



Master 2 Informatique  
Spécialité Géomatique

---

# **Conceptualisation et développement d'outils pour cartographier le risque**

**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	Le "modèle" SIEL . . . . .	4
2.2.2	Architecture SIEL . . . . .	4
2.3	Problématiques du stage . . . . .	5
<b>3</b>	<b>Conexte thématique</b>	<b>6</b>
3.1	Environnement-santé . . . . .	6
3.1.1	Définitions . . . . .	6
3.1.2	Facteurs de risque de transmission du paludisme . . . . .	8
3.2	Approches logicielles existantes . . . . .	12
3.2.1	Cas d'utilisation concret dans le contexte environnement-santé : . . . .	13
3.2.2	Récapitulatif existant / besoins . . . . .	13
<b>4</b>	<b>Méthodologie</b>	<b>14</b>
4.1	Compréhension de la problématique . . . . .	14
4.1.1	Données disponibles . . . . .	16
4.1.2	Modèle simplifié . . . . .	16
4.2	Raisonnement . . . . .	18
4.2.1	Description des données et des traitements . . . . .	18
4.2.2	Les catégories de données et de traitements . . . . .	21
4.2.3	Conceptualisation de la chaîne de traitements . . . . .	22
4.3	Le bloc élémentaire . . . . .	25
4.3.1	Librairies et bibliothèques utilisées . . . . .	25
4.3.2	Architecture informatique de la chaîne de traitements . . . . .	28
4.3.3	Implémentation de la chaîne de traitements . . . . .	29
4.3.4	Architecture informatique du logiciel ouvert . . . . .	33
4.3.5	Implémentation de lu logiciel ouvert . . . . .	33
<b>5</b>	<b>Résultats</b>	<b>34</b>
5.1	Les chaînes de traitements . . . . .	34
5.2	Discussion . . . . .	34
5.3	Perspectives . . . . .	34
<b>6</b>	<b>Conclusion</b>	<b>35</b>
<b>7</b>	<b>Glossaire et définitions</b>	<b>36</b>

<b>8</b>	<b>Annexes</b>	<b>37</b>
8.0.1	Les traitements effectués . . . . .	37

# 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

Le présent stage s'inscrit dans un contexte de coopération entre l'équipe SIC et l'équipe AIMS.

Depuis plusieurs années au sein de l'UMR Espace-Dev 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 d'évaluer le risque de dégradation de la végétation en milieu aride et permet de produire des indices environnementaux spatialisés sous la forme de cartes.

L'objectif du présent stage est d'élargir les potentialités du SIEL en définissant dans un premier temps des facteurs de risque de transmission dans le contexte environnement-santé et plus précisément du paludisme. Le développement d'une chaîne de traitements indépendante et basée sur une plateforme Open Source devrait permettre à long terme de rendre le fonctionnement du SIEL plus souple, plus ouvert et plus facile face à des problématiques diverses comme par exemple l'environnement-santé ou la déforestation.

Un travail supplémentaire a été effectué et a abouti à un logiciel ouvert qui permet d'exécuter chaque traitement de la chaîne indépendamment et de combiner les traitements selon les besoins

---

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

de l'utilisateur et donc adaptable à de nouvelles problématiques. Ce logiciel ouvert est une autre étape importante dans le développement d'une nouvelle architecture informatique du SIEL.

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 domaine de l'environnement-santé il y a une forte demande en terme d'automatisation des traitements respectivement dans l'automatisation de la création d'indicateurs.

Il est donc particulièrement intéressant d'intégrer la construction automatisée d'indicateurs dans le contexte environnement-santé au sein du SIEL. 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".

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 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 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 Le "modèle" SIEL

Le SIEL est un modèle environnemental instrumenté (Loireau et al. 20056a) qui spatialise les pratiques d'exploitation des ressources sur un territoire, modélise les paysages et évalue les risques de dégradation des ressources. (référence guide utilisateur)

Le SIEL se base sur le fonctionnement interactif milieux/sociétés sur un territoire dans lequel des acteurs, selon leurs représentations de leur territoire et leurs usages (agricole, pastoral, forestier), ont des pratiques qui permettent d'exploiter des ressources naturelles (physiques, biologiques) sur un territoire donné (Guide utilisateur). Les interactions entre l'espace, les acteurs, les usages et les ressources produisent des paysages qui vont composer les territoires. Ces paysages évoluent en fonction des interactions entre facteurs naturels et humains. Pour mettre en oeuvre ce phénomène, le modèle SIEL décompose le territoire en Unités Spatiales dites « de Référence » (USR). Ces USR correspondent à une délimitation spatiale des différents paysages du territoire par voie de modélisation.

Un minimum de données est nécessaire pour alimenter la chaîne de traitements :

- données socio-économiques (acteurs, stratégies, pratiques d'exploitation...)
- données biophysiques (ressources, végétation...)

Enfin, le modèle SIEL met en œuvre une méthode spatiale interdisciplinaire qui permet une cartographie dynamique des paysages et le calcul d'indices environnementaux associés (Loireau et al, 20077).

### 2.2.2 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, en faisant différents traitements, créent des indices environnementaux spatialisés qui permettent la cartographie des risques de dégradation des ressources. Les données utilisées et les résultats sont stockés dans une base de données.

## 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 : Choix de l'architecture informatique , conceptualisation des traitements et implémentation des outils (chaîne de traitements et logiciel ouvert).

Concernant l'environnement-santé, la problématique principale est la compréhension de la problématique, la définition des facteurs de risque et la construction d'un modèle conceptuel du cycle du paludisme.

Plusieurs problématiques au niveau informatique peuvent être soulevées : Choix de l'architecture informatique, choix du langage informatique principal, proposition d'un prototype, élaboration de modèles de données et de traitements et implémentation de la chaîne. Le développement d'un logiciel ouvert consiste un travail supplémentaire, chaque traitement peut être exécuté indépendamment et combiné selon les besoins de l'utilisateur. L'idée est d'ouvrir la boîte de la chaîne de traitements et de déstructurer la chaîne en boîtes élémentaires réutilisables. Le logiciel ouvert permet d'exécuter chaque traitement de la chaîne indépendamment ainsi que de les combiner selon les besoins de l'utilisateur.

Le travail de Nadine Dessay sur l'élaboration de cartes de risques de paludisme va servir comme base lors de la définition des traitements à réaliser. Dans ce travail, elle définit les étapes nécessaires pour l'élaboration de ce type de cartes par rapport à la problématique du paludisme. Ces traitements sont réalisés sous ArcGIS, l'objectif est de trouver des fonctionnalités similaires avec des outils "OpenSource" et d'automatiser l'exécution de tous ces traitements et calculs à l'intérieur de la chaîne de traitements. Le logiciel ouvert permettra d'exécuter chacun de ces traitements indépendamment.

Nadine Dessay a également extraits plusieurs données à partir d'une image de haute résolution **Quickbird** de la ville de Bandiagara au Mali qui serviront comme "données d'essai" pour le prototype de la chaîne de traitements.

En résumé une problématique générale du stage se dégage :  
**Conceptualisation et développement d'outils pour cartographier le risque.**

# Chapitre 3

## Conexite thématique

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

### 3.1 Environnement-santé

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

#### 3.1.1 Définitions

##### Environnement-santé

L'environnement-santé peut être défini comme l'influence de la qualité de l'environnement physique, chimique et biologique sur la santé des hommes et des animaux. 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. (afsset)

##### Paludisme

Le paludisme, appelé également malaria, est une parasitose due à un protozoaire transmis par la piqûre de la femelle d'un moustique (**anophèle**), provoquant des fièvres intermittentes. La maladie est la cause d'environ deux millions de décès chaque année dans le monde, principalement dans les régions tropicales et en Afrique sub-saharienne. (5). (cf Annexe carte risque transmission)

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 transformation qui lui permettent d'échapper au système immunitaire humain. Au final, lorsqu'un moustique non-infectée 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 de

risque, 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 plus gravement affectée que la moyenne par un aléa (dans notre cas d'étude une source de maladie), soit en raison d'une sensibilité aux effets de l'aléa ou sinon suite à une exposition plus importante que la moyenne.

## **Aléa**

Pour définir l'aléa pour le risque sanitaire, nous nous appuyons sur les travaux de (C.Aschan-Leygonie et S.Baudet-Michel (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 dépend de la capacité d'une population ou d'un système de faire face à des menaces (aléas). 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).

## **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 ?.

## **Facteur de risque**

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.

## **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.

### 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 :** La précipitation est un des facteurs de risque les plus importants du paludisme. 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. 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. (cf 3.1.2)

**Température :** Le pourcentage de survie des moustiques sur un territoire est étroitement lié à la température. (16). 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 (10) 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.

**Distance point eau** Les moustiques anophèles peuvent se déplacer d'un maximum de 7 kilomètres par rapport à un point d'eau. 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 par avant.

**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 (13) 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 %.

**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.

**Végétation :** Certains types de végétation servent comme habitat pour les moustiques adultes. (à revoir)

### **Facteurs biologiques**

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.

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. 3.1.2

**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 (1). 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. (1) 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.(19). 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 liés 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.(14)

**Qualité système de santé** La qualité du système de santé ou l'accès au système de santé influe directement sur le risque du paludisme (19). 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 également être regroupés en deux catégories différentes : les facteurs causant un aléa et les facteurs causant 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.

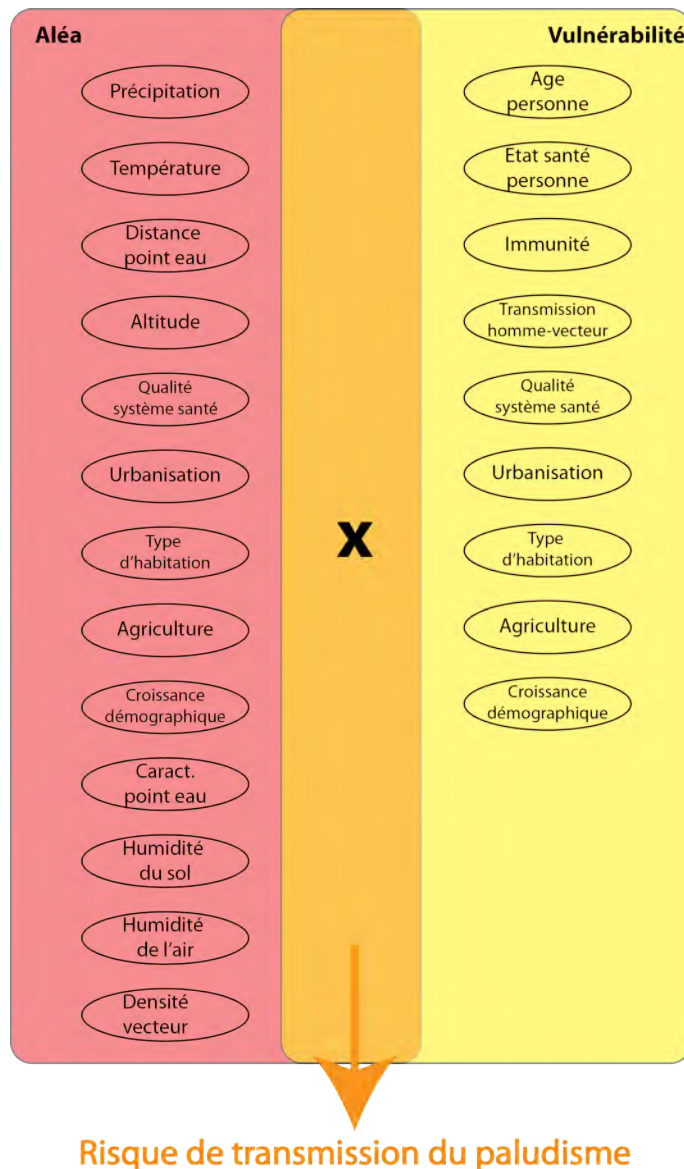


FIGURE 3.1 – Alea et Vulnérabilité selon les facteurs

## 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.

### 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.



## **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. (Peut-être plutôt dans partie Méthodologie ? ?)

### **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 (biblio) 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 Récapitulatif existant / besoins**

En terme d'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 de la télédétection et des outils informatiques à la lutte contre paludisme. (cf...)

Il est donc particulièrement intéressant de proposer un prototype incluant les traitements permettant de cartographier les zones de risque de paludisme. En se basant sur cette chaîne, nous avons développé un logiciel qui permet aux utilisateurs d'utiliser les traitements selon leurs besoins. Ce logiciel servira dans le futur comme base pour retravailler le SIEL et pour intégrer de nouvelles problématiques au sein du SIEL et pour développer une architecture informatique OpenSource pour le SIEL. Cette nouvelle architecture facilitera l'utilisation notamment car il ne sera plus nécessaire d'acquérir une licence ArcGIS.

# Chapitre 4

## Méthodologie

L'objectif principal du stage est de proposer 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.

Dans un premier temps la conceptualisation du cycle de paludisme sous forme d'un modèle conceptuel UML m'a permis de m'appréhender de comprendre la problématique. Ce modèle nous permet par la suite 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.

Par la suite, nous avons conduit un raisonnement qui permet à partir des données disponibles d'obtenir le résultat souhaité : Une chaîne de traitements respectivement un logiciel ouvert qui rend accessible facilement chacun des traitements. Pour ce faire, il est nécessaire de comprendre le fonctionnement des traitements et des données. Des modèles conceptuels des traitements et des données sont proposés à la fin de cette sous-partie.

Au niveau informatique, nous présentons les traitements utilisés au sein de la chaîne de traitements. 4 modèles présentent le fonctionnement de la chaîne, en partant d'une représentation la plus large possible pour arriver à une instanciation de tous les traitements.

A partir des différents modèles conceptuels, nous avons choisi les architectures informatiques de la chaîne de traitements. Nous présentons les bibliothèques et libraires utilisées ainsi que l'implémentation de la chaîne.

Finalement, nous présentons également l'architecture informatique et l'implémentation du logiciel ouvert. Le développement de ce logiciel est un travail supplémentaire.

### 4.1 Compréhension de la problématique

Le modèle conceptuel du cycle du paludisme 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.



## Précision modèle conceptuel du cycle du paludisme

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 virus peut également infecter une moustique non infectée par avant.

### 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 :

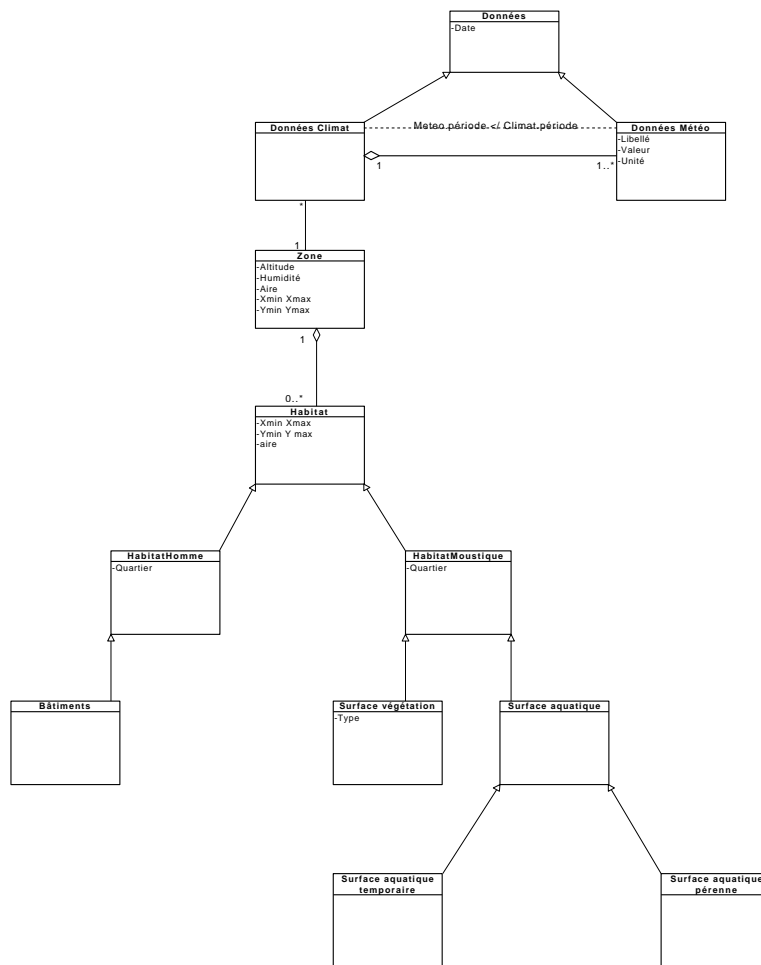


FIGURE 4.2 – Modèle UML simplifié du paludisme

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 :

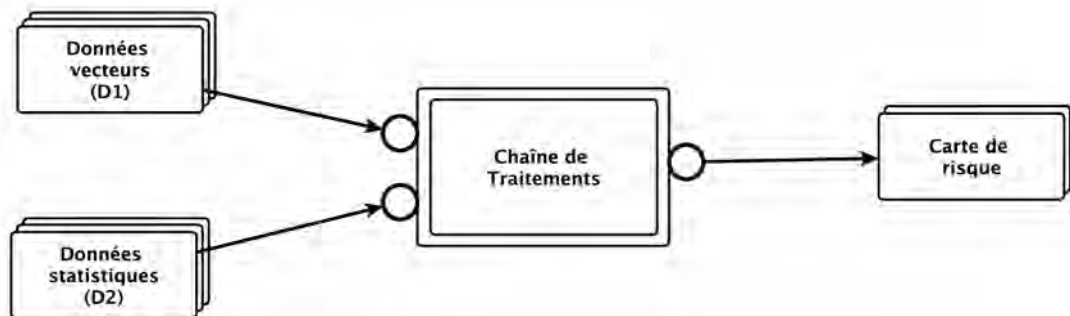


FIGURE 4.3 – Description de ce que fait un traitement

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.

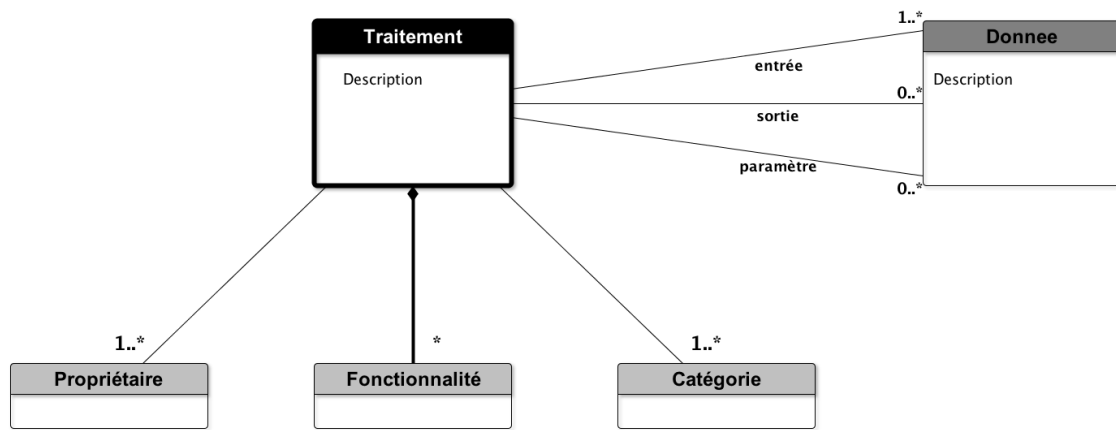


FIGURE 4.4 – Description du fonctionnement d'un traitement

**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. Conformément aux catégories de traitements définis par (Yuan Thèse), un traitement correspond aux catégories suivantes : pré-traitements, traitements et post-traitements.

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 : Pré-traitements
- 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, un format et une ou plusieurs fonctionnalités.

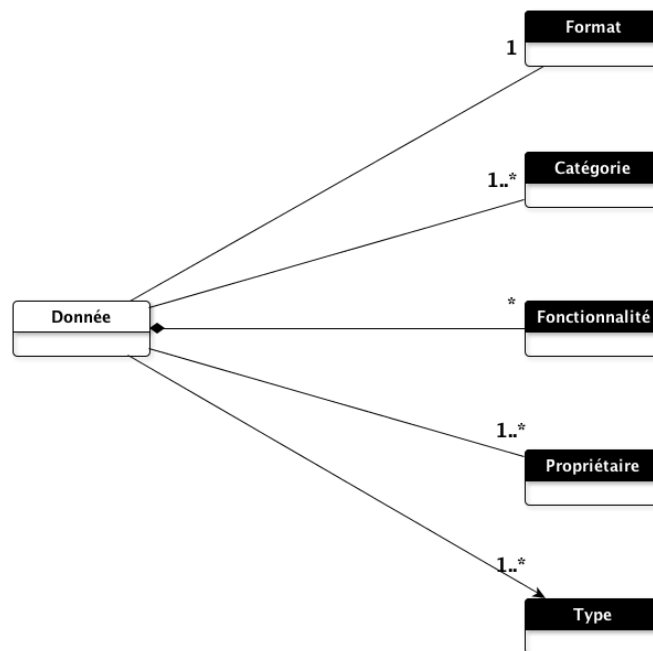


FIGURE 4.5 – Description des données

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 caractérisé 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
- Fonctionnalité(s) : Dans notre cas : Calculer densité de population
- Catégorie : Capteurs satellites
- Type : Polygone
- 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.

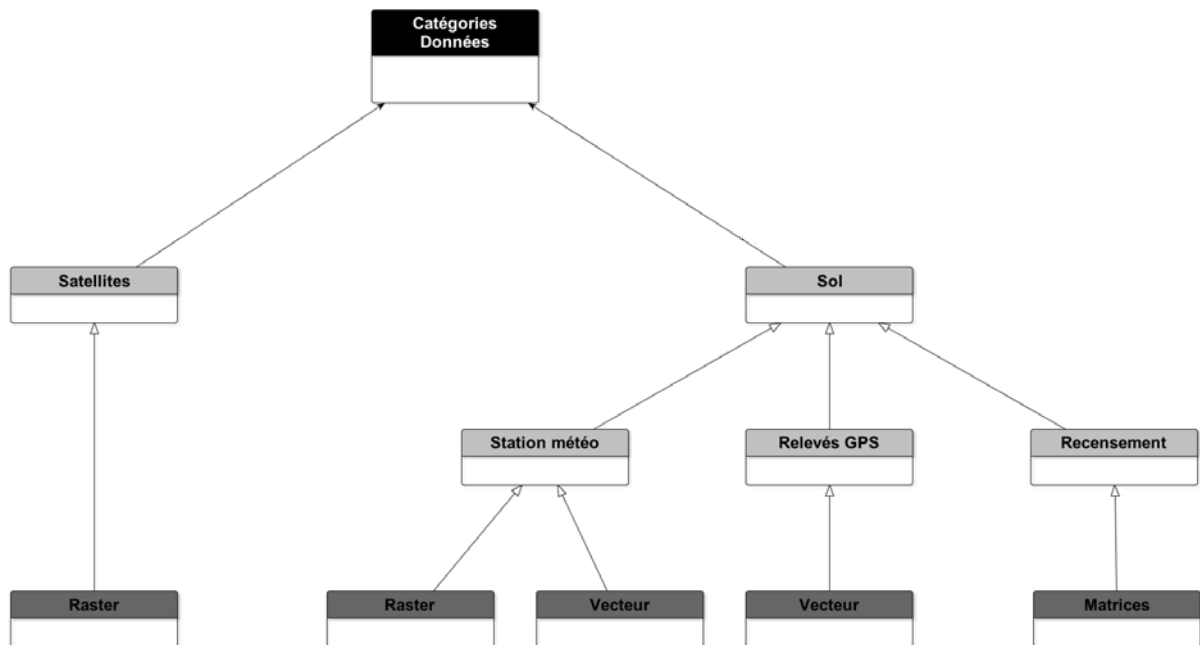


FIGURE 4.6 – Modèle conceptuel des catégories de données de la chaîne de traitements

Les données de la chaîne de traitements sont issues de sources diverses. Les **données issus de satellites** correspondent aux données obtenues à partir de satellites (passifs ou actifs). Les **données issus 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 différents formats. Par exemple, une station météo peut fournir des données au format vecteur ou au format raster. Des relevés GPS fournissent des fichiers vecteurs. Des données matricielles ?? 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'ils puissent être traitées. Dans le cadre de ce travail il est nécessaire de stocker les données, de les reprojeter et d'effectuer des requêtes SQL en fonction des données qui seront utilisées par les 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 deux sous-catégories : les traitements **statistiques** et les traitements **d'analyse spatiale**.

Les traitements statistiques concernent les calculs de densités de population et/ou de points respectivement le calcul de la taille de pixel des couches raster.

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

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.

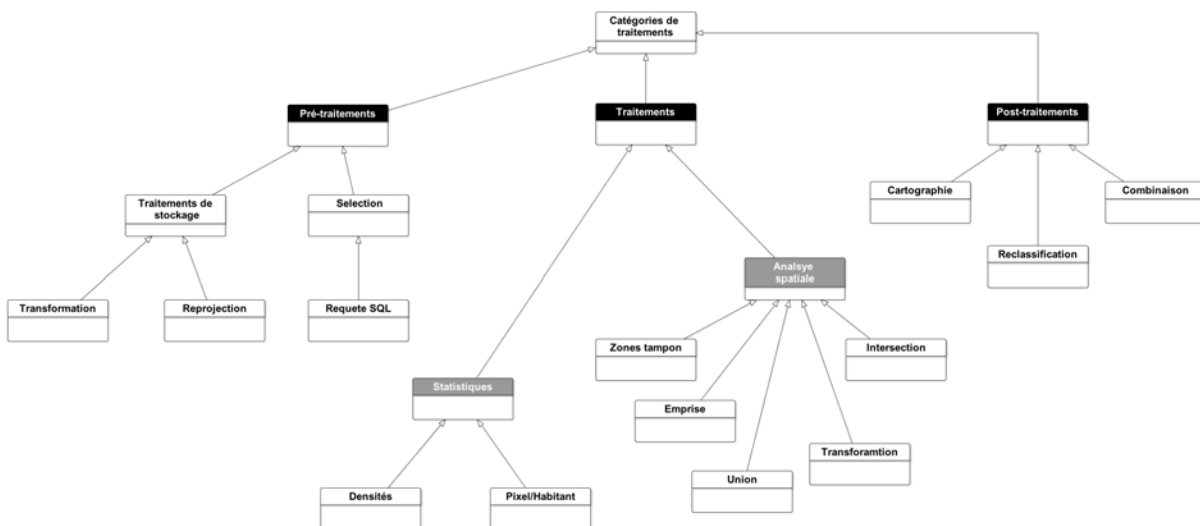


FIGURE 4.7 – Modèle conceptuel des hiérarchies de catégories de traitements de la chaîne de traitements

### 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.). Ce 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 4.8)

## Description générale

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

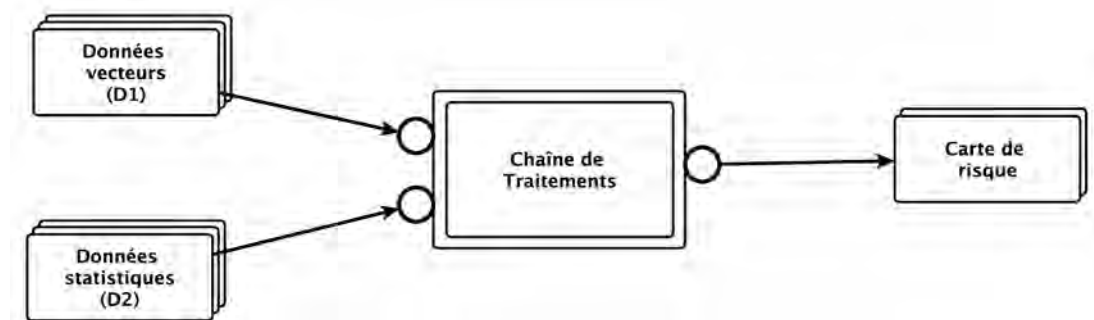


FIGURE 4.8 – Description générale

## Instanciation des données et des traitements

Dans un premier temps, les données vecteurs sont stockés (traitements de stockage) dans une base de données. Par la suite, des **sélection** 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.

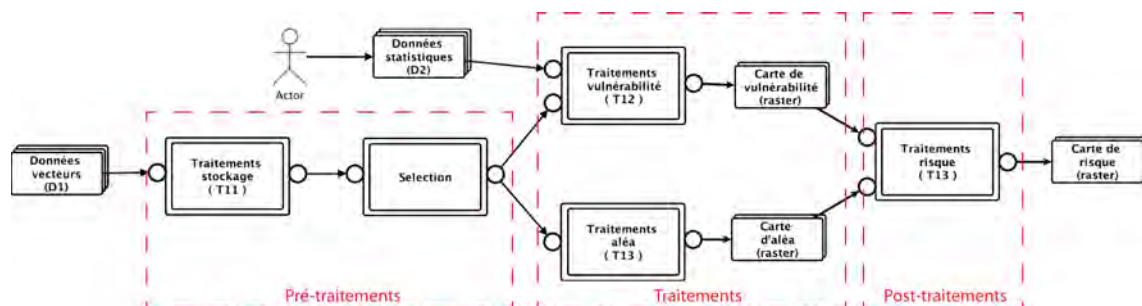


FIGURE 4.9 – Instanciation des données et des traitements

## Instanciation des catégories 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 par la suite insérées dans une base de données spatiale (créée au cours de l'exécution de la chaîne de traitements). Des requêtes SQL permettent de sélectionner 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 à partir de traitements statistiques et d'analyse sur les données vecteurs (bâtiments) et les données statistiques.

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 des cartes de vulnérabilité et d'aléa. La carte de risque est le résultat de l'ensemble des traitements.

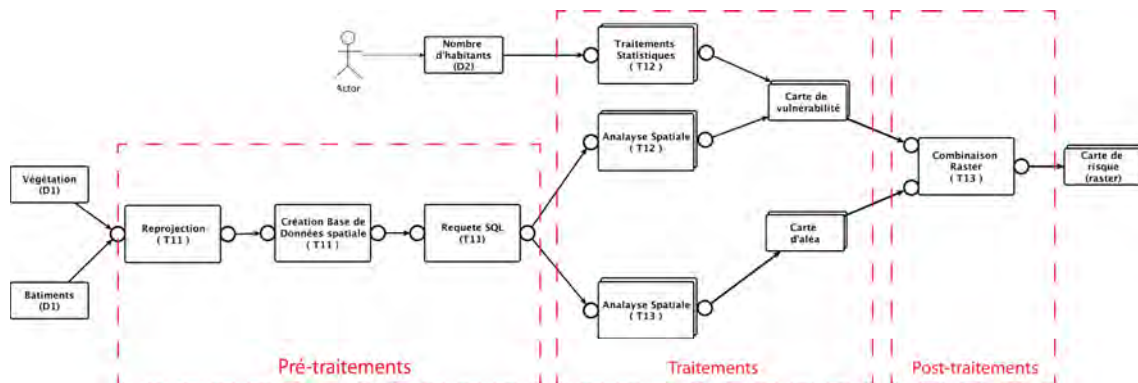


FIGURE 4.10 – Instanciation des catégories de traitements

## Instanciation de la chaîne de traitements

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'un pixel pour un habitant. La couche "bâtiments" est ensuite rastérisée en fonction de la taille de pixel calculée par avant. Chaque pixel est transformé en points et 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.

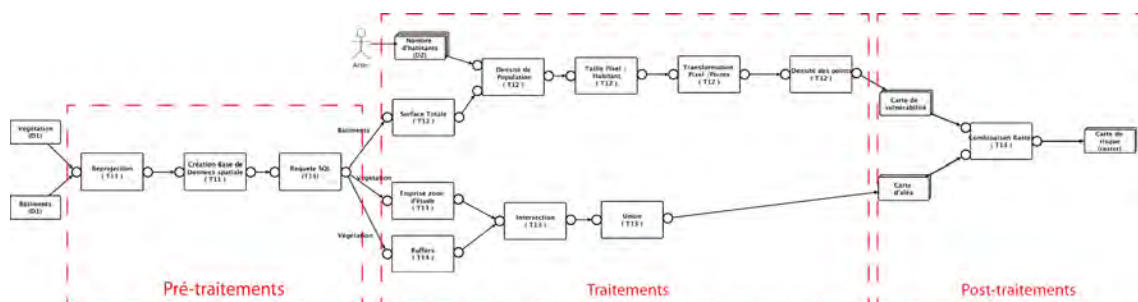


FIGURE 4.11 – Instanciation de la chaîne de traitements

## 4.3 Le bloc élémentaire

Pour valider sur le plan opérationnel de 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 tout d'abord les librairies et bibliothèque que nous avons utilisés pour développer les deux outils. Dans un second temps nous présentons l'implémentation des deux outils, respectivement l'architecture informatique des deux outils.

### 4.3.1 Définition, 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." ((?))

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 bibliothèques 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").

### 4.3.2 Architecture informatique de la chaîne de traitements

Le prototype de la chaîne de traitements appelle 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 du prototype.

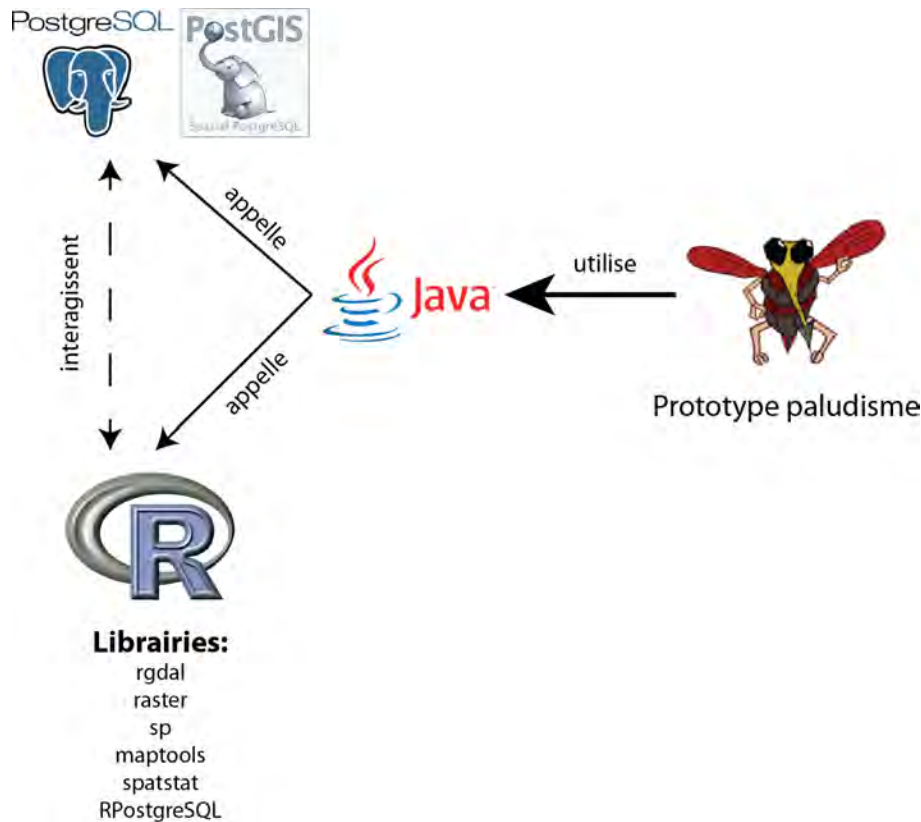


FIGURE 4.12 – Architecture informatique du prototype

Dans un premier temps, toutes les données disponibles sont insérées dans une base de données PostgreSQL (incluant l'extension PostGIS). Cette base de données est créée automatiquement lors de l'exécution de la chaîne de traitements. Les différents outils utiliseront les données dans cette base de données et interagissent entre eux. 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 est donc basée exclusivement sur des outils Open Source et permet 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 de format différentes grâce aux nombreuses bibliothèques disponibles pour cet outil. La bibliothèque "spatstat" nous permet par exemple de calculer une densité de population ou d'effectuer une reclassification d'une image raster. La bibliothèque "rgdal" nous permet de disposer des fonctionnalités de **GDAL** et de charger des données raster ou vectorielles à partir de la 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 problématique (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.

### 4.3.3 Implémentation de la chaîne de traitements

**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) et au maximum 5 données.

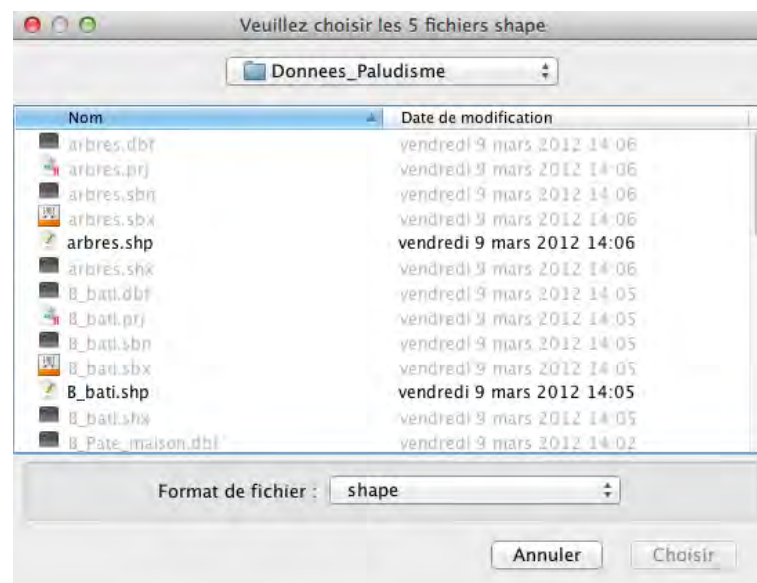


FIGURE 4.13 – Choix fichiers d'entrée

**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.

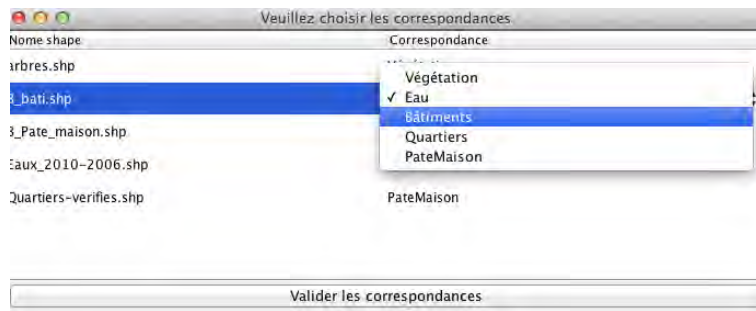


FIGURE 4.14 – Correspondances des données

**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 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.

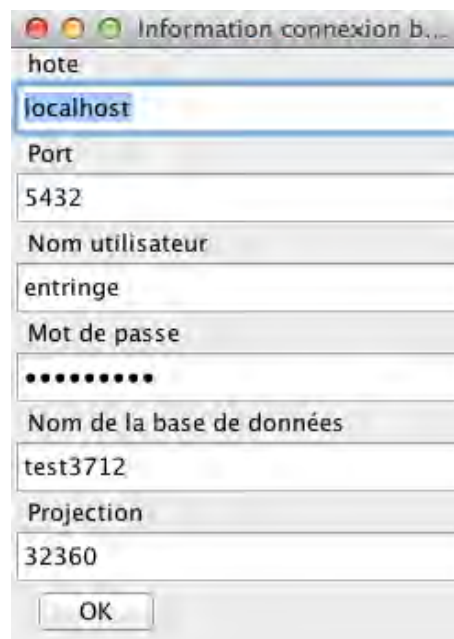


FIGURE 4.15 – Informations création base de données

**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".

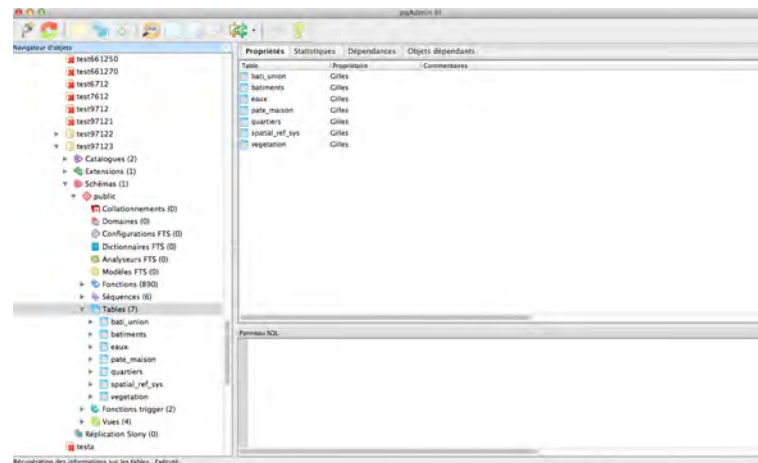


FIGURE 4.16 – Base de données avec données insérées

**Nombre d’habitants** L’outil demande à l’utilisateur le nombre d’habitants pour la zone d’étude.

FIGURE 4.17 – Nombre d’habitants

**Calcul densité de population et taille pixel** L’outil calcule automatiquement, à partir de la couche "bâtiments", la densité de population de la zone d’étude et la taille de pixel théorique pour un habitant.

```
Gilles — java — 80x24
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 4.18 – Taille Pixel / Densité de population

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

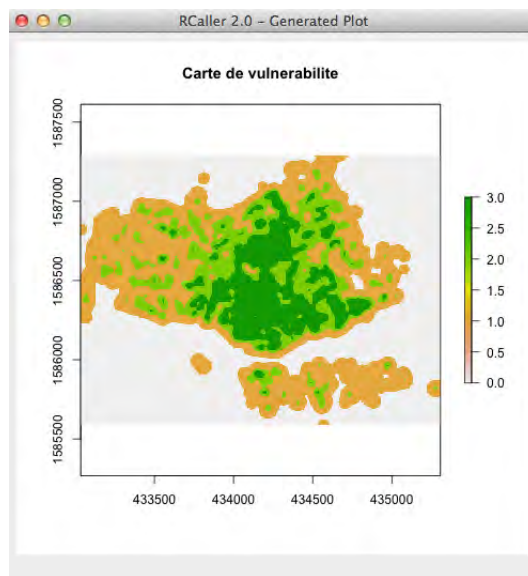


FIGURE 4.19 – Carte de vulnérabilité

**Suppression des surfaces aquatiques polluées** Dans le cadre des données que nous possédons, il est nécessaire de supprimer des surfaces aquatiques polluées. L'outil permet donc de sélectionner quelles surfaces sont polluées à partir d'une colonne qui permet d'identifier ces surfaces.

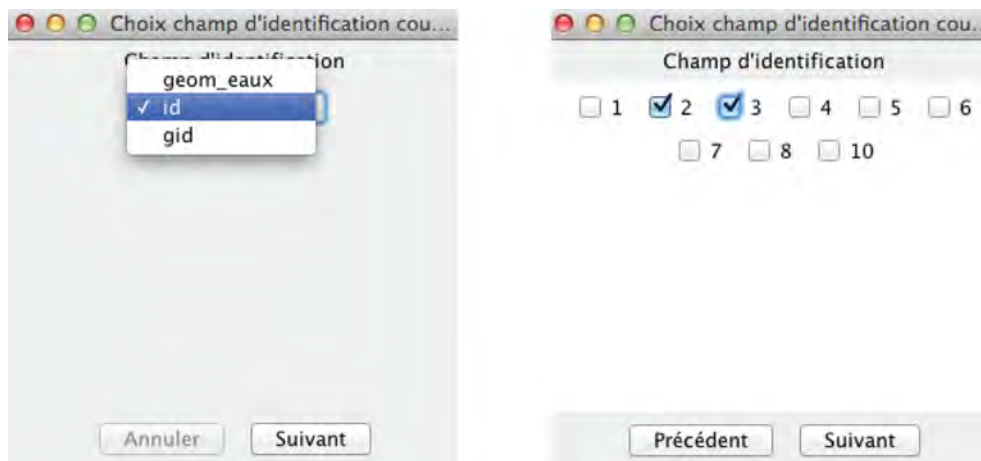


FIGURE 4.20 – Supprimer surfaces aquatiques polluées

**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 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.

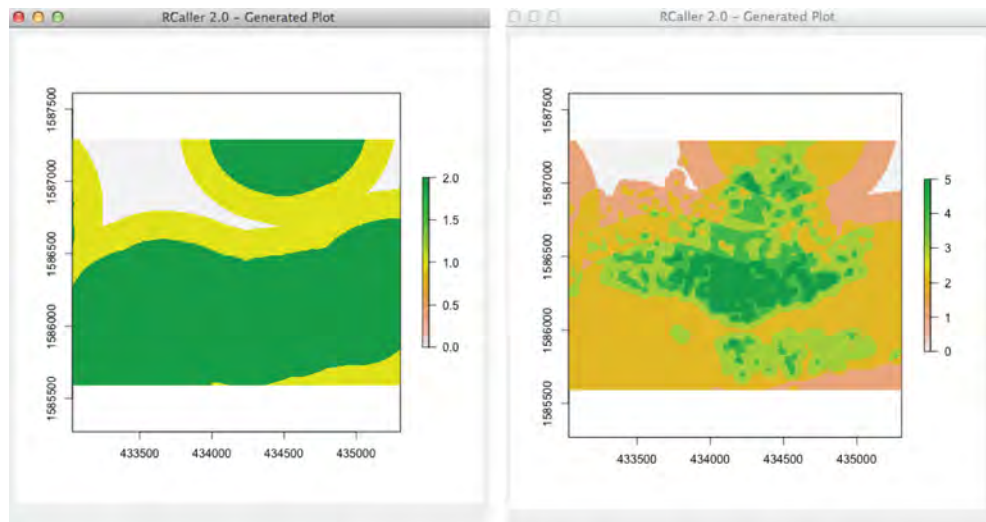


FIGURE 4.21 – Exemple de carte d'aléa et carte de risque

### 4.3.4 Architecture informatique du logiciel ouvert

### 4.3.5 Implémentation de lu logiciel ouvert

#### Conclusion

Les différents traitements permettent de créer des cartes de risque de transmission du paludisme. Les pré-traitements permettent de préparer les différents outils et données. Les traitements statistiques sont effectués afin de disposer de toutes les informations utiles et nécessaires sur le territoire d'étude.

à revoir en fonction des modifications apportées à la chaîne...

# **Chapitre 5**

## **Résultats**

### **5.1 Les chaînes de traitements**

### **5.2 Discussion**

### **5.3 Perspectives**

# **Chapitre 6**

## **Conclusion**

# 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



# Chapitre 8

## Annexes

### 8.0.1 Les traitements effectués

Pour chaque traitement effectué, la chaîne utilise différentes bibliothèques 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.

**Bibliothèques 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 la 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 :** PostgreSQL, 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 PostgreSQL / Postgis :

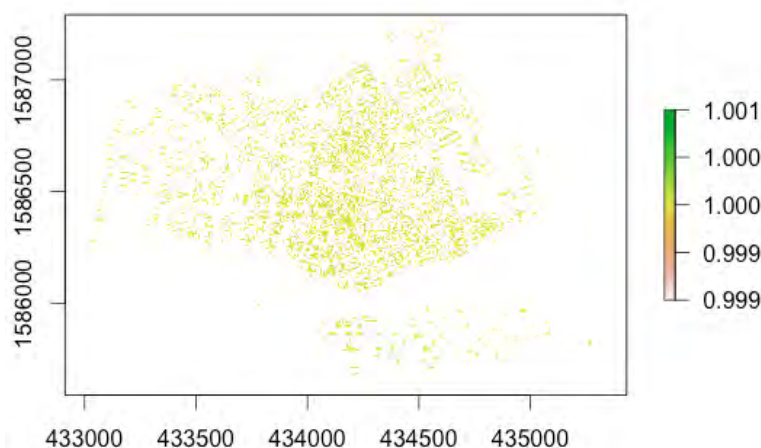


FIGURE 8.1 – Raster chargé dans R

### 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

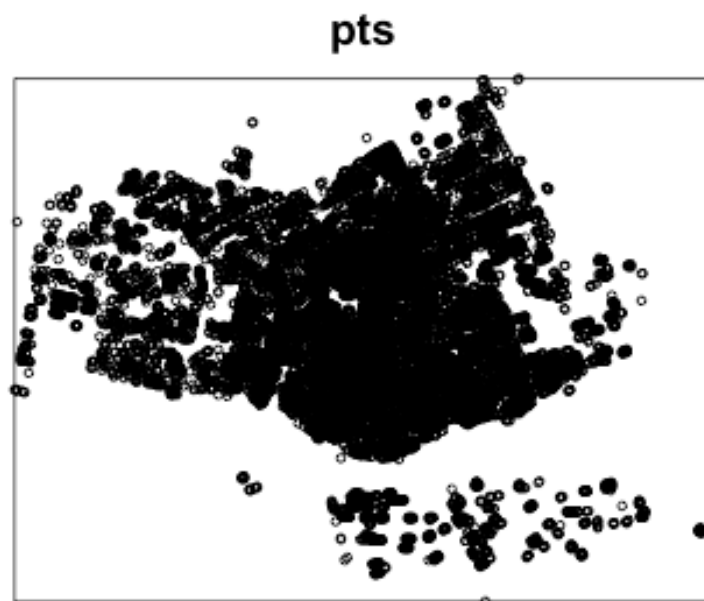


FIGURE 8.2 – Raster transformé en points

### 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 :

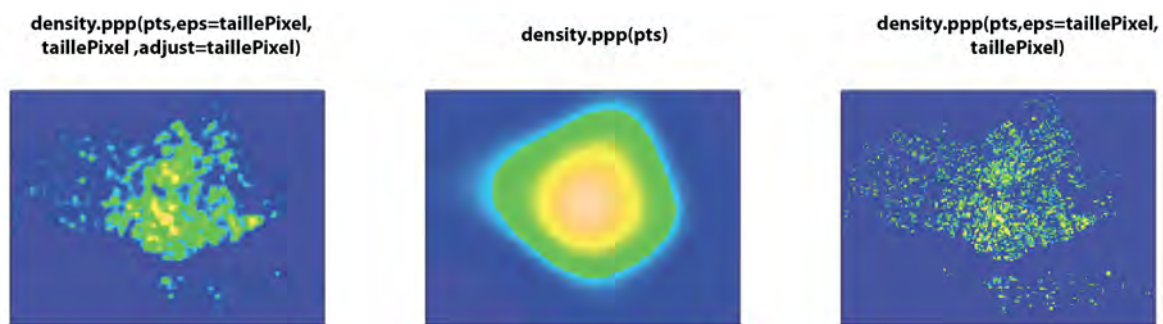


FIGURE 8.3 – Densités des points selon les paramètres

### 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 <- max(ras)
minim <- min(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 :

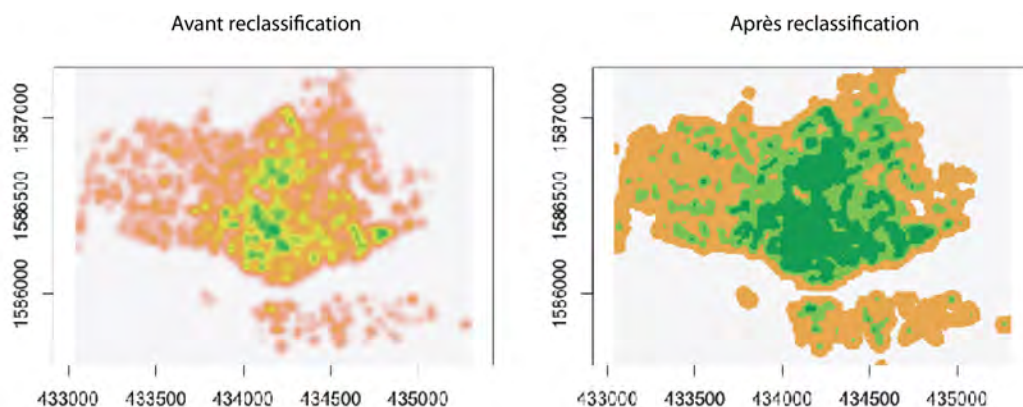


FIGURE 8.4 – Reclassification d'une image raster

## 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.

# Bibliographie

Faire reculer le PALUDISME, O. m. d. l. S. (2010). Qu'est-ce que le paludisme ? Rapport technique, Faire reculer le paludisme, Organisation mondiale de la Santé.