outils permettant de cartographier les indicateurs relatifs à divers risques environnementaux.

# Chapitre 4

## Méthodologie

L'objectif principal du stage est de proposer un outil et une architecture logicielle permettant d'automatiser les traitements nécessaires pour cartographier le risque dans le contexte du paludisme (environnement-santé). Calculer le risque ou les indicateurs revient à définir des chaînes de traitements. Celles-ci sont conceptualisées à partir de l'expertise du domaine.

Une première partie d'analyse est dédiée à la compréhension du phénomène (paludisme) et à la conceptualisation des chaînes de traitements. Un modèle conceptuel UML des éléments nécessaires à l'élaboration des indicateurs permet de dégager les facteurs de risque les plus pertinents. A partir des données disponibles, un deuxième modèle UML simplifié servira comme base pour la définition des traitements nécessaires à l'élaboration et au développement de la chaîne. Dans un second temps, des modèles conceptuels des traitements permettent de d'analyser et de définir les différentes étapes nécessaires pour le développement informatique.

Dans une deuxième partie nous présentons l'architecture informatique des outils ainsi que tout le travail de conception mené autour du développement informatique.

Finalement, nous présentons l'opérationnalisation et l'implémentation des deux outils.

### 4.1 Analyse

Le cycle du paludisme permet de comprendre de façon simplifiée le développement du paludisme.

Schema paludisme

Le modèle conceptuel des éléments nécessaires à l'élaboration des indicateurs (4.1) permet d'avoir une vue d'ensemble des facteurs qui interviennent dans le développement du paludisme et qui peuvent être une source de risque. Les grandes caractéristiques du paludisme ont été définies dans une démarche participative lors d'une réunion avec des experts de différents domaines. A partir de ces caractéristiques j'ai élaboré un modèle conceptuel UML.

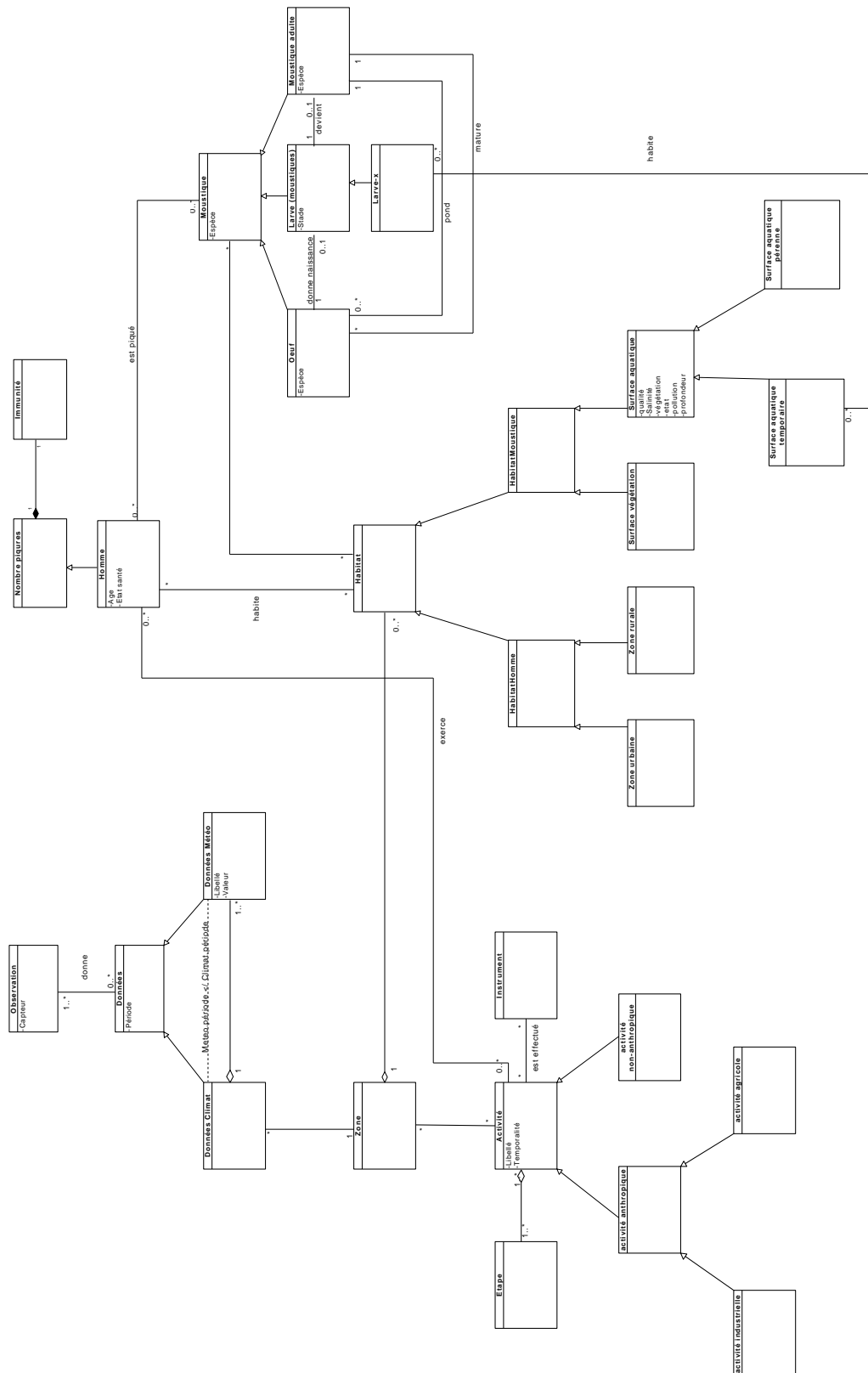


FIGURE 4.1 – Modèle conceptuel UML des éléments nécessaires à l'élaboration des indicateurs

## Précision modèle conceptuel

Des observations ont été effectuées sur un territoire (dans notre cas le village de Bandiagara au Mali). Les **observations** sont réalisées à l'aide de capteurs, comme par exemple de satellites. Une observation crée des **données** pour une période donnée. Dans le contexte de ce travail, il faut différencier les **données climat** et les **données météo**, c'est-à-dire que nous allons plutôt utiliser des données météo, qui elles font partie de données climatologiques. Les données climatologiques (généralement 30 ans) sont enregistrées sur une période plus longue, pour définir les grands principes du climat sur un territoire. Une donnée climatologique comprend donc plusieurs données météo, par exemple la température moyenne sur le territoire pendant le mois de Janvier 2008. Une donnée climatique correspond à une **zone** définie. Sur cette zone sont exercés des **activités**. Ces activités sont effectuées à l'aide d'**instruments**. Une activité se divise en plusieurs étapes et peut être de type **anthropique** ou de type **non-anthropique**. Une activité anthropique est une activité relative à l'activité de l'homme comme par exemple les activités industrielles ou agricoles.

Les hommes habitent des **zones rurales** ou des **zones urbaines**. Cette différenciation est très importante comme en zone urbaine le risque d'être infecté par le paludisme est beaucoup plus faible qu'en zone rurale car généralement les **eaux de surface** sont trop polluées. Les moustiques ont besoin d'**habitats** et d'endroits pour pondre leurs œufs. Les *larves* nécessitent des surfaces aquatiques. La qualité, la salinité etc. d'une surface d'eau, la pollution et la profondeur jouent un rôle déterminant dans le cycle du développement de la larve. Généralement dans les zones à risque les **surfaces aquatiques pérennes** sont trop polluées et ne permettent pas le développement du moustique. Ainsi, un des facteurs de risque majeurs du paludisme est la présence de **surfaces aquatiques temporaires** car ils servent comme lieu de pontage des œufs de moustique jusqu'à ce que les larves deviennent des moustiques. Les **moustiques adultes** généralement habitent des surfaces de végétation. Finalement les moustiques piquent les hommes. En fonction des nombres de piqûres, de l'âge de l'homme et de son état de santé, l'homme développe une certaine immunité. Un **homme** porteur du parasite peut également infecter un moustique non infecté auparavant.

### 4.1.1 Données disponibles

Nous disposons d'un certain nombre de données et d'informations relatives au territoire d'étude (Ville de Bandiagara). Ces données ont été extraites par Nadine Dessay à partir d'une image à très haute résolution Quick Bird (précision...). Les données correspondent aux facteurs de risque suivants :

- Végétation
- Eau
- Bâtiments

En plus, nous disposons de données statistiques (données brutes) issu d'un recensement de la population de la ville de Bandiagara en 2004 (14133 habitants).

### 4.1.2 Modèle simplifié

A partir des données disponibles, nous avons élaboré un modèle simplifié qui a servi comme point de départ pour la conceptualisation et l'élaboration de la chaîne de traitements. La partie concernant les activités effectuées sur une zone donnée a été écartée du modèle car il sera quasi-

ment impossible de disposer de ces données. Il en est de même pour le cycle de vie du moustique qui n'interviendra pas dans les traitements. Finalement le modèle suivant a été retenu :

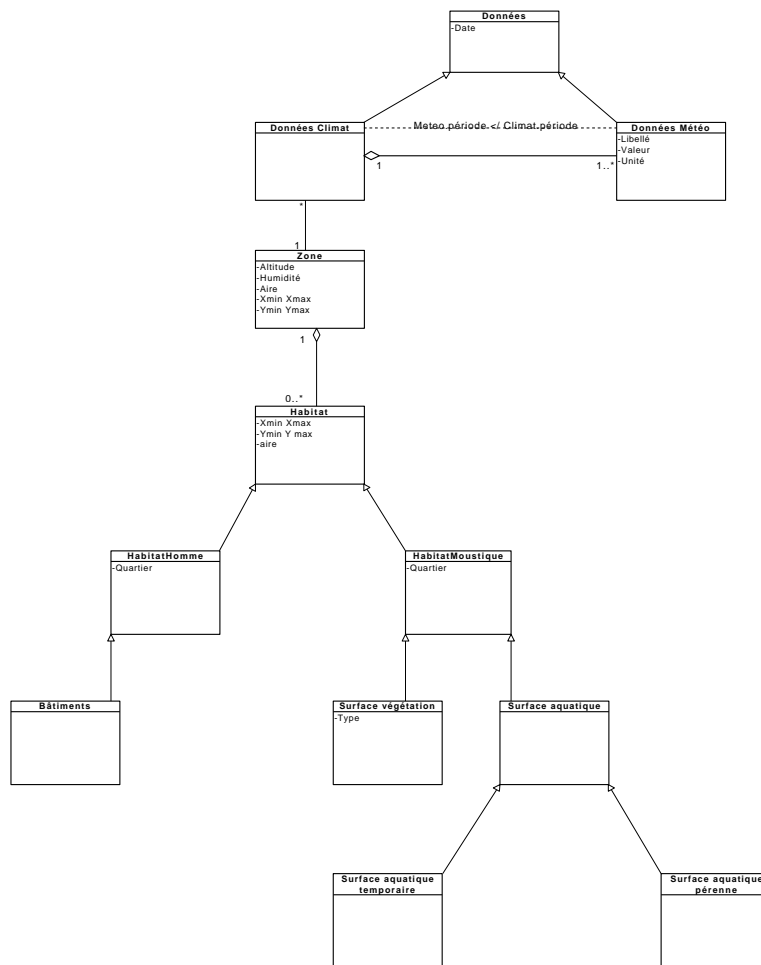


FIGURE 4.2 – Modèle UML simplifié des indicateurs du paludisme

Le modèle conceptuel simplifié comprend les catégories de risques suivantes :

- Données climatologiques (précipitations, température etc). => aléa
- Présence de surface d'eau (temporaires ou pérennes) => aléa
- Présence de surface végétales => aléa
- Habitations humaines => vulnérabilité

La présence humaine est indispensable pour qu'il y ait un risque (vulnérabilité) dans une zone. La présence de bâtiments a été retenue comme facteur de risque lié à cette présence humaine. Les surfaces végétales et les surfaces d'eau sont retenues comme lieu d'habitation potentiel pour les moustiques anophèles respectivement pour les larves de moustiques anophèles.

Ce modèle UML simplifié nous a servi comme base pour définir dans un premier temps les traitements nécessaires et pour conceptualiser la chaîne de traitements.

## 4.2 Raisonnement

A partir des données disponibles, du modèle conceptuel simplifié et en collaboration avec des experts dans le domaine de l'environnement-santé et de la télédétection le raisonnement nous a permis de définir les traitements nécessaires pour l'élaboration de cartes de risque du paludisme. Nous nous sommes basés sur les travaux réalisés par Nadine Dessay. Une fois les traitements nécessaires définis nous nous sommes intéressés à la conceptualisation des traitements et des données. Cette partie est indispensable pour comprendre le futur fonctionnement de la chaîne de traitements.

De façon très schématique le fonctionnement de la chaîne de traitements correspond au modèle suivant :

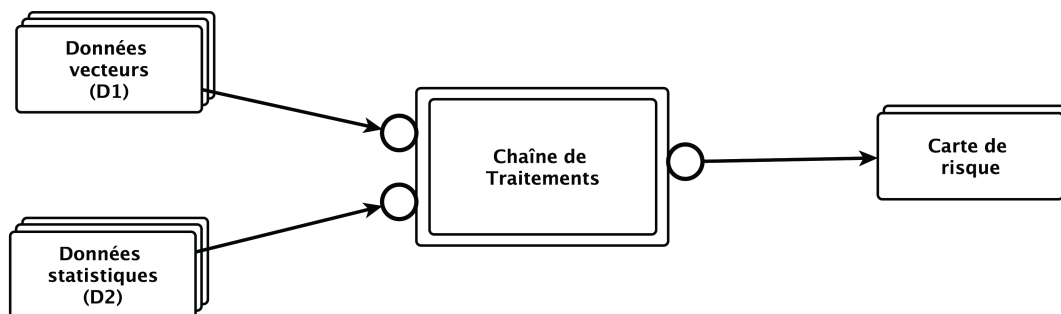


FIGURE 4.3 – Description générale de la chaîne de la chaîne de traitements

Des traitements sont effectués sur des données au format vecteur et sur des données statistiques pour obtenir en sortie une carte de risque. Nous expliquons dans un premier temps ce que fait un traitement et comment fonctionne une donnée. Ceci est fait à partir de **diagrammes de classes**.

Par la suite, des catégories de données et des catégories de traitements de la chaîne sont proposées à l'aide de **modèles conceptuels des hiérarchies**. Ceci permet de mieux comprendre le fonctionnement et les différentes étapes de la chaîne de traitements.

### 4.2.1 Description des données et des traitements

Nous allons décrire et expliquer les traitements réalisés au sein de la chaîne. Cette description peut être regroupée en 2 parties :

- La description globale de ce que fait un traitement
- La description des données

Dans la suite de cette partie, nous expliquons successivement les deux aspects cités ci-dessus. Le modèle (4.4) qui suit décrit le fonctionnement d'un traitement.



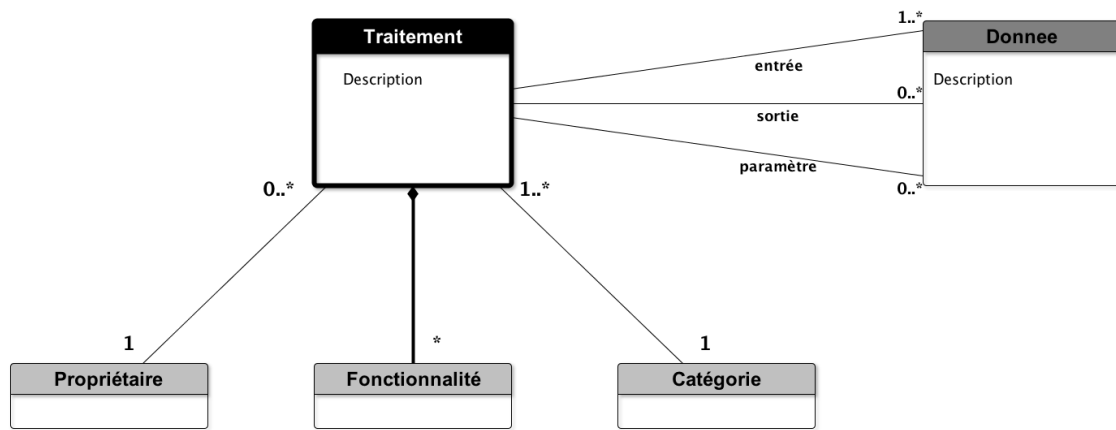


FIGURE 4.4 – Description du fonctionnement d'un traitement

**1) Description globale de ce que fait un traitement** Un traitement correspond à un propriétaire (contact), celui qui a développé respectivement défini le traitement. Chaque traitement correspond également à une ou plusieurs fonctionnalités et à une catégorie de traitements. Une fonctionnalité d'un traitement est par exemple d'ouvrir un fichier ou de le transformer. Un traitement correspond aux catégories de traitement suivantes : pré-traitements, traitements et post-traitements. (cf. (Lin, 2011))

Un traitement peut être relatif à trois types de données : des données d'entrée, des données de sortie et des données de paramétrage. Les données seront expliqués plus en détail dans le paragraphe suivant.

Exemple d'un traitement : shp2pgsql

- Propriétaire : PostGIS / GDAL
- Fonctionnalité(s) : Transformation du format, reprojection
- Catégorie : Traitement
- Donnée d'entrée : Fichier au format vecteur (shape)
- Donnée en sortie : Fichier au format ".sql"
- Donnée de paramétrage : postgis.sql, spatial\_ref.sql

**2) Description des données** Décrire les traitements d'un système demande également de décrire les données utilisées. Une donnée correspond à une ou plusieurs catégories de données, un ou plusieurs propriétaires et un format. (cf. (Abd-el Kader, 2006))

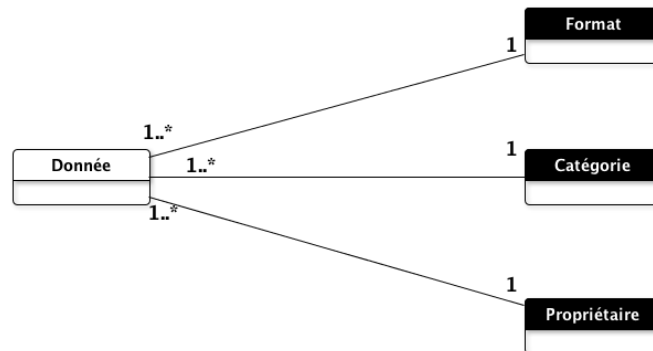


FIGURE 4.5 – Description des données

Dans le contexte de ce travail, les formats de données utilisés sont les suivantes : Format vecteur (fichiers de forme géoréférencés), format raster (images composites géoréférencés) et format tableur (statistiques). Ces données correspondent à des fonctionnalités comme par exemple la création de zones tampons. Chaque donnée est également caractérisée par un propriétaire (contact) et une catégorie de données. Nous pouvons regrouper les données utilisés en deux catégories principales : les données issus de capteurs satellites et les données récoltés sur le sol (stations ou capteurs au sol). Une donnée correspond également à un type de données. Une donnée vecteur peut être par exemple de "type" ligne ou polygone.

Exemple de donnée : b\_bati.shp

- Propriétaire : Nadine Dessay
- Catégorie : Raster
- Format : vecteur (shape)

## 4.2.2 Les catégories de données et de traitements

### Les catégories de données

Le modèle conceptuel des catégories de données (4.6) décrit de façon formelle les catégories de données qui sont utilisées par la chaîne de traitements.

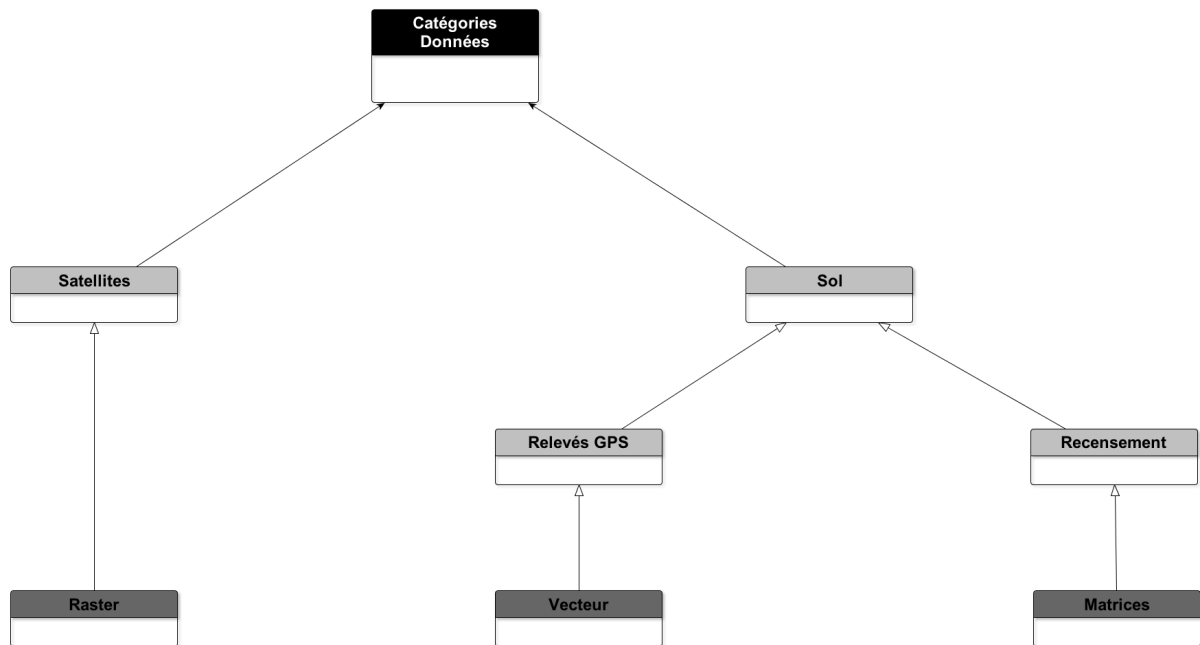


FIGURE 4.6 – Modèle conceptuel des catégories de données de la chaîne de traitements

Les données de la chaîne de traitements sont issues de sources diverses. Les **données issues de satellites** correspondent aux données obtenues à partir de satellites (passifs ou actifs). Les **données issues du sol** sont les données qui proviennent de capteurs ou de toute autre source se situant sur le sol terrestre.

Les satellites produisent des **images raster**, à partir desquelles, en faisant recours à des traitements de télédétection, peuvent être extraites de nombreuses informations. Pour notre chaîne de traitements, les informations suivantes ont été extraites par des experts en télédétection (Nadine Dessay) :

- Eau : Les surfaces d'eau
- Végétation : Les arbres etc.
- Urbanisation : Bâtiments

Les données récoltées sur le **sol** peuvent correspondre à différentes catégories. Par exemple, une station météo peut fournir des données de catégorie vecteur ou raster. Des relevés GPS fournissent des vecteurs. Des statistiques (catégorie matrices) sont obtenues à partir de recensements de population par exemple.

### Les catégories de traitements

Nous proposons trois grandes catégories de traitements : Les **pré-traitements**, les **traitements** et les **post-traitements**. La catégorie des pré-traitements recouvre l'ensemble des opé-

rations qui préparent les données pour qu'elles puissent être traitées. Dans le cadre de ce travail il est nécessaire de stocker les données et d'effectuer des requêtes SQL afin de sélectionner les données à en fonction des traitements.

La catégorie des traitements recouvre les traitements proprement dits, c'est à dire les traitements qui vont créer ou modifier des données. Cette catégorie présente trois sous-catégories : les traitements **statistiques**, les traitements **d'analyse spatiale** et les traitements de **transformation**.

Les traitements statistiques concernent les calculs de densités de population, de densité de points et le calcul de la taille de cellule des couches raster.

Les traitements d'analyse spatiale sont les traitements géographiques sur les **fichiers de forme** (shape) (par exemple intersection entre deux fichiers de forme vecteurs).

Les traitements de transformation correspondent aux traitements qui modifient le format des données. Par exemple, pour insérer un fichier vecteur dans une base de données spatiale, il est nécessaire d'effectuer une transformation vers le format ".sql".

Les post-traitements concernent tout ce qui est représentation des résultats obtenus, par exemple une reclassification d'une image raster pour faire ressortir une information plus clairement.

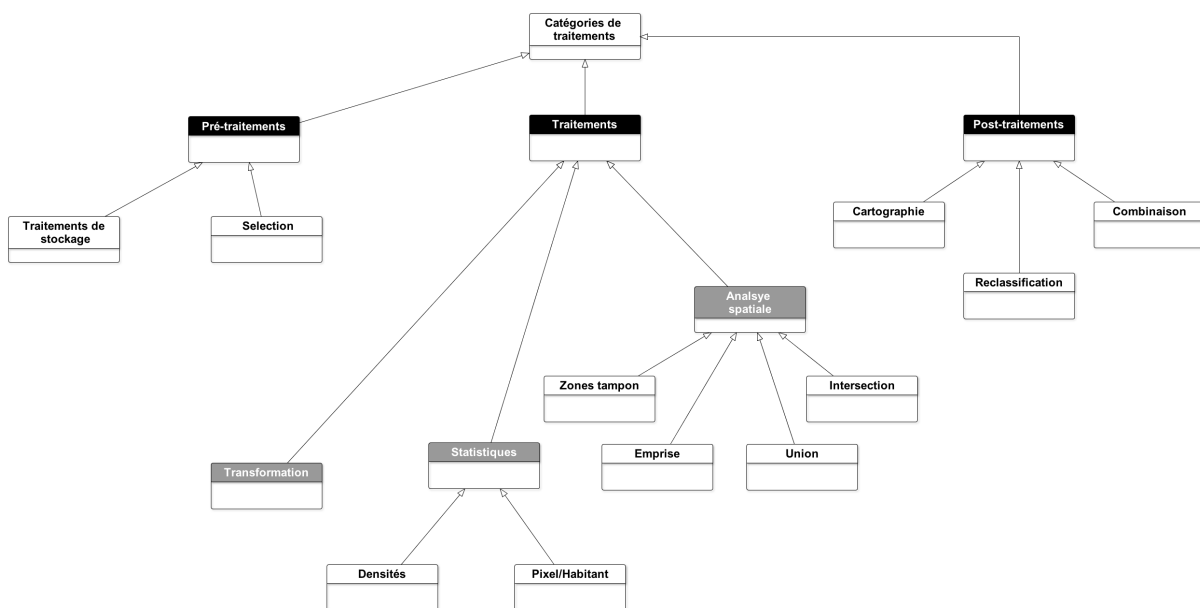


FIGURE 4.7 – Modèle conceptuel des hiérarchies de catégories de traitements de la chaîne de traitements

### 4.2.3 Conceptualisation de la chaîne de traitements

Le modèle conceptuel des traitements permet de représenter la dynamique de la chaîne de traitements, c'est-à-dire les opérations et traitements qui sont réalisées en fonction d'autres événements. Ce modèle permet donc de représenter et d'expliquer de façon conceptuelle le fonctionnement du système sans faire référence aux choix qui ont été réalisés (par exemple quelle librairie a été utilisée, quel langage informatique etc.). Le modèle explique donc les traitements qui sont effectués dans la chaîne mais il n'explique pas comment ils sont effectués.

Pour illustrer les différentes opérations de la chaîne de traitements, nous allons détailler successivement comment se construit la chaîne de traitements à partir des descriptions précédentes et comment l’instanciation est réalisable à partir du modèle de départ (cf 4.9)

Pour faciliter la représentation des chaînes abstraites à partir de , nous utiliserons le langage graphique proposé par Yuan Lin (4.8). Les ports de données associés à un traitement abstrait ont pour objectif de distinguer les différents flux de données en terme d’entrée / sorties (Lin, 2011).

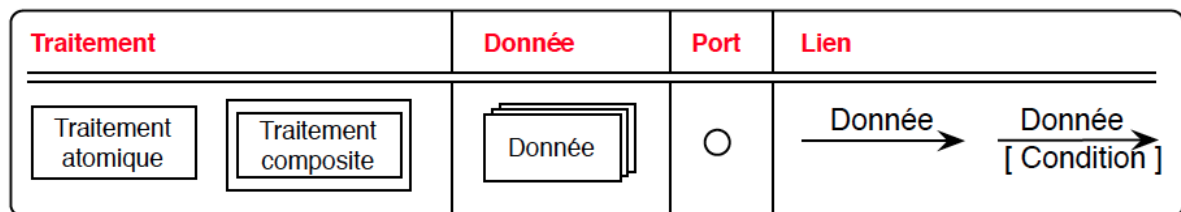


FIGURE 4.8 – Langage graphique proposé par Yuan Lin (Lin, 2011)

### Description générale

Pour créer une carte de risque, il est nécessaire d’effectuer des traitements sur des données. Les données d’entrée correspondent à deux formats différents : format vecteur et format statistique. Ces données sont utilisées par la chaîne de traitements pour effectuer les traitements nécessaires pour cartographier le risque.

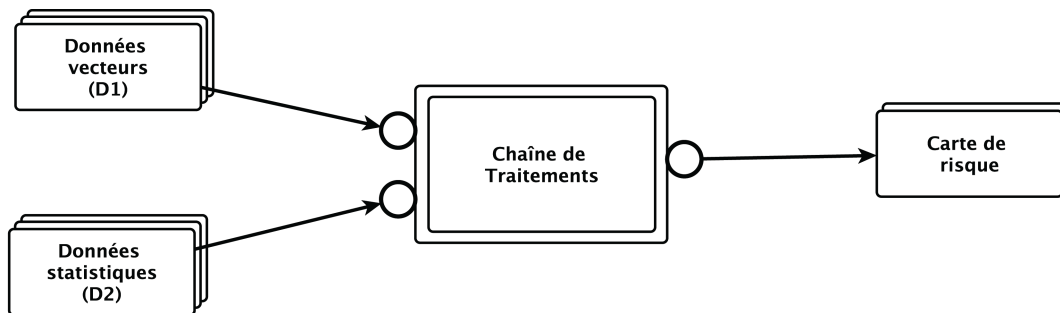


FIGURE 4.9 – Description générale

### Chaîne abstraite

Dans un premier temps, les données vecteurs sont stockées (traitements de stockage) dans une base de données. Par la suite, des **sélections** permettent de définir précisément les données qui seront traitées lors de chaque traitement.

Le risque est la combinaison entre l’**aléa** et la **vulnérabilité**. Deux grandes catégories de traitements sont effectuées : les traitements de vulnérabilité et les traitements d’aléa. Le traitement de risque permet de créer des cartes de risque.

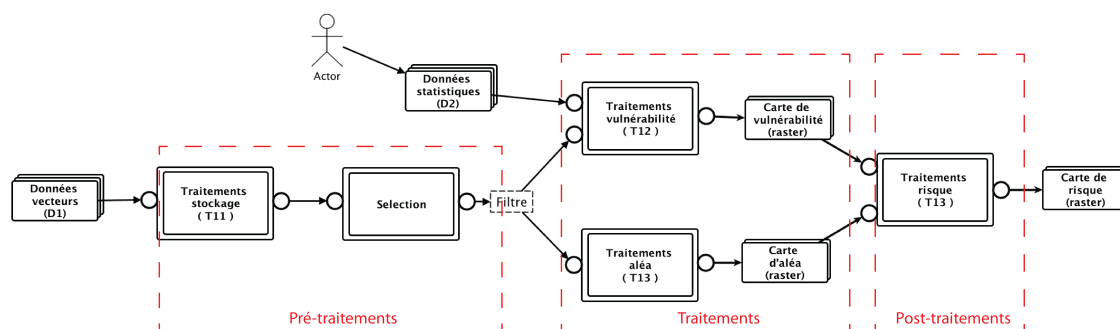


FIGURE 4.10 – Chaîne abstraite

## Traitements élémentaires de la chaîne de traitements

Chaque donnée vectorielle stockée dans la base de données est d'abord reprojété, la projection de la base de données spatiale et de toutes les données insérées sera la même. Les données sont insérées dans une base de données spatiale (créée au cours de l'exécution de la chaîne de traitements). Des requêtes SQL permettent de sélectionner (filtrer) les données à utiliser pour chaque traitement.

Les traitements de vulnérabilité correspondent à deux catégories de traitements : des traitements d'analyse spatiale et des traitements statistiques. La chaîne de traitements demande à l'utilisateur le nombre d'habitants de la zone d'études (= données statistiques). La chaîne de traitements crée une carte de vulnérabilité en combinant les deux catégories de traitement.

Les traitements d'aléa correspondent à des traitements d'analyse spatiale sur les données "eau" (format vecteur). L'ensemble de ces traitements permettent de créer une carte d'aléa.

Les traitements de risque correspondent à la combinaison de la carte de vulnérabilité et de la carte d'aléa. La carte de risque est le résultat de l'ensemble des traitements.

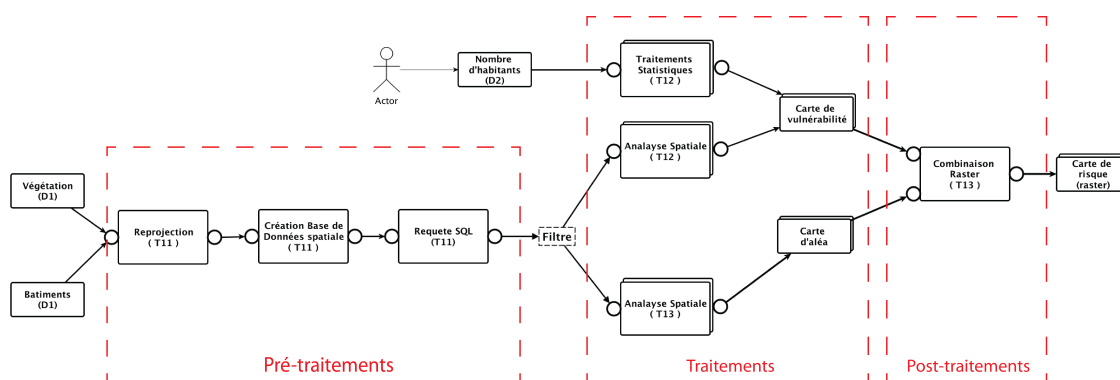


FIGURE 4.11 – Chaîne abstraite avec traitements élémentaires

## Chaîne de traitements de plus en plus détaillées

Des requêtes SQL permettent de calculer la surface totale de la couche "bâtiments". A partir des informations saisies par l'utilisateur (nombre d'habitants), la chaîne calcule la densité de population de la zone d'étude. A partir de ces informations, la surface qu'occupe en théorie un

habitant permet de calculer la taille théorique d'une cellule pour un habitant. La couche "bâtiments" est ensuite rastérisée en fonction de la taille de la cellule calculée par avant. Pour chaque cellule de ce raster, un point est créé et le calcul de la densité des points permet de créer une carte de vulnérabilité.

En ce qui concerne les traitements d'analyse spatiale de l'aléa, plusieurs traitements sont successivement effectués : définition de l'emprise de la zone d'étude, zone tampons autour de la couche "eaux" (400 et 600 mètres), union et intersection des résultats pour arriver à la création d'une carte d'aléa.

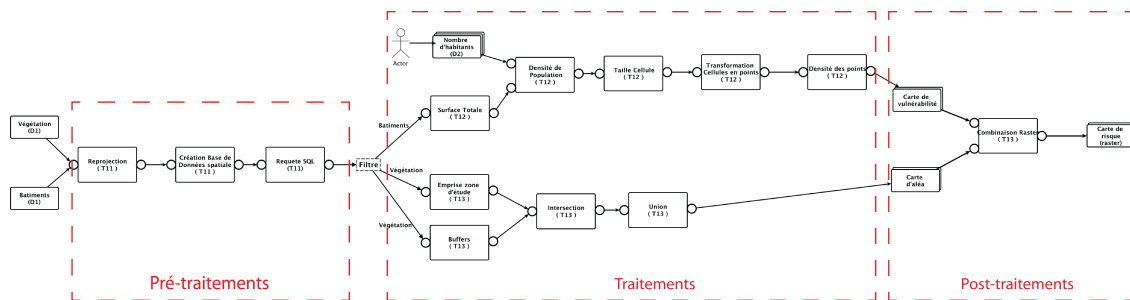


FIGURE 4.12 – Chaîne de traitements de plus en plus détaillée

### Chaîne de traitements instanciée

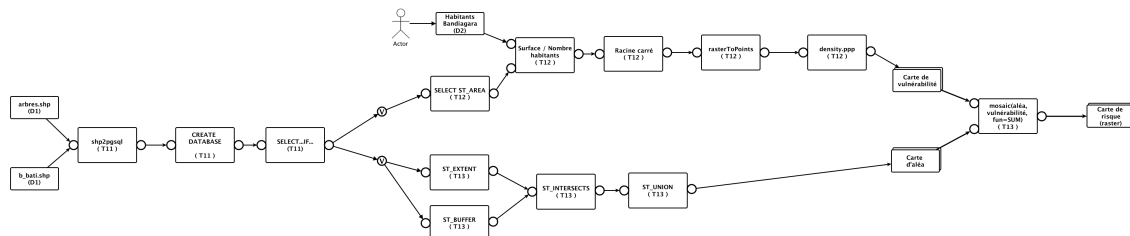


FIGURE 4.13 – Chaîne de traitements instanciée

## 4.3 Conception

Pour valider sur le plan opérationnel notre approche, nous avons réalisé deux types d'outils. Dans un premier temps, nous avons développé une chaîne automatisée de traitements. L'utilisateur choisit les données d'entrée et la chaîne effectue automatiquement tous les traitements nécessaires pour la création d'une carte de risque.

Dans un second temps, nous avons développé un logiciel ouvert, l'utilisateur peut ainsi choisir les traitements qu'il souhaite effectuer.

Dans cette partie du travail, nous présentons l'implémentation des deux outils, respectivement l'architecture informatique des deux outils. Les librairies et bibliothèques utilisées sont présentées en Annexe.

### 4.3.1 Architecture informatique de la chaîne de traitements et du logiciel ouvert

Le prototype de la chaîne de traitements et le logiciel ouvert appellent les fonctionnalités des outils PostgreSQL / Postgis (via le JDBC PostgreSQL) et R (via RCaller) via le langage Java. Le schéma suivant explique de façon très simplifiée l'architecture informatique des deux outils.

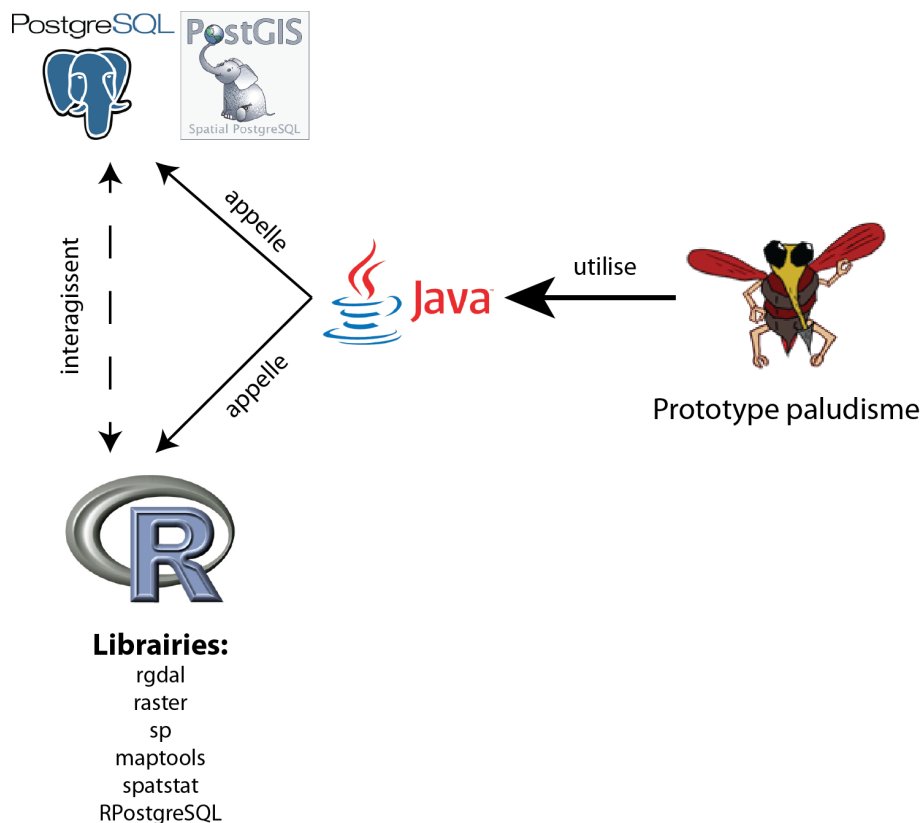


FIGURE 4.14 – Architecture informatique du prototype

#### Présentation de l'outil vision "fermée" (boîte noire)

Les données utilisées sont insérées dans une base de données PostgreSQL (incluant l'extension PostGIS). Une nouvelle base de données spatiale est créée automatiquement lors de chaque exécution de la chaîne de traitements. Les différentes bibliothèques utiliseront les données dans cette base de données et interagissent entre elles. Par exemple, des données se trouvant dans la base de données PostgreSQL vont être chargées dans R afin de permettre de calculer la densité de population de la zone d'étude.

L'architecture informatique est basée exclusivement sur des outils et bibliothèques Open Source et a permis de créer un prototype générique et modulaire. Les fonctions de PostgreSQL respectivement de Postgis permettent d'effectuer la majorité des traitements sur des données vecteurs comme des intersections, des buffers ou des unions. Depuis la version 2.0, Postgis gère également les fichiers au format raster. Ainsi, il est désormais possible d'insérer des fichiers raster dans la base de données PostgreSQL ou de transformer des fichiers au format vecteur en format raster. Cette version de Postgis est encore en cours de développement, de nouvelles fonctionnalités raster vont certainement être rajoutées dans le futur.



L'outil R permet de manipuler des données aux formats différentes grâce aux nombreuses librairies disponibles pour cet outil. La librairie "spatstat" nous permet par exemple de calculer une densité de population ou d'effectuer une reclassification d'une image raster. La librairie "rgdal" nous permet de disposer des fonctionnalités de la librairie **GDAL** et de charger dans R les données raster ou vecteurs stockées dans une base de données.

Certaines fonctionnalités sont disponibles dans les deux outils et nous avons faits plusieurs tests pour trouver l'outil le plus pertinent par rapport à une problématiques. Par exemple la transformation d'un vecteur en raster nécessite beaucoup plus de temps sous R qu'en utilisant la fonction "ST\_Raster" de Postgis.

La multiplicité des fonctions disponibles avec ces deux outils, sachant qu'il existe des librairies R par rapport à de nombreuses problématiques, garantit que cette architecture pourra servir comme base pour un futur développement information du SIEL et être réutilisable par des experts de différents domaines.

### **Présentation du logiciel vision "ouvert" (boîte blanche)**

Le logiciel ouvert fonctionne sur le même principe que la chaîne de traitements. Le logiciel a été développé en Java et se connecte à un **SGBD** via un JDBC et au logiciel R via RCaller. En fonction des traitements que l'utilisateur veut effectuer, le logiciel appelle les fonctionnalités PostgreSQL/PostGIS ou R.

Néanmoins, un certain nombre de fonctionnalités supplémentaires a été rajouté. Le logiciel est à ce jour un logiciel de création ou de gestion de données, permettant d'insérer, de créer, de modifier ou de supprimer des données dans une base de données spatiale. Ainsi, le logiciel peut intéresser un grande nombre de personnes, spécialistes et non-spécialistes dans le domaine des SIG ou des base de données.

## **4.4 Opérationnalisation**

### **4.4.1 Fonctionnement chaîne de traitements "fermée"**

#### **Choix des fichiers d'entrée**

L'utilisateur choisit les données d'entrée. Seulement des données au format shape (vecteur) peuvent être sélectionnées. L'utilisateur doit au moins deux données : les données correspondant aux bâtiments et les données correspondant aux surfaces aquatiques de la zone d'étude (conformément aux facteurs de risque retenus). Au maximum l'utilisateur peut choisir cinq données (correspondant aux données dont nous disposons).

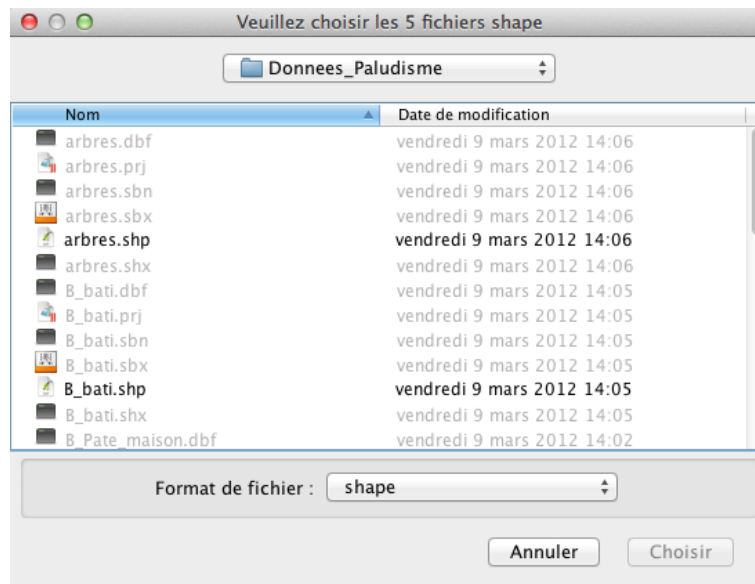


FIGURE 4.15 – Choix fichiers d'entrée

## Correspondance des couches

Par rapport aux données sélectionnées par l'utilisateur, il est nécessaire de connaître quelle information correspond à quelle donnée. L'utilisateur choisit donc par exemple que la couche "b\_bati" correspond à l'information sur les bâtiments de la zone d'étude et la chaîne gardera en mémoire cette information pour la suite des traitements.

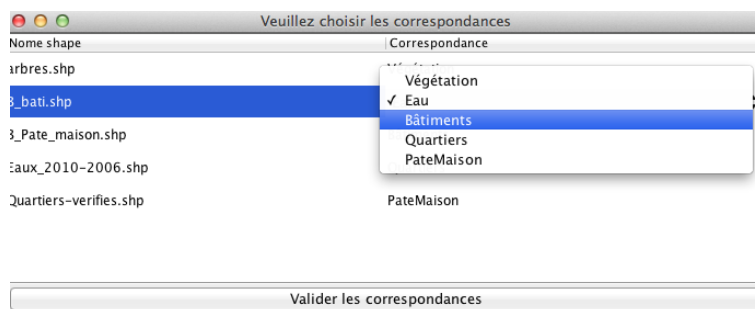


FIGURE 4.16 – Correspondances des données

## Création d'une nouvelle base de données spatiale

La chaîne de traitements crée une nouvelle base de données. Ceci nécessite un certain nombre d'informations : L'hôte sur lequel PostgreSQL / PostGIS sont installés (localhost si installation en local sur l'ordinateur), le port sur lequel PostgreSQL / PostGIS sont installés, le nom d'utilisateur pour la base de données, le nom de la nouvelle base de données, le mot de passe et la projection souhaités pour la base de données.

Information connexion b...

hote

localhost

Port

5432

Nom utilisateur

entringe

Mot de passe

.....

Nom de la base de données

test3712

Projection

32360

OK

FIGURE 4.17 – Informations création base de données

### Insertion des données vectorielles dans la base de données

A l'aide du module de transformation shp2pgsql les données vectorielles sont insérées dans la base de données. Ce module transforme les fichiers au format ".shp" en fichiers ".sql" qui sont par la suite insérés dans la base de données à l'aide de la fonctionnalité "pgsql".

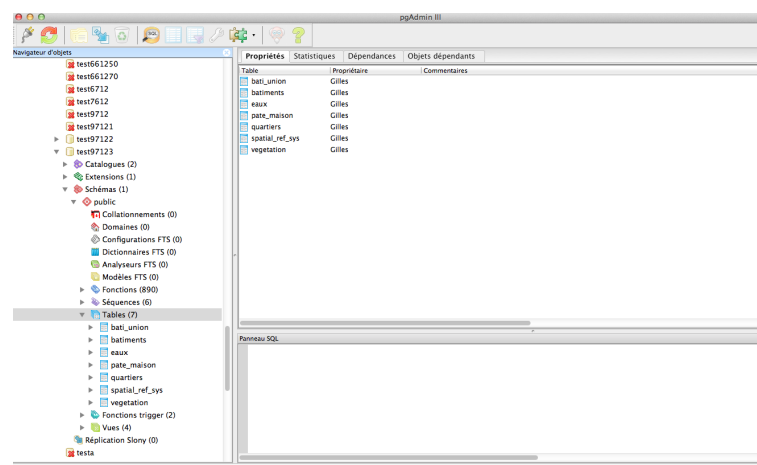


FIGURE 4.18 – Base de données avec données insérées

### Nombre d'habitants

L'outil demande à l'utilisateur le nombre d'habitants pour la zone d'étude.

FIGURE 4.19 – Nombre d'habitants

### Calcul densité de population et taille pixel

L'outil calcule automatiquement, à partir de la couche qui correspond aux bâtiments, la densité de population de la zone d'étude et la taille de cellule qu'un habitant occupe en théorie. Par exemple, pour une population de 15.000 habitants sur une superficie totale de 300.000 m<sup>2</sup>, nous effectuons le calcul suivant :

$$300.000/15.000 = 20 \text{ m}^2 \text{ par habitants}$$

$$1 \text{ habitant} = \sqrt{(20)} \text{ 1 habitant} = 1 \text{ cellule} = 4.47 * 4.47 \text{ m}$$

Un habitant correspond donc en théorie à 20 mètres carrés et à une taille de cellule théoriques de 4.47 \* 4.47 mètres.

```

INSERT 0 1
INSERT 0 1
INSERT 0 1
INSERT 0 1
INSERT 0 1
INSERT 0 1
INSERT 0 1
INSERT 0 1
INSERT 0 1
INSERT 0 1
INSERT 0 1
INSERT 0 1
INSERT 0 1
INSERT 0 1
INSERT 0 1
INSERT 0 1
INSERT 0 1
INSERT 0 1
INSERT 0 1
COMMIT
Surface: 249394.703579867
Nombre Habitants: 14133.0
Taille pixel: 4.200746106818336
4.200746106818336
test97124 Gilles
/Users/Gilles/Documents/sql

```

FIGURE 4.20 – Taille de cellule / Densité de population

### Calcul et affichage carte de vulnérabilité

A partir des différentes opérations et calculs, l'outil calcule une densité des points, affiche et stocke une carte de vulnérabilité.

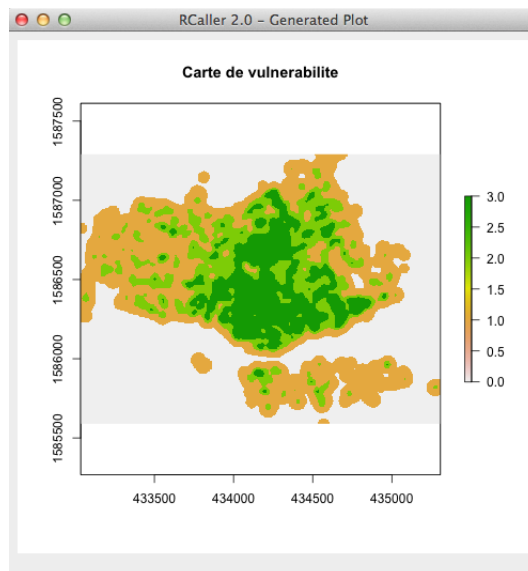


FIGURE 4.21 – Carte de vulnérabilité

### Suppression des surfaces aquatiques polluées

Par rapport aux données que nous possédons, il est nécessaire de supprimer des surfaces aquatiques polluées car les moustiques ne peuvent pas survivre dans les eaux polluées. L'outil permet donc de sélectionner quelles surfaces sont polluées à partir d'une colonne qui permet à l'utilisateur d'identifier ces surfaces.

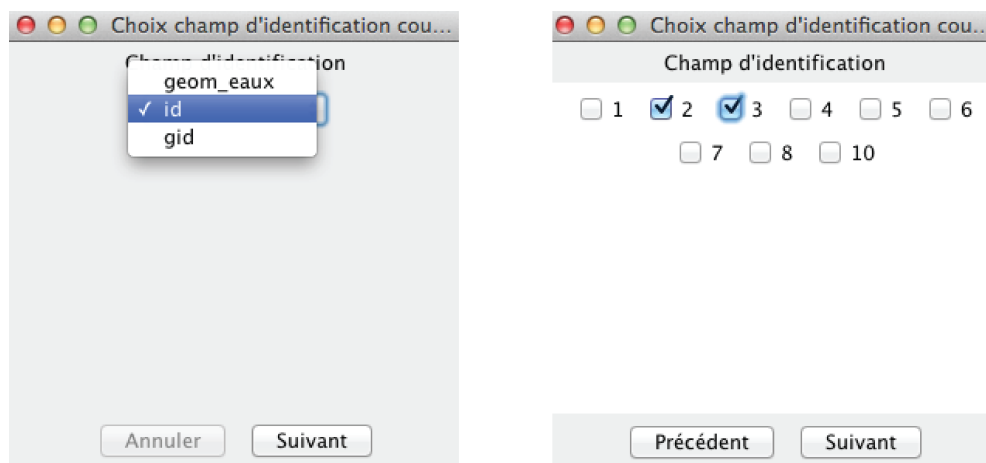


FIGURE 4.22 – Supprimer surfaces aquatiques polluées

### Carte d'aléa

La carte d'aléa est calculée en combinant plusieurs opérations et traitements d'analyse spatiale : Intersection, Création de zones tampon, Union, Extent (emprise géographique). Comme il n'y aura pas de risque dans les zones où il n'y a pas de vulnérabilité, il est intéressant de supprimer les parties concernant l'aléa qui n'intersectent pas avec la carte de vulnérabilité. Finalement nous obtenons donc une carte d'aléa. Cette carte, en combinaison avec la carte de vulnérabilité crée une carte de risque de transmission, affiché à la fin du traitement.

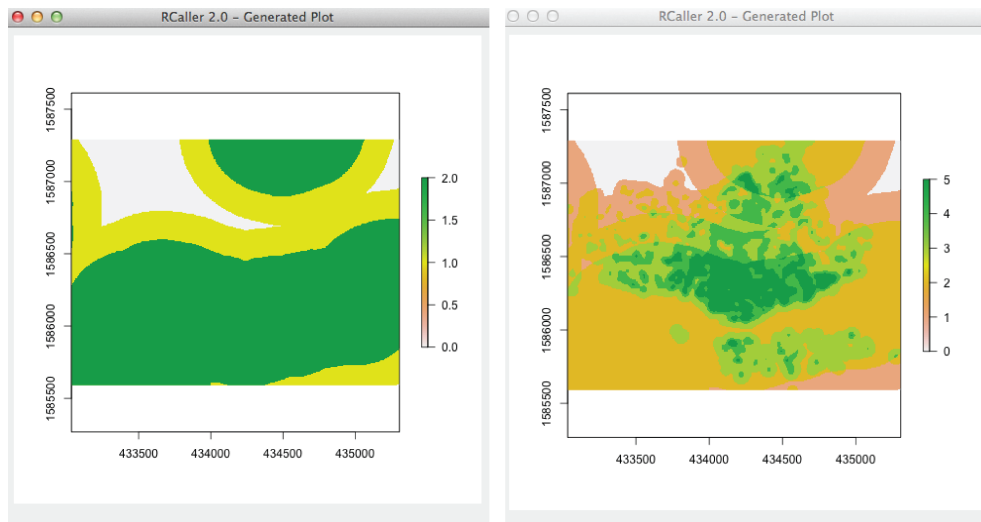


FIGURE 4.23 – Exemple de carte d'aléa et carte de risque

## 4.4.2 Fonctionnement du logiciel "ouvert"

Le logiciel ouvert utilise les mêmes librairies et bibliothèques (R, PostGis, PostgreSQL) que la chaîne de traitements. Au fur et à mesure de l'avancement du développement du logiciel, nous avons rajoutés certaines fonctionnalités que nous avons jugé important pour l'utilisation du logiciel, comme par exemple l'export des données ou la possibilité de renommer les tables de la base de données. Dans cette partie sera illustrée l'implémentation de l'outil à partir des fonctionnalités les plus intéressants et innovants.

### Choix emplacement des données

L'utilisateur dispose de trois choix possibles par rapport à l'emplacement des données qu'il souhaite traiter :

- Données dans une base de données PostGIS existante : Les données stockées dans la base sont utilisées
- Données sur disque dur : Une nouvelle base de données est créée
- Données sur disque dur et dans base de données : Des données sont insérées dans une base de données existantes

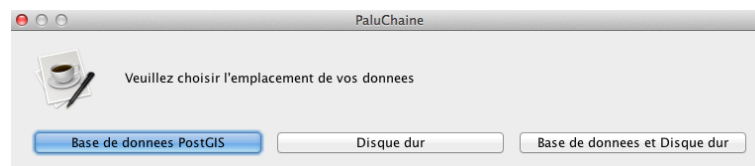


FIGURE 4.24 – Choix de l'emplacement des données

### Information connexion base de données

L'utilisateur indique les informations relatifs à la base de données. En fonction du choix précédent (emplacement des données), les données stockées dans une base de données existante sont chargés ou une nouvelle base de données est créée et les fichiers sélectionnés par l'utilisateur sont stockés dedans. De plus, l'utilisateur indique la projection de la base de données. Lors de la création d'une nouvelle base de données, toutes les données sont projetés par rapport à cette information, dans le cas d'une base de données existante, toutes les données stockées dans la base sont analysés et, si nécessaire, reprojetés.

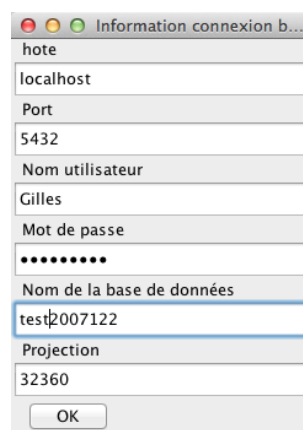


FIGURE 4.25 – Informations base de données

## Menu du logiciel

Un menu présente les différents traitements en fonction des catégories de traitements définies par avant (cf 4.2.2). L'utilisateur peut choisir indépendamment les traitements qu'il souhaite effectuer.

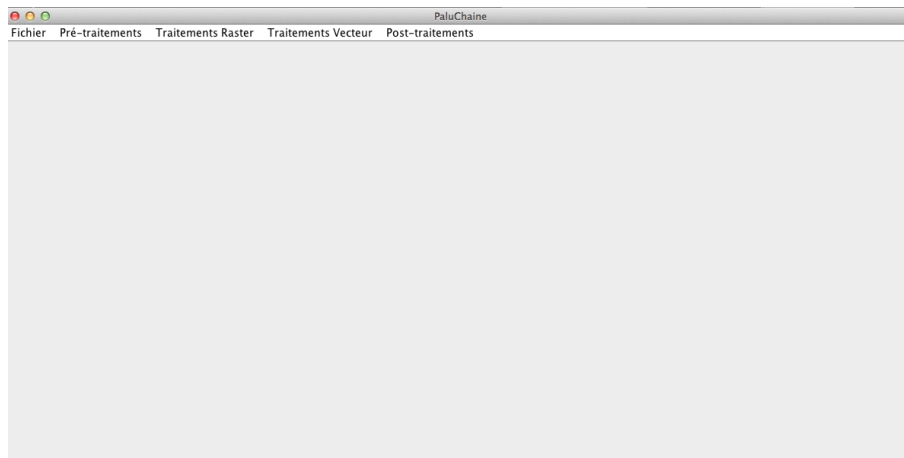


FIGURE 4.26 – Le menu de l'outil

## Les traitements possibles

En fonction des catégories de traitements définis par avant, les traitements suivants peuvent être effectués :

### **Pré-traitements :**

- Insertion des rasters et des vecteurs dans la base de données
- Supprimer des rasters et des vecteurs stockés dans la base de données

### **Traitements :**

#### *Analyse Spatiale*

- Intersection entre des couches vecteurs
- Union entre des couches vecteurs
- Création des zones tampons
- Calcul la différence entre deux couches vecteur
- Création une couche représentant l'emprise d'une couche

#### *Statistique*

- Calcul la surface totale
- Calcul la taille théorique d'une cellule (cf. Chaîne de traitements)
- Calcul une densité des points à partir d'une couche vecteur

#### *Transformation*

- Rastérisation de couches vecteurs
- Changement de la projection de la base de données

### **Post-traitements**



- Supprimer certains contenus d'une couche
- Reclassification un raster
- Combinaison de plusieurs rasters
- Renommage des données dans la base (raster et vecteur)
- Visualisation les vecteurs et les rasters

## **Spécificités techniques du logiciel**

### **Mémoire**

Le logiciel se rappelle des traitements déjà effectués. Par exemple, il n'est pas possible de calculer une taille théorique d'une cellule sans calculer la surface totale de la donnée auparavant. Lorsque l'utilisateur a calculé la taille de cellule et veut rasteriser une donnée au format vecteur, le logiciel se rappelle du calcul de la cellule effectué par avant. L'utilisateur peut bien évidemment également saisir manuellement la taille de pixel qu'il souhaite appliquer pour le traitement.

Après avoir calculé la superficie pour une donnée, elle n'apparaît plus dans la liste des données disponibles pour effectuer ce traitement. Il en est de même pour le calcul de la taille de pixel.

### **Affichage**

L'utilisateur peut afficher à tout moment les données stockées dans la base de données. Cet affichage se fait sous la forme de la fonctionnalité "plot" du logiciel R.

### **Exporter des données**

Les données stockées dans la base de données peuvent être exportées. Exemple : L'utilisateur, après avoir créé des zones tampons, peut exporter cette nouvelle donnée (au format shape) et l'utiliser dans le SIG de son choix.

### **Gestion des données**

A tout moment, des nouvelles données (vecteur ou raster) peuvent être insérées dans la base de données. Il est également possible de supprimer des données et de renommer les données (tables).

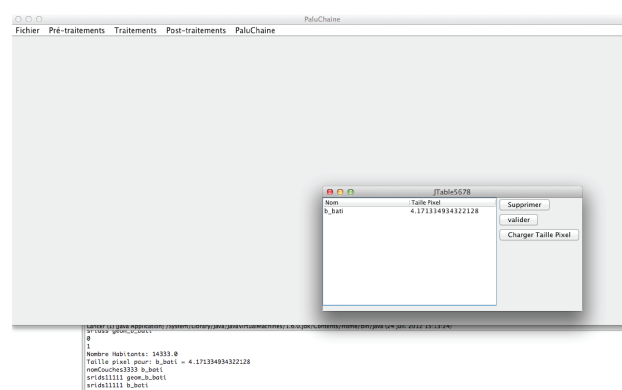
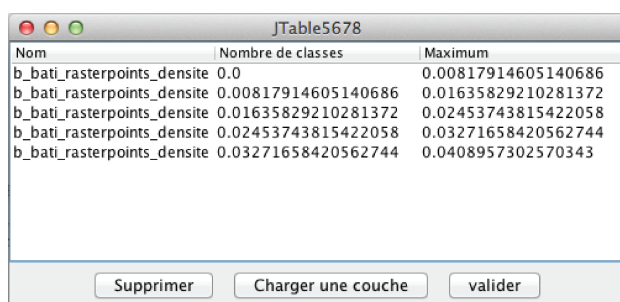
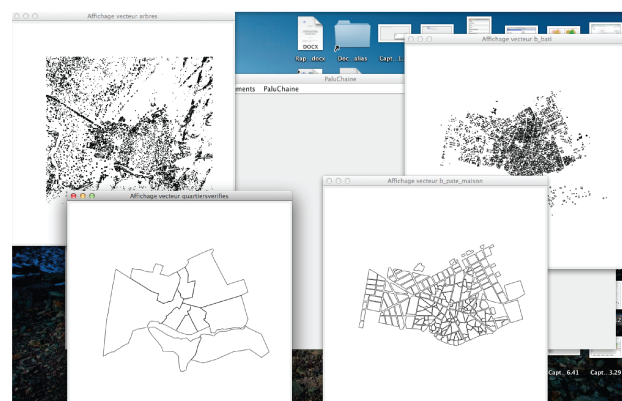
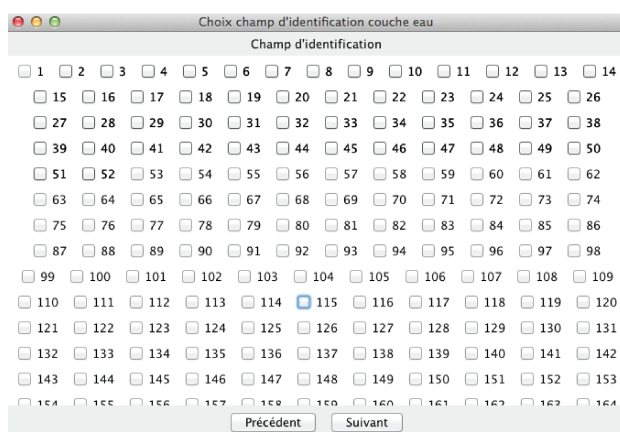
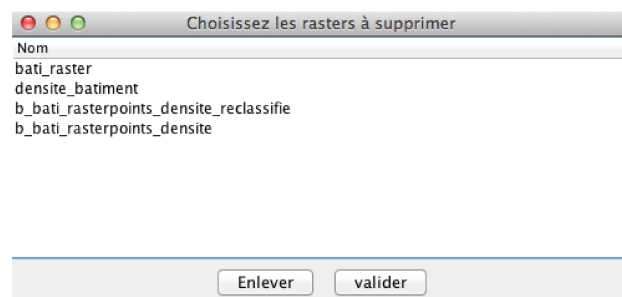
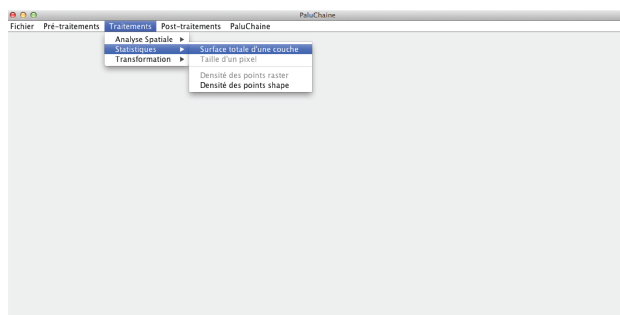
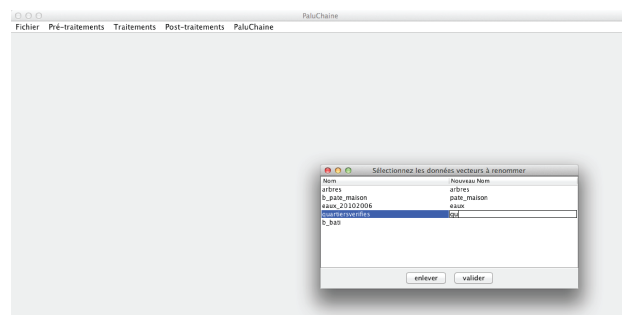
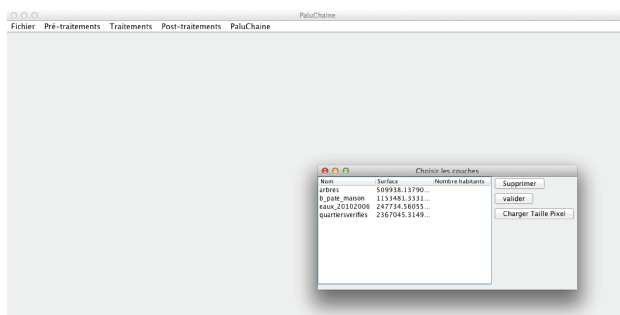


FIGURE 4.27 – Aperçus du logiciel libre

# Chapitre 5

## Résultats

### 5.1 Résultat

Nous avons développé deux outils, basés sur des plateformes Open Source. Le premier outil, une chaîne de traitements permet d'effectuer des traitements dans un ordre bien définis pour cartographier des indicateurs de risque environnemental. Le deuxième outil, un **logiciel ouvert**, donne à l'utilisateur la possibilité d'exécuter chacun des traitements de la chaîne selon ses besoins.

### 5.2 Validation des résultats

Pour valider les résultats et le fonctionnement notamment de la chaîne de traitements, nous comparons une carte de risque créée par la chaîne à une carte de risque élaborée par Nadine Dessay en utilisant le logiciel ArcGIS.

### 5.3 Difficultés rencontrés

Tout au long du développement informatique de la chaîne de traitements et du logiciel ouvert, nous avons été confrontés à de nombreuses difficultés. Dans cette partie j'expliquerai les problèmes majeurs afin de faciliter le futur développement informatique de l'outil. Les difficultés sont regroupées en trois sous-parties : Difficultés liées au développement en Java, difficultés liées à l'utilisation de PostgreSQL / PostGIS et difficultés liées au logiciel R.

#### 5.3.1 Java

Sur internet, il existe un nombre infini de forum et de tutoriels sur tous ce qui est programmation en Java. Ainsi, j'ai pu m'approprier de nouvelles connaissances dans ce domaine. Néanmoins ce qui a été très compliqué, était de comprendre le fonctionnement de Rcaller, permettant d'utiliser les fonctionnalités de R sous Java. Il existe en effet que très peu de documentations et informations sur internet pour cette extension de Java.

La prise en main de Rcaller a nécessité un certain temps. Par exemple, nous avons dû faire face au problème que lors de l'affichage d'une donnée (plot), la fermeture d'une fenêtre causait automatiquement la fermeture de tout le programme. Pour résoudre ce problème, le résultat (l'image affichée) doit être stockée temporairement et par la suite être affichée dans une nouvelle fenêtre.

Une autre difficulté a été de comprendre comment intégrer des variables Java dans le code "R".

Par exemple, en ce qui concerne la reclassification d'une image raster, l'utilisateur peut choisir le nombre de classes et les valeurs minimales et maximales de chaque classe. La commande "addDoubleArray" de RCaller permet d'intégrer un tableau composé de valeurs du type double dans le code R. En fonction des choix de l'utilisateur, Rcaller crée à l'intérieur du code R une "matrice" comportant les valeurs du tableau Java.

Il est également possible de récupérer les résultats de calculs effectués dans R. On peut seulement récupérer le résultat de la dernière ligne du code R. Par exemple, si on calcule deux moyennes différentes, il est nécessaire de stocker les deux moyennes dans une "liste" (moyenne <- list(moyenne1,moyenne2) sous R). Avec la commande "runAndReturnResult (moyenne)" il est par la suite possible de récupérer le résultat des calculs stocké dans la liste "moyenne".

Ces illustrations sont seulement quelques-unes des nombreuses difficultés que nous avons rencontrées en utilisant RCaller. Finalement il faut préciser que l'utilisation de cette extension est en général très lente, notamment si l'utilisateur ne dispose pas d'une machine performante.

### 5.3.2 PostgreSQL/PostGIS

En ce qui concerne l'utilisation de PostGIS et de PostgreSQL, ce qui a été le plus compliqué était certainement l'installation. Lorsque nous avons démarré le développement informatique de la chaîne de traitements, nous utilisions une version antérieure de PostgreSQL et de PostGIS. Le 3 avril 2012 est sortie PostGIS 2.0 avec comme principale nouveauté la gestion des images rasters. Le passage à la nouvelle version a été assez compliqué, notamment sous le système d'opération Linux. En effet, toutes les librairies utilisées par PostGIS comme GDAL, PROJ etc. ont dû être mises à jour, tout comme la version de PostgreSQL (minimum 8.4). Sous Windows et MacOS ceci a posé moins de problèmes car des exécutables d'installation sont disponibles sur internet et les anciennes versions peuvent être supprimées plus facilement.

Souvent il n'est pas possible d'effectuer des opérations sur des données spatiales stockées dans la base de données, probablement parce qu'il y a eu un erreur lors de la construction de la donnée. Pour résoudre ce problème, il est nécessaire pour chaque traitement PostGIS (par exemple Intersection entre deux couches) de créer une donnée "temporaire" à l'aide la fonctionnalité PostGIS "ST\_Buffer". Cette fonctionnalité crée théoriquement des zones tampons. En créant des zones de tampons de taille zéro, la fonctionnalité permet également de résoudre ce genre de problème.

### 5.3.3 R

Par rapport à l'utilisation de R, nous avons été confrontés à une difficulté majeure qui est l'utilisation de la librairie "rgdal". Sous Linux, il a été assez facile d'installer cette librairie à l'aide de la commande R "install.packages("rgdal")", néanmoins il est important de vérifier que lors de l'installation les versions de GDAL, PROJ et GEOS soient mises à jour car sinon un certain nombre de fonctionnalités risquent de ne pas fonctionner.

Sous MacOS, il est nécessaire de faire l'installation à partir de la source du package. Sur le site <http://www.r-bloggers.com/installing-rgdal-on-mac-os-x-2/> sont expliquées les informations nécessaires pour cette opération.

Sous Windows, l'installation de rgdal est facile et se fait à l'aide de la commande "install.packages". Malheureusement, il est impossible d'influencer la version GDAL utilisée lors de l'installation du package. Ainsi, il est actuellement impossible de charger des données dans R à partir d'une base de données PostgreSQL en utilisant rgdal. Ainsi, sous Windows, la chaîne de traitements ne fonctionne pas encore et le logiciel libre présente des fonctionnalités limitées (tous les traitements liés à l'utilisation de R ne fonctionnent pas).

## 5.4 Perspectives

L'objectif principal du stage était la conceptualisation et le développement informatique d'une chaîne de traitements permettant de cartographier le risque de transmission du paludisme. La définition des facteurs de risque de transmission du paludisme a été une autre étape indispensable de ce travail.

Le modèle conceptuel ainsi que les listes concernant les différents facteurs de risque de transmission permettent aux experts, mais également à des non-experts, de comprendre le cycle du paludisme, comment le paludisme se développe.

A partir de ce modèle conceptuel, nous avons pu définir les traitements nécessaires pour cartographier le risque. Par la suite, de nombreuses réflexions autour de l'architecture informatique ont été menées. L'objectif était que la chaîne soit réutilisable pour le plus grand nombre de personnes. Créer un logiciel basé sur une architecture OpenSource et libre était donc indispensable. Nous avons décidé de développer la chaîne en Java. Par la suite, nous avons recherché pour chaque traitement les librairies et outils nécessaires. Très rapidement, nous nous sommes aperçu que PostgreSQL, PostGIS (traitements d'analyse spatiale) et R (notamment pour les traitements statistiques et les traitements raster) permettent d'effectuer tous les traitements que nous devons enchaîner.

Au bout de quatre mois, le développement informatique a abouti à une première chaîne de traitements. Cette chaîne de traitements effectue un certain nombre de traitement dans un ordre bien précis, l'utilisateur n'a aucune influence sur le déroulement des traitements.

Avec les différents encadrants, nous avons décidé qu'il serait intéressant de développer, en se basant sur les traitements de la chaîne de traitements, un logiciel qui permettra d'effectuer chacun des traitements indépendamment.

Le logiciel qui a été développé permet d'imaginer de nombreuses perspectives. A court terme, il peut être très intéressant de rédiger un mode d'emploi pour l'outil. Aussi, serait-il intéressant d'améliorer l'affichage des données (avec Grass par exemple) et d'améliorer la gestion. La possibilité de se connecter à une base de données externe (hébergé sur un serveur internet) permet d'imaginer l'utilisation de l'outil dans de nombreux projets. Beaucoup de gens sont intéressés par les fonctionnalités de PostGIS, mais ne disposent pas des connaissances nécessaires pour son utilisation. Le logiciel leur permet dans un premier temps de créer une base de données, d'insérer des vecteurs et des rasters dans cette base et d'effectuer un certain nombre de traitements sur ces données. A long terme, l'intégration de nouveaux traitements pourrait faire de ce logiciel un outil intéressant pour les gens souhaitant d'effectuer des traitements d'analyse spatiale de base (intersection, union, zones tampon etc.).

A long terme, il peut être intéressant de créer un équivalent au "ModelBuilder" d'ArcGIS. L'utilisateur pourrait donc créer ses propres chaînes de traitements en fonction de ses besoins.

En plus, il est envisageable que ces chaînes soient réutilisables.

## 5.5 Conclusion

Au terme des six mois de stage, la collaboration avec des experts dans des domaines divers (informatique, environnement, environnement-santé, biologie etc.) m'a permis d'acquérir un grand nombre de connaissances et de compétences et de travailler sur une multitude de thématiques.

L'élaboration d'un modèle conceptuel du paludisme a nécessité une recherche bibliographique importante ainsi qu'une longue phase de réflexion. Ce travail était donc très scientifique et conceptuel.

Dans un deuxième temps, le développement d'une chaîne de traitements respectivement d'un logiciel ouvert m'a permis d'approfondir mes connaissances dans plusieurs domaines de l'informatique : Gestion des base de données (spatiales), R et surtout programmation en Java sans oublier bien évidemment la conceptualisation et toutes les étapes de réflexion de la chaîne en amont.

Toutes les étapes, les erreurs et les leçons tirées de ce stage m'ont beaucoup aidé à évoluer dans le monde des SIG et plus particulièrement dans le domaine de l'informatique.

# **Chapitre 6**

## **Discussion**

# Chapitre 7

## Glossaire et définitions

SGBD Quickbird SRID :Le SRID est un entier qui identifie de façon unique le système de références spatiales (par exemple le SRID 32360 correspond à ...)

diagrammes de classes

GDAL

anophèle

indices environnementaux spatialisés

logiciel ouvert



# Chapitre 8

## Annexes

### 8.0.1 Définitions, librairies et bibliothèques utilisées

#### Traitement

Un traitement est un outil qui permet de transformer une information ou une donnée fournie en entrée et livrée en sortie. Comme nous nous intéressons aux traitements informatiques, nous nous intéresserons principalement à la partie logicielle (software) de ces traitements. "Un traitement réalise une fonctionnalité. Pour être précis, il faudrait dire qu'un traitement spécifie la façon de réaliser une fonctionnalité. Cette spécification peut être exprimée en langage informatique." ((Abd-el Kader, 2006))

Dans ce mémoire, nous excluons donc une partie des traitements comme la saisie et la création de données (par exemple à partir d'une image satellite). Nous nous intéresserons aux traitements informatiques géographiques nécessaires pour créer une carte du risque. Les traitements informatiques géographiques regroupent les traitements qui manipulent les données géographiques.

#### Java

Java est un langage orienté objet, c'est-à-dire que le programme est vu comme un ensemble d'entités (de classes). Au cours de l'exécution du programme, les entités collaborent entre elles pour arriver à un but commun.

#### Eclipse

L'environnement de programmation en Java le plus connu est le projet "Eclipse" de la fondation Eclipse. Ce logiciel simplifie la programmation grâce à un certain nombre de raccourcis et notamment grâce à la possibilité d'intégrer de nombreuses extensions. Au fur et à mesure de l'avancement du code, Eclipse compile automatiquement le code et signale les problèmes qu'il détecte.

#### PostgreSQL

PostgreSQL est un système de gestion de bases de données relationnelles objet (Manuel PostgreSQL). PostgreSQL est un outil Open Source et disponible gratuitement, compatible avec les systèmes d'opérations les plus connus (Linux, Unix (Mac OSX, Solaris etc.) et Windows). PostgreSQL propose des interfaces de programmations pour des langages de programmation comme Java, C++, Python etc. Le développement de PostgreSQL a débuté en 1986 (appelé à l'époque Postgres). En 1995, les développeurs ajoutent un interpréteur de langage SQL à l'outil. A partir de 1996, l'outil s'appelle PostgreSQL afin de souligner le lien entre Postgres et le

langage SQL. PostgreSQL peut être facilement étendu par l'utilisateur en ajoutant de nouvelles fonctions, de nouveaux opérateurs ou même de nouveaux langages de procédure.

## **PostGIS**

PostGIS est une extension du système de gestion de base de données PostgreSQL qui permet de stocker des données (objets) géographiques dans la base de données. Cette extension permet d'utiliser une base de données PostgreSQL comme une base de données dans n'importe quel projet SIG. Depuis avril 2012, la dernière version de PostGIS (PostGIS 2.0) offre de nombreuses améliorations. La nouveauté la plus "innovante" est celle que PostGIS gère désormais les données raster (données images). PostGIS est compatible avec des nombreux autres outils comme par exemple Mapserver.

## **JDBC**

JDBC est une API (Application Programming Interface) java. JDBC est un nom déposé et non un acronyme, en général on lui donne la définition suivante : Java DataBase Connectivity. (<http://java.developpez.com/faq/jdbc/...>) Cette API est constituée d'un ensemble d'interfaces et de classes qui permettent l'accès, à partir de programmes java, à des systèmes de gestion de bases de données relationnelles (par exemple PostgreSQL).

## **R**

R est un langage et un environnement pour des calculs et des graphiques statistiques qui fournit une grande variété de techniques statistiques. R est Open Source et disponible gratuitement. R ne fonctionne pas sous le principe du "clique bouton" mais doit être compris comme un langage informatique. Un grand nombre d'extensions (packages) est disponible et permet d'utiliser R dans des domaines très divers. Ces extensions permettent également d'appeler R à partir d'autres langages de programmation.

## **RCaller**

RCaller est une extension de Java permettant d'appeler le logiciel R et d'utiliser les fonctionnalités de R dans un programme Java. Cet outil est particulièrement intéressant comme il permet d'appeler les librairies de R comme par exemple rgdal et donc d'utiliser les fonctionnalités de GDAL dans notre chaîne de traitements.

## **GDAL**

GDAL est une bibliothèque libre permettant de lire et de traiter les fichiers raster (images) à partir des langages de programmation de haut niveau (C++, Java, Perl, Python ou le langage R). Un sous-ensemble de GDAL est OGR qui permet de gérer les données vectorielles. La fonctionnalité principale de GDAL et d'OGR est la conversion entre plusieurs formats et permet donc d'assurer la compatibilité entre les différents formats existants.

## **Packages R**

*rgdal*

Rgdal est une extension de R et permet d'utiliser les fonctionnalités de la bibliothèque GDAL à partir du langage de programmation R. Cette extension permet donc tout d'abord

de gérer et de représenter des fichiers spatiaux, mais également d'effectuer des traitements statistiques sur les objets géographiques.

### *RPostgreSQL*

RPostgreSQL fournit une interface de gestion permettant d'accéder aux bases de données PostgreSQL et au driver PostgreSQL pour R. Cet outil est particulièrement intéressant car il permet de se connecter à la base de données spatiale de la chaîne de traitements et d'y récupérer les données nécessaires pour les traitements sous R.

### *spatstat*

Spatstat est un libraire de R permettant de traiter et d'analyser des données spatiales. Ainsi, il permet par exemple de calculer une densité des points.

### *raster*

Raster est un paquet de R permettant de traiter les images raster. Ce paquet permet également de créer des fichiers raster à partir de données d'entrée de formats diverses.

### *maptools*

Paquet permettant de lire et de traiter des données géographiques comme par exemple des données au format shape (".shp").

## 8.0.2 Les traitements effectués

Pour chaque traitement effectué, la chaîne utilise différentes librairies et outils. Dans cette partie du travail, chaque traitement sera expliqué en détail et quelles opérations informatiques ont été utilisées pour effectuer les différents traitements. Les traitements sont divisés en cinq catégories cf (4.7) :

- pré-traitements
- traitements statistiques
- traitements de fichiers raster
- traitements de fichiers vecteurs
- post-traitements

### Pré-traitements

#### Modification du système de coordonnées

L'utilisateur peut choisir individuellement les données d'entrée (ici les fichiers de forme Végétation, Bâtiments et Eau) en fonction de ses besoins. Il est donc nécessaire d'homogénéiser les systèmes de coordonnées de ces données. Ceci est fait lors de la transformation des fichiers vecteurs (shape), à l'aide de la commande `shp2pgsql`.

**Librairies utilisées :** PostgreSQL, PostGIS, JDBC

**Données en entrée :** Fichiers de forme géoréférencés (vecteur)

**Données temporaires :** Fichiers .sql

**Données en sortie :** Données avec système de coordonnées dans base de données

---

#### Algorithme 1: `shp2pgsql`

---

Syntaxe : `shp2pgsql -d -g nomGeom -s SRID fichier.shp nomTable -> fichier.sql`

---

Pour la modification de la projection, c'est l'option `-s SRID` de la commande "`shp2pgsql`" qui va effectuer cette opération. Cette option définit le système de coordonnées ou, s'il existe déjà, le modifie. A l'intérieur de la chaîne de traitements, cette option est définie une seule fois lors de la création de la base de données et toutes les données vectorielles insérées dans la base de données auront le même système de coordonnées afin de garantir l'homogénéité de l'ensemble des données.

#### Transformation du format de fichier

Afin de pouvoir insérer des fichiers de forme vecteur dans une base de données spatiale il est nécessaire de les transformer. Cette transformation est également effectuée à l'aide du module de transformation `shp2pgsql` intégré dans PostGIS. L'option `-d` de cette fonction indique au module de supprimer la table si elle existe déjà dans le base de données et de la recréer par la suite. Cette option permet donc de garantir que les données en entrée seront insérées correctement dans la base de données. L'option `-g` permet de choisir le nom de la colonne géométrique (créée automatiquement pour toute couche géométrique insérée dans une base de données PostgreSQL / PostGIS) ce qui facilitera les traitements par la suite. L'option `-s` a été détaillée ci-dessus et permet de choisir le système de coordonnées de la donnée. L'option "`nomtable`" attribue le nom de la table dans la base de données correspondant à la donnée d'entrée. Le paramètre "`fichier.sql`" indique l'emplacement du fichier ".sql" en sortie qui sera par la suite utilisé pour insérer les

données dans la base de données.

**Librairies utilisées :** PostgreSQL, PostGIS, JDBC

**Données en entrée :** fichiers de forme géoréférencés (vecteur)

**Données en sortie :** Fichiers au format .sql

---

**Algorithme 2:** shp2pgsql

---

Syntaxe : shp2pgsql -d -g nomGeom -s SRID fichier.shp nomTable -> fichier.sql

---

La fonctionnalité "ogr2ogr" de la bibliothèque GDAL peut également être utilisée, notamment pour des conversions vers d'autres formats que le format ".sql".

### Création base de données

Il existe plusieurs moyens pour créer une base de données. Nous utilisons la méthode suivante : Connexion à la base de données de maintenance "postgres" et création d'une nouvelle base de donnée. Attention, cette base de données n'est pas encore compatible avec les données spatiales (géographiques).

**Librairies utilisées :** PostgreSQL, PostGIS, JDBC

**Données en entrée :** /

**Données en sortie :** Base de données PostgreSQL

---

**Algorithme 3:** create database

---

Syntaxe : CREATE DATABASE nomBase

---

### Initialisation de la base de données aux fonctionnalités PostGIS

Il est nécessaire d'inclure 3 fichiers dans la base de données afin de la rendre compatible avec les fonctionnalités PostGIS : postgis.sql, spatial\_ref\_sys.sql, rtpostgis.sql.

- postgis.sql : Permet de charger dans la base de données, les types, les fonctions géographiques et les tables associées qui sont utilisées avec PostGIS.
- spatial\_ref\_sys.sql : Permet de charger dans la base de données, la table spatial\_ref\_sys qui contient environ 300 systèmes de référence spatiales et les détails nécessaires à leur transformation.
- rtpostgis.sql : Permet de charger dans la base de données les fonctionnalités nécessaire pour qu'elle supporte les fichiers raster.

**Librairies utilisées :** PostgreSQL, PostGIS, JDBC

**Données en entrée :** Base de données PostgreSQL

**Données en sortie :** Base de données PostgreSQL compatible PostGIS

---

**Algorithme 4:** Initialisation base de données PostGIS

---

Syntaxe : psql -d nomBase -f postgis.sql (pareil pour spatial\_ref\_sys.sql et rtpostgis.sql)

---

## Traitements statistiques

### Calcul de la surface totale d'une couche de données

La surface totale d'une couche de données est calculée à l'aide de la fonction PostGIS suivante : ST\_AREA(geom\_couche). Cette fonction calcule la surface totale pour chaque enregistrement d'une colonne géométrique d'une table. L'unité correspond automatiquement à l'unité du système de coordonnées de la base de données. A partir de ce calcul on obtient facilement la surface totale à l'aide d'une requête SQL classique (select SUM...).

---

**Algorithme 5:** Calcul surface

---

Syntaxe : SELECT ST\_AREA(geom\_table) from table as somme

---

### Calcul densité de population et taille pixel

Dans les zones de risque, les habitations sont généralement des habitations sur un seul étage. Ainsi, il est possible de calculer à partir de la surface totale des bâtiments et de la population totale de la ville, la superficie qu'occupe en théorie un seul habitant.

Par exemple, pour une population de 15.000 habitants sur une superficie totale de 300.000 m<sup>2</sup>, nous effectuons le calcul suivant :

$$300.000/15.000 = 20 \text{ m}^2 \text{ par habitants.}$$

Un habitant correspond donc en théorie à 20 mètres carrés. A partir de ce calcul, nous pouvons déterminer la taille théorique d'un pixel. Pour les données d'essai utilisées, ce sont donc les calculs suivants qui sont effectués lors du lancement de la chaîne de traitements :

$$14\ 133 \text{ habitants} = 249394,70358 \text{ mètres carrés}$$

$$1 \text{ habitant} = 17,646 \text{ mètres carrés}$$

$$1 \text{ pixel} = 4.2 \text{ m} * 4.2 \text{ m}$$

Aucune fonctionnalité particulière n'est nécessaire pour ce calcul statistique, Java ou R peuvent facilement effectuer ces calculs de base.

## Traitements fichiers Vecteur

### Insertion fichiers de forme dans une base de données spatiale

Pour l'insertion d'un fichier de forme dans une base de données les informations suivantes sont nécessaires : le serveur (en général localhost), le port sur lequel la base de données est installé (en général 5432), le nom de l'utilisateur qui est le propriétaire de la base de données, le nom de la base de données dans laquelle seront insérées les fichiers de forme et enfin le fichier ".sql" qui comporte les informations relatives au fichier de forme transformé par avant.

**Librairies utilisées :** PostgreSQL, PostGIS, JDBC

**Données en entrée :** Fichiers au format .sql(vecteur)

**Données en sortie :** Table dans la SGBD

---

**Algorithme 6:** psql

---

Syntaxe : psql -h serveur -p port -U Utilisateur -d nomBase -f fichier.sql

---

## Buffer

Afin de pouvoir déterminer et représenter l'aléa lié au risque, il est nécessaire de créer des buffers (zones tampons) autour des points d'eau. Ce traitement est effectué à l'aide de la fonctionnalité Postgis "ST\_Buffer". Les informations obtenues sont stockées dans une nouvelle table à l'intérieur de la base de données, ce qui permet en même temps de créer une nouvelle table de données dans la base de données. A noter qu'il faut utiliser la colonne géométrique de la couche par rapport à laquelle on veut calculer les buffers.

**Librairies utilisées :** PostgreSQL, PostGIS, JDBC

**Données en entrée :** Table géométrique dans une BDS

**Données en sortie :** Table géométrique dans une BDS

---

### Algorithme 7: Buffer 400 mètres

---

Syntaxe pour un buffer de 400 mètres autour de la couche eau : create table  
buffer\_eaux400 as select gid, st\_buffer(geom\_eaux,400) from eaux ;

---

Cette opération crée des zones tampons autour de chaque point d'eau de la couche de source et donc plusieurs géométries différentes.

## Union

La création de buffers crée plusieurs géométries différentes à l'intérieur d'une table. Pour la suite des traitements il est nécessaire de créer une seule géométrie à partir des différentes géométries. Il faut donc disposer d'une géométrie unique pour la table buffer\_eaux400 par exemple. Cette opération est effectuée à l'aide de la fonctionnalité "ST\_UNION" de Postgis. Le résultat de cette opération est stocké dans une nouvelle table de la base de données.

**Librairies utilisées :** PostgreSQL, PostGIS, JDBC

**Données en entrée :** Table géométrique dans une BDS

**Données en sortie :** Table géométrique dans une BDS

---

### Algorithme 8: Union de plusieurs géométries

---

Syntaxe : create table tableNom as select st\_union(st\_geom) from table ;

---

## Intersection

La création de zones tampons crée logiquement des géométries se situant à l'extérieur de notre zone d'étude. Pour enlever cette partie d'information, qui va seulement ralentir le déroulement de notre chaîne de traitements, nous l'enlevons en utilisant la fonctionnalité Postgis "ST\_INTERSECTION". Nous créons une intersection des géométries des buffers par rapport à "l'emprise" de notre zone d'étude.

**Librairies utilisées :** PostgreSQL, PostGIS, JDBC

**Données en entrée :** Table géométrique dans une BDS

**Données en sortie :** Table géométrique dans une BDS

---

### Algorithme 9: Intersection de plusieurs géométries

---

Syntaxe : create table raster4001 as select st\_intersection(st\_geom,st\_envelope) from  
table1,extend ;

---

## Différence entre deux couches géométriques

Il peut être nécessaire de calculer la différence entre deux géométries. La fonctionnalité à utiliser dans ce cas est la fonctionnalité Postgis "ST\_SymDifference".

**Librairies utilisées :** PostgreSQL, PostGIS, JDBC

**Données en entrée :** Table géométrique dans une BDS

**Données en sortie :** Table géométrique dans une BDS

---

### Algorithme 10: Différence entre deux géométries

---

Syntaxe : `select ST_SymDifference(a.st_geom,b.st_geom) from table1 a, table2 b`

---

## Traitements fichiers raster

### Rastérisation d'un fichier de forme

Afin de traiter la couche "batiments" il est nécessaire de la transformer en raster. Ceci est fait en utilisant la commande PostGIS "select ST\_AsRaster" qui transforme la colonne géométrique d'une table en colonne de type "raster". Au stade de développement actuel de Postgis, il est indispensable que la nouvelle colonne s'appelle "rast", sinon les autres fonctionnalités (rgdal par exemple) n'arrivent pas à lire et à traiter l'information. La taille du pixel calculée par avant est également réutilisée et chaque pixel du raster aura la taille calculée par avant.

**Librairies utilisées :** PostgreSQL, PostGIS, JDBC

**Données en entrée :** Fichiers au format .sql(vecteur)

**Données en sortie :** Table dans la SGBD

---

### Algorithme 11: Transformation table batiments en table raster

---

Syntaxe : `CREATE TABLE bati_raster as SELECT`

`ST_asraster(geom_batiments,taillePixel,-taillePixel) as rast from batiments")`

---

Cette opération peut également être effectuée sous R à l'aide de la librairie "raster" et de la fonction "rasterize". Les tests ont néanmoins démontré que Postgis est beaucoup plus rapide pour faire ce genre de traitement.

## Emprise d'une couche

Afin de délimiter d'autres données et informations par rapport à notre zone d'étude, il est nécessaire de connaître l'étendu de la zone. La fonctionnalité Postgis "ST\_ENVELOPE" effectue cette opération. En sortie, nous obtenons une géométrie qui représente (sous forme de rectangle) l'emprise d'une couche. Par la suite nous pouvons intersecter les autres données par rapport à cette géométrie. (cf 9 )

**Librairies utilisées :** PostgreSQL, PostGIS, JDBC

**Données en entrée :** Table dans une BDS(vecteur)

**Données en sortie :** Table dans une BDS

---

### Algorithme 12: transformation table géométrique en table raster

---

Syntaxe : `create table extend as SELECT st_envelope(rast) from densite_batiment`

---



## Chargement d'une donnée raster dans R à partir d'une base de données spatiale

Pour charger une donnée raster stockée dans une base de données spatiale nous utilisons la fonctionnalité "readGDAL" de la librairie "rgdal" du logiciel R.

**Librairies utilisées :** PostgresSQL, PostGIS, R, rgdal, RPostgreSQL

**Données en entrée :** Table dans une BDS(vecteur)

**Données en sortie :** Variable R (raster)

---

**Algorithme 13:** Charger données raster dans R à partir d'une base de données spatiale

---

Syntaxe : `layers <- readGDAL("PG :host="hote" user="username" dbname="nomBase" password="mot de passe" port="port" table=coucheRaster"`

---

La même opération peut être effectuée avec des données au format vectoriel et la fonctionnalité "readOGR" de la même librairie R.

Illustration de la couche "batiments" chargée à partir de la base de données PostgresSQL / Postgis :

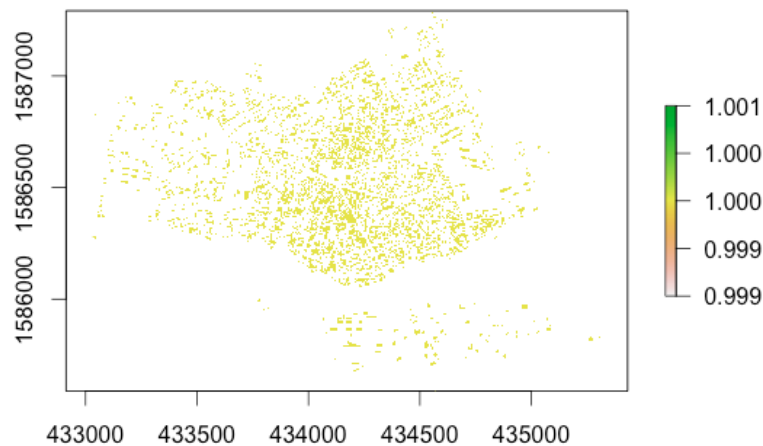


FIGURE 8.1 – Raster chargé dans R

## Transformation d'une couche raster en points

Afin de représenter la densité de population, nous allons calculer une densité des points. Il est donc nécessaire de transformer le raster (pour lequel un pixel correspond aux calculs effectués par avant) en points. Chaque pixel correspondra donc à un point. Le logiciel R avec sa librairie "raster" permet d'effectuer ce genre d'opération grâce à la commande "rasterToPoints".

**Librairies utilisées :** R, RCaller, rgdal

**Données en entrée :** Table raster dans BDS

**Données en sortie :** Fichier de points dans R

---

**Algorithme 14:** Transformation d'une couche raster en points

---

Syntaxe : `rasterToPoints(raster,fun=function(x)x>0,spatial=TRUE)`

---

L'opération nécessite tout d'abord une variable R au format R (chargée depuis la base de donnée spatiale dans notre cas). Par la suite la fonction "fun=function(x)x>0" permet de ne pas

prendre en compte tous les pixel "vides".

Illustration de la couche "batiments" chargée à partir de la base de données PostgreSQL / Postgis

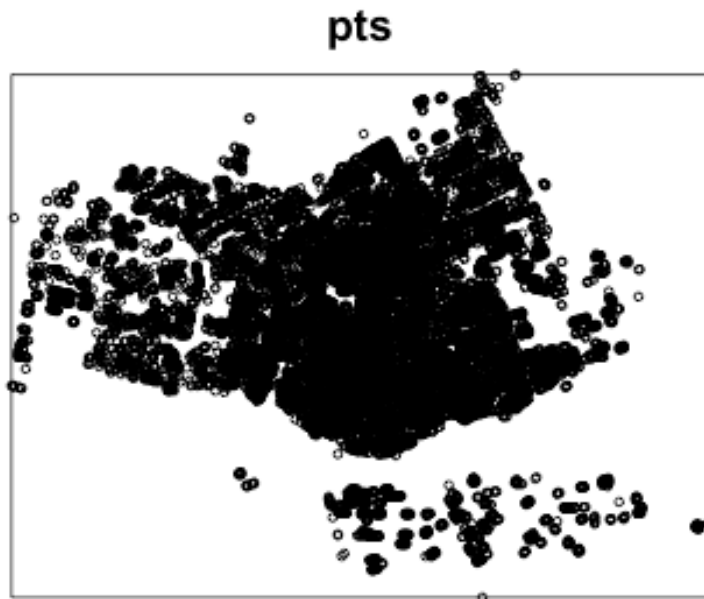


FIGURE 8.2 – Raster transformé en points

### Calcul densité des points

Pour calculer la densité des points, nous utilisons la librairie "spatstat" de R. 2 opérations différentes sont nécessaires : D'abord il faut transformer la variable au format points vers le format compatible spatstat qui est "ppp (point pattern)". Par la suite on peut calculer la densité à partir de cette variable à l'aide de la commande "density.ppp".

**Librairies utilisées :** R, spatstat

**Données en entrée :** Variable R (points)

**Données en sortie :** Variable R (points pattern, densité des points)

---

#### Algorithme 15: Densité des points

---

Syntaxe : `pts <- as(points,ppp)`

`dsty <- density.ppp(pts,eps=taillePixel,taillePixel ,adjust=taillePixel)`

---

L'argument "eps" permet de définir la taille des pixels de la variable en sortie. Si on n'ajoute pas cet argument, R calcule automatiquement la taille des pixels en sortie. Il est donc important d'indiquer la taille de pixel calculée par avant pour obtenir le résultat souhaité. L'argument "adjust" permet à l'utilisateur de déterminer le "lissage du pixel en sortie". La fonctionnalité "density.ppp" offre un large nombre de possibilité pour régler ce paramètre. Plusieurs essais ont montré que le plus facile est d'utiliser l'argument "adjust" qui est une combinaison entre toutes les autres possibilités. Cet argument va par exemple multiplier le "sigma" automatiquement par cette valeur. A nouveau, il faut attribuer à cet argument comme valeur la taille du pixel calculée par avant.

Illustration des résultats selon les arguments utilisés :

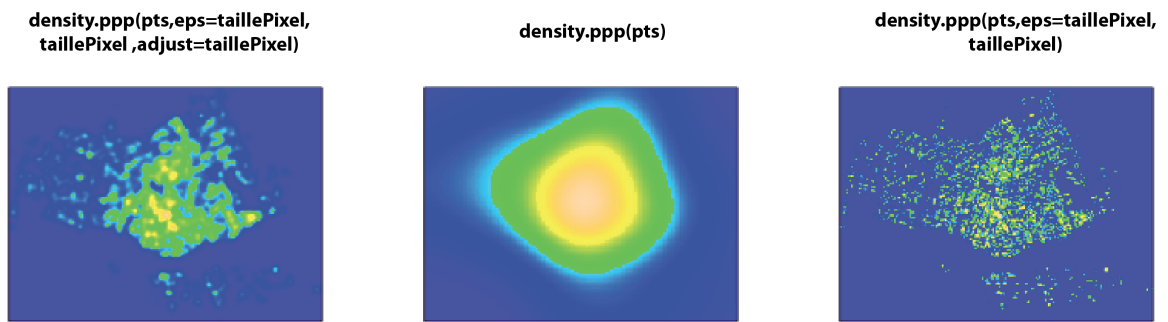


FIGURE 8.3 – Densités des points selon les paramètres

### Rastériser une couche de points

Afin de pouvoir exporter une variable R représentant une densité de points (la variable R est au format "ppp") il est nécessaire tout d'abord de la transformer vers le format "SpatialGridDataFrame". Par la suite, il est facilement possible de transformer ce format en format raster à l'aide de la fonction "raster(...)".

**Librairies utilisées :** R, spatstat, raster

**Données en entrée :** Variable R (densité de points)

**Données en sortie :** Variable R (raster)

---

#### Algorithme 16: Transforamtion points en raster

---

```
Syntaxe : ab <- as(densitePoints,"SpatialGridDataFrame")
ras <- raster(ab)
```

---

### Reclassification d'une couche raster

Chaque pixel du raster créé contient une valeur en fonction de la densité des points calculée par avant. Afin d'optimiser le résultat, il est nécessaire de faire une reclassification de l'image raster. Nous allons classifier l'image en 4 catégories : les pixels vides auront la valeur 0. Les autres pixels seront regroupés en 3 classes égales.

**Librairies utilisées :** R, raster

**Données en entrée :** Variable R (raster)

**Données en sortie :** Variable R (raster reclassifié)

---

#### Algorithme 17: Transformation points en raster

---

```
Syntaxe : maxim <- maxValue(ras)
minim <- minValue(ras)
somme <- maxim+minim
sixieme <-somme/6
tiers <-sixieme*2
m <- c(minim-1,0,0,0.0001, sixieme,1, sixieme,tiers,2, tiers,maxim,3)
rclmat <- matrix(m, ncol=3, byrow=TRUE)
rc <- reclass(ras,rclmat)
```

---

Illustration des résultats avant et après la reclassification :

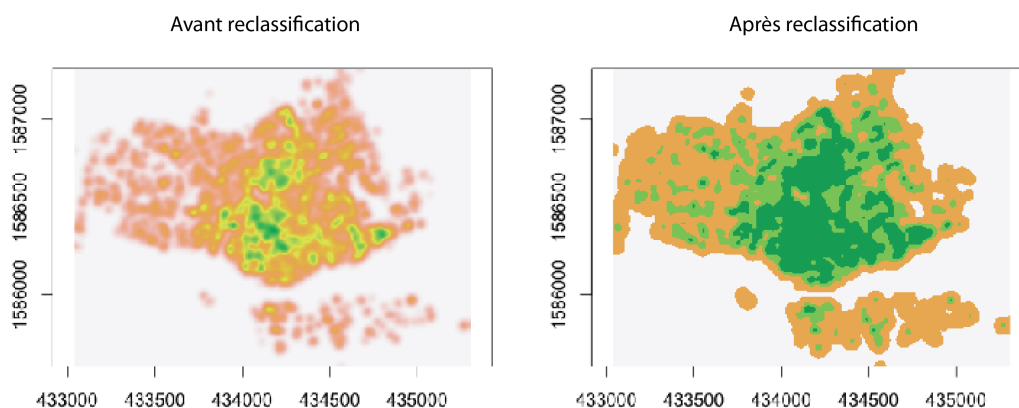


FIGURE 8.4 – Reclassification d'une image raster

## Post-traitements

### Combinaison de plusieurs couches raster

La combinaison de deux couches raster nous permet d'obtenir une combinaison entre la carte de vulnérabilité (=carte de densité des points) et la carte d'aléas (=carte buffers). La fonctionnalité à utiliser est "mosaic" de la librairie "raster".

**Librairies utilisées :** R, raster

**Données en entrée :** 2 données raster

**Données en sortie :** 1 donnée raster

---

#### Algorithme 18: Combinaison de deux données raster

---

Syntaxe : `mosaic(raster1,raster2,fun=sum)`

---

L'argument "fun" de cette fonction indique la méthode selon laquelle les deux raster sont combinés. Nous utilisons la méthode "sum" (somme) qui va donc calculer la somme des valeurs des pixels des raster combinés.

### 8.0.3 Configuration prérequis pour l'utilisation des outils

Avant d'utiliser le logiciel, l'utilisateur doit tout d'abord s'assurer d'avoir un environnement Java, ou **JRE** sur sa machine. Les explications nécessaires peuvent être trouvées sur le site suivant : <http://www.siteduzero.com>

L'utilisateur doit également disposer d'un accès à une base de données spatiale PostGIS. Pour ceci, il y a deux possibilités : L'utilisateur installe ou a déjà installé PostgreSQL et PostGIS en local sur sa machine (en local). Sinon, il existe également la possibilité de se connecter à une base de données stockée en ligne. Afin de pouvoir traiter et gérer des données raster, il faut que la version 2.0 de PostGIS soit installée. La possibilité de se connecter à une base de données distante augmente les possibilités de réutilisation du logiciel.

Finalement, l'utilisateur doit également avoir installé sur sa machine le logiciel R, ainsi les "packages" suivant :

- rgdal

- RPostgreSQL
- raster
- spatstat
- maptools

A l'instant, toutes les fonctionnalités peuvent être utilisées sous Linux et sous MacOS. Sous Windows, un certain nombre de fonctionnalités ne sont pas disponibles (cf Difficultés rencontrées).

# Bibliographie

- Abd-el KADER, Y. (2006). *Conception et exploitation d'une base de métadonnées de traitements informatiques, représentation opérationnelle des connaissances d'expert*. Thèse de doctorat.
- ADJUIK, M., BAGAYOKO, M., BINKA, F., COETZEE, M., COX, J., CRAIG, M., DEICHMAN, U., DESAVIGNY, D., FONDJO, E., FRASER, C., GOUWS, E., KLEINSCHMIDT, I., LEMARDELEY, P., LENGELER, C., LESUEUR, D., OMUMBO, J., SNOW, B., SHARP, B., TANSER, F., TEUSCHER, T. et TOURÉ, Y. (1998). Towards an atlas of malaria risk in africa. Rapport technique, MARA/ARM Collaboration.
- C. ASCHAN-LEYGONIE, S.-M. (2009). Risque, vulnérabilité, résilience : comment les définir dans le cadre d'une étude géographique sur la santé et la pollution atmosphérique en milieu urbain. *Vulnérabilités sociétales, risques et environnement*, page 8.
- ERMERT, V., FINK, A. H., JONES, A. E. et MORSE, A. P. (2011). Development of a new version of the liverpool malaria model. i. refining the parameter settings and mathematical formulation of basic processes based on a literature review. *Malaria Journal*, page 17.
- Faire reculer le PALUDISME, O. m. d. l. S. (2010). Qu'est-ce que le paludisme ? Rapport technique.
- GAUDART, J., DESSAY, N., TOURÉ, O., FANÉ, M., SABATIER, P. et DOUMBO, O. (2005). Développement d'une méthodologie appliquée à l'évaluation du risque palustre, au mali. Rapport technique.
- KREFIS, A. C., SCHWARZ, N. G., NKURUMAH, B., ACQUAH, S., LOAG, W., OLDELAND, J., SARPONG, N., ADU-SARKODIE, Y., RANFT, U. et MAY, J. (2011). Spatial analysis of land cover determinants of malaria incidence in the ashanti region, ghana. *Plosone*, 6(3).
- LIN, Y. (2011). *Méthodologie et composants pour la mise en oeuvre de workflows scientifiques*. Thèse de doctorat.
- LOIREAU, M., SGHAIER, M., FÉTOUI, M., BA, M., ABDELRAZIK, M., D'HERBÈS, J.-M., DESCONNETS, J.-C., LEIBOVICI, D., DEBARD, S. et DELAITRE, [U+FFFD] (2007). Système d'information sur l'environnement à l'échelle locale (siel) pour évaluer le risque de désertification : situations comparées circumsahariennes (réseau roselt). *Sécheresse*, pages 328–335.
- LOIREAU MAUD, Sghaier Mongi, F. M. O. M. T. H. O. B. A. (2012). Guide utilisateur siel.
- MACHAULT, V., VIGNOLLES, C., BORCHI, F., VOUNATSOU, P., PAGES, F., BRIOLANT, S., LACAUX, J.-P. et ROGIER, C. (2011). The use of remotely sensed environmental data in the study of malaria. *Geospatial Health*, 5(2):151–168.

- MINAKAWA, N., MUTERO, C. M., GITHURE, H. I., BEIER, J. C. et YAN, G. (1999). Spatial distribution and habitat characterization of anopheline mosquito larvae in western kenya. *AMm. J. Trop. Med. Hyg*, (61(6)).
- ORGANIZATION, W. H. (2012). World malaria report 2011. Rapport technique, World Health Organization.
- PICHERAL, H. (2001). Dictionnaire raisonné de géographie de la santé.
- PROTOPOPOFF, N., BORTEL, W. V., SPEYBROECK, N., GEERTRUYDEN, J.-P. V., BAZA, D., D'ALESSANDRO, U. et COOSEMANS, M. (2009). Ranking malaria risk factors to guide malaria control efforts in african highla. *PLoS ONE*, 4(11).
- SANTÉ, A. N. d. e. d. e. (2002). Construction et utilisation des indicateurs dans le domaine de la santé. Rapport technique, Agence Nationale d'Accréditation et d'Évaluation en Santé.
- SANTÉ, A. N. d. S. (2006). Santé et environnement : définitions et évolutions récentes. Rapport technique.