



UNIVERSITÉ
LUMIÈRE
LYON 2

N°d'ordre NNT : 2016LYSE2005

THESE de DOCTORAT DE L'UNIVERSITÉ DE LYON

Opérée au sein de

L'UNIVERSITÉ LUMIÈRE LYON 2

École Doctorale : ED 512 : Informatique et Mathématiques

Informatique

Soutenue publiquement le 27 janvier 2016, par :

Fatima Zohra YOUNSI

Mise en place d'un Système d'Information Décisionnel pour le Suivi et la Prévention des Epidémies

Devant le jury composé de :

Karim BOUAMRANE, Professeur d'université, Université d'Oran 1, Président

Gilles ZURFLUH, Professeur des universités, Université Toulouse 1, Rapporteur

Sidi Mohamed BENSLIMANE, Professeur d'université, Université de Sidi Bel Abbès, Rapporteur

Fadila BENTAYEB, Maître de conférences HDR, Université Lumière Lyon 2, Examinatrice

Omar BOUSSAID, Professeur des universités, Université Lumière Lyon 2, Directeur de thèse

Djamila HAMDADOU, Maître de conférences HDR, Université d'Oran 1, Co-Directrice de thèse

UNIVERSITÉ LUMIÈRE LYON 2
UNIVERSITÉ D'ORAN 1

THÈSE

pour obtenir le titre de

Docteur en Informatique

Fatima Zohra YOUNSI

Mise en place d'un Système d'Information Décisionnel pour le Suivi et la Prévention des Epidémies

Sous la direction de
M. Omar BOUSSAID et Mme Djamila HAMDADOU

préparée au sein du Laboratoire ERIC-Université Lumière Lyon 2 et Laboratoire
LIO- Université d'Oran

soutenue publiquement le 27 Janvier 2016

Jury :

<i>Rapporteurs :</i>	Gilles ZURFLUH	- Professeur, Université de Toulouse
	Sidi Mohamed BENSLIMANE	- Professeur, Ecole Supérieure d'Informatique de Sidi Bel Abbes
<i>Examineurs :</i>	Karim BOUAMRANE	- Professeur, Université d'Oran 1 – Ahmed Ben Bella
	Fadila BENTAYEB	- MCF-HDR, Université Lumière, Lyon 2
<i>Directeurs :</i>	Omar BOUSSAID	- Professeur, Université Lumière, Lyon 2
	Djamila HAMDADOU	- MCF-HDR, Université d'Oran 1 – Ahmed Ben Bella
<i>Invité :</i>	Ahmed BOUNEKKAR	- MCF-HDR, Université Claude Bernard, Lyon 1

Table des matières

Liste des Figures	vi
Liste des Tables	vii
Remerciements	viii
Résumé	ix
Acronymes	x
Introduction Générale	1
 I La modélisation mathématique et les réseaux complexes	 8
1 Concepts et définitions de base	9
1.1 Introduction	9
1.2 La modélisation mathématique	10
1.2.1 Concepts de base	10
1.2.2 La modélisation mathématique en épidémiologie	11
1.2.3 Quelques modèles mathématiques de transmission de maladie	12
1.3 Les réseaux complexes	16
1.3.1 L'épidémiologie par réseaux de contacts	17
1.3.2 L'analyse structurelle des réseaux	20
1.4 Conclusion	23
 2 Etat de l'art sur les modèles d'épidémie	 25
2.1 Introduction	25
2.2 Principe de la modélisation mathématique en épidémiologie	25
2.3 Les modèles mathématiques d'épidémie simple	26
2.4 Les modèles compartimentaux déterministes	27
2.4.1 Les modèles compartimentaux déterministes simples	27
2.4.2 Les modèles compartimentaux déterministes spatiaux	28
2.5 Modèles compartimentaux stochastiques	30
2.6 Les modèles mathématiques d'épidémie complexes	31
2.6.1 Les modèles à base d'agent / individus-centrés	31
2.6.2 Les modèles de réseau	32

2.6.3	Les modèles de métapopulation	34
2.6.4	Modèles à base des structures de contact	35
2.7	Conclusion	36
3	Modèle de propagation d'épidémie proposé	39
3.1	Introduction	39
3.2	Les objectifs du modèle de propagation SEIR-SW	40
3.3	L'identification du problème	40
3.3.1	Comprendre la dynamique de la maladie	40
3.3.2	Comprendre la réparation spatiale et la dynamique de la population	41
3.3.3	Définition des hypothèses	41
3.4	La description du modèle de propagation SEIR-SW	42
3.4.1	Le cadre générale de la démarche de modélisation	42
3.4.2	Le modèle SEIR-SW	42
3.5	Simulation et résultats expérimentaux	50
3.5.1	Jeux de données	50
3.5.2	L'estimation des paramètres du modèle SEIR-SW	50
3.5.3	Résultats de la simulation	52
3.5.4	Estimation du nombre de reproduction de base : \mathcal{R}_0	53
3.5.5	La validation du modèle SEIR-SW	54
3.5.6	L'analyse du réseau social petit monde (SW)	55
3.6	Conslusion	58
II	Les systèmes d'Information Décisionnels	59
4	Généralités sur les SID	60
4.1	Introduction	60
4.2	Qu'est-ce qu'un Système d'Information Décisionnel SID ?	61
4.3	L'architecture d'un Système d'Information Décisionnel (SID)	61
4.4	Entrepôts de données (ED) « Datawarehouse»	62
4.4.1	Définitions	62
4.4.2	Les objectifs d'un ED	62
4.4.3	Les composants d'un ED	63
4.4.4	Les phases de construction d'un ED	64
4.5	Schéma multidimensionnel : le cube Olap (On-Line Analytical Processing)	67
4.5.1	Les Opérations de manipulation d'un cube	67
4.5.2	Les architectures OLAP	69
4.5.3	Outils de visualisation OLAP	70
4.5.4	Spatial OLAP (SOLAP)	72
4.6	Conclusion	73
5	Etat de l'art sur les Systèmes d'Information Décisionnels	74
5.1	Introduction	74
5.2	Les Systèmes de surveillance d'épidémies/ Les systèmes d'épidémiosur-	
	veillance	75
5.2.1	Des systèmes à l'échelle nationale et mondiale	75
5.2.2	Travaux de recherche connexes	76
5.3	Les systèmes de surveillance à référence spatiale	77
5.3.1	Le SIG comme outil d'aide à la décision	77

5.3.2	Le SIG et les modèles mathématiques	79
5.4	Les SID épidémiologiques	80
5.4.1	Les travaux traitant les entrepôts de données médicales	80
5.4.2	Les travaux traitant les entrepôts de données médicales et l'OLAP	82
5.4.3	Les travaux traitant les entrepôts de données et les réseaux sociaux	83
5.5	Autres stratégies de lutte contre les épidémies	84
5.5.1	Surveillance active ou passive	84
5.5.2	Surveillance exhaustive	84
5.6	Conclusion	85
6	Le système décisionnel proposé SYDSEP	87
6.1	Introduction	87
6.2	Le système proposé : SYDSEP	88
6.2.1	Les objectifs du système SYDSEP	88
6.2.2	La description du système SYDSEP	88
6.3	La démarche décisionnelle adoptée par SYDSEP	92
6.4	Modélisation UML de SYDSEP	93
6.4.1	Diagramme de cas d'utilisation	93
6.4.2	Diagramme de séquence	93
6.5	Description de l'entrepôt de données	94
6.5.1	Les sources de données	94
6.5.2	Modélisation multidimensionnelle de l'ED spatio-temporel	96
6.6	Mise en œuvre de SYDSEP	99
6.6.1	L'environnement technique	99
6.6.2	Le découpage spatial administratif de la région d'étude	100
6.6.3	SIG et analyse spatiale	100
6.6.4	Création des cubes d'analyse en ligne	101
6.6.5	Suivi et surveillance de la grippe saisonnière	101
6.6.6	La prédiction	104
6.7	Le système SYDSEP : Discussion	107
6.8	Conclusion	108
	Conclusion Générale	109
III	Annexes	113
A	L'épidémiologie	114
A.1	Quelques définitions	114
A.2	Formes épidémiologiques des maladies	115
A.3	Indicateurs épidémiologiques	115
B	Les facteurs climatologiques	117
B.1	Description des données climatiques	117
C	L'analyse multidimensionnelle	119
C.1	Description du schéma MEDICAL_STATISTICS et ses dimensions	119
C.2	Description du schéma ENVIRONMENT et ses dimensions	119
C.3	Description du schéma HEALTH_STATUS_OF_POPULATION et ses dimensions	120

C.4	Description du schéma DEMOGRAPHY et ses dimensions	121
C.5	Description du schéma SIMULATION et ses dimensions	122
C.6	Description du schéma STATUT et ses dimensions	123
Bibliographie		135

Table des figures

1.1	Représentation du processus de contagion	11
1.2	Schéma de transmission du modèle SI	12
1.3	Schéma de transmission du modèle SIR	13
1.4	Exemple du graphe aléatoire d'Erdős et Rényi	18
1.5	Le modèle petit monde (Small World) proposé par Watts Strogatz	19
1.6	Le réseau de Kleinberg	20
1.7	Le diagramme des différentes étapes de l'analyse des réseaux de contact	21
2.1	L'évolution des variables d'états dans le temps pendant une épidémie : (a) : le modèle SIR (Pas de phase de latente), (b) : le modèle SEIR (avec le temps de latente)	27
2.2	Représentation schématique du modèle de métapopulation [Meloni <i>et al.</i> , 2011]	35
2.3	Le schéma du modèle. (A) La couche spatiale, basée sur l'approche de la métapopulation, décrit la structure de l'espace et la mobilité des individus. (B) La couche sociale décrit la structure de contact au sein de chaque sous-population	36
3.1	Schéma général de la démarche classique de modélisation	42
3.2	Le modèle SEIR-SW proposé	43
3.3	Représentation schématique de la liste d'adjacence et le graphe généré	46
3.4	Distribution des degrés dans un graphe SW	47
3.5	Représentation schématique du modèle de diffusion épidémique SEIR	47
3.6	L'évolution du nombre d'infectés de la grippe	50
3.7	Représentation schématique décrivant le processus de l'algorithme d'estimation pour une vague d'épidémie	51
3.8	L'ajustement des données réelles et simulées	52
3.9	Visualisation du réseau SW pour la première vague de la grippe : (a) degré de distribution, (b) modularité, et (c) coefficient de clustering	57
3.10	Visualisation du réseau SW pour la deuxième vague de la grippe : (a) degré de distribution, (b) modularité, et (c) coefficient de clustering	57
4.1	Les principaux composants d'un SID	61
4.2	Les principales composantes d'un entrepôt de données [Kroenke et Auer, 2011]	63
4.3	Un exemple de modèle multidimensionnel d'une vente	64
4.4	Une présentation schématique du processus ETL	66

4.5	Architecture d'un système ROLAP	69
4.6	Architecture d'un système MOLAP	69
4.7	Architecture d'un système HOLAP	70
4.8	Les composants d'un système de reporting [Kroenke et Auer, 2011]	71
5.1	EPI-SDSS : une vue d'ensemble [Younsi <i>et al.</i> , 2014b]	80
5.2	Le modèle multidimensionnel proposé par [Bouba <i>et al.</i> , 2013]	81
6.1	Le système décisionnel SYDSEP proposé	89
6.2	Le processus interactif de SEIR-SW	90
6.3	Le processus ETL	91
6.4	La démarche décisionnelle proposée	92
6.5	Le diagramme de cas d'utilisation	93
6.6	Le diagramme de séquences	94
6.7	Représentation schématique des données constituant l'entrepôt de données .	95
6.8	Schéma du modèle multidimensionnel d'épidémiosurveillance	99
6.9	Le découpage administratif de la ville d'Oran	100
6.10	Représentation du cube multidimensionnel des statistiques médicales	102
6.11	Représentation des dimensions hiérarchisées	102
6.12	Représentation des opérations <i>Slice</i> et <i>Dice</i> sur le cube	103
6.13	Représentation graphique de requêtes générées : (a) l'incidence; (b) le taux d'attaque	103
6.14	Affichage cartographique du résultat de la requête des incidences : (a) la semaine 52/2009, (b) les mois entre 10/2009 jusqu'à 12/2009	104
6.15	Représentation du cube multidimensionnel des simulations	105
6.16	Représentation graphique de requêtes générées : (a) du taux d'attaque; (b) l'évolution de la grippe	106
6.17	Représentation cartographique de : (a) la distribution de taux d'attaque, (b) l'évolution de la grippe dans la commune d'Arzew	106
6.18	Représentation cartographique de la distribution des facteurs climatologiques	107
B.1	La table d'échelle de Beaufort	118
C.1	Modèle Dimensionnel en étoile du fait « MEDICAL_STATISTICS »	119
C.2	Modèle Dimensionnel en étoile de l'activité « ENVIRONMENT »	120
C.3	Modèle Dimensionnel en étoile de l'activité « HEALTH_STATUS_OF_POPULATION »	121
C.4	Modèle Dimensionnel en étoile de l'activité « DEMOGRAPHY »	121
C.5	Modèle Dimensionnel en étoile de l'activité « SIMULATION »	122
C.6	Modèle Dimensionnel en étoile de l'activité « STATUT »	123

Liste des tableaux

1.1	Quelques extensions du modèle SIR	14
1.2	Quelques estimations du nombre de reproduction \mathcal{R}_0	16
1.3	Présentation des différentes structures avec leurs mesures	23
2.1	Comparaison entre la modélisation par équations et par agent	32
2.2	La classification des différents types de réseaux	33
3.1	Synthèse des paramètres et des variables d'état utilisés par SEIR-SW . . .	48
3.2	Les valeurs des paramètres et des variables d'état utilisées par SEIR-SW .	52
3.3	La corrélation entre les données des simulations numériques de SEIR-SW et les données observées	54
3.4	Illustration des résultats du test de Kolmogorov–Smirnov	55
3.5	L'analyse du réseau SW pour les deux vagues de la grippe	56
4.1	Comparaison des systèmes ROLAP, MOLAP et HOLAP [Boulil, 2012] . . .	70
5.1	Synthèse des applications des SIG en santé publique	78
5.2	Synthèse des travaux appliquant les ED et la fouille de données en santé publique	82
5.3	Synthèse des travaux appliquant les ED et l'analyse des réseaux sociaux en santé publique	83

Remerciements

CETTE THÈSE a été réalisée en cotutelle entre le laboratoire ERIC de l'Université Lumière Lyon 2 et le laboratoire LIO de l'Université d'Oran 1 Ahmed Ben Bella. Mes remerciements s'adressent particulièrement à toutes les personnes qui ont contribué au bon déroulement de ce travail.

En premier lieu, je tiens à remercier profondément mes encadreurs de thèse Omar BOUSSAID et Djamila HAMDADOU qui m'ont guidée tout au long de ces années. Je tiens à exprimer ici ma reconnaissance pour leurs soutiens, leurs encouragements et pour les discussions toujours constructives, j'ai réellement apprécié de travailler à ses côtés. Je les remercie aussi pour leurs accueils au sein de leurs laboratoire de recherche (le laboratoire LIO " Université d'Oran 1" et le laboratoire ERIC " Université de Lumière Lyon 2"). Je tiens également à remercier les membres des deux laboratoires pour leur présence et leur collaboration. J'adresse aussi mes remerciements à toute l'équipe Modélisation Spatio-Temporelle et Vision Artificielle : du Capteur à la Décision, dirigée par Mme Djamila HAMDADOU et dont je suis membre.

Je tiens également à remercier les personnes qui ont accepté de participer à mon jury de thèse :

Monsieur Karim BOUAMRANE pour m'avoir fait l'honneur de présider ce jury ainsi que pour le temps qu'il m'a accordé en acceptant de participer à ce jury.

J'adresse mes plus sincères remerciements aux rapporteurs de cette thèse, Monsieur Gilles ZURFLUH et Monsieur Sidi Mohamed BENSLIMANE qui ont accepté d'évaluer ce travail de recherche.

Un merci particulier à Madame Fadila BENTAYAB d'avoir bien voulu être examinatrice dans ce jury. Qu'elle trouve ici toute ma reconnaissance.

Un grand merci également à Ahmed BOUNEKKAR pour toutes ses discussions enrichissantes et de l'aide qu'il m'a apporté surtout pour la première partie de cette thèse.

Finalement, je réserve des remerciements très particulière à mes parents et ma famille en leur exprime ici toute ma gratitude pour leurs soutiens sans limite, sans lesquels ce travail n'aurait pu être réalisé. Je tiens à remercier, aussi, les autres personnes, qui ont contribué de façon directe ou indirecte dans la réalisation de cette thèse.

A tous, Merci ...

Résumé

Les maladies infectieuses représentent aujourd'hui un problème majeur de santé publique. Devant l'augmentation des résistances bactériennes, l'émergence de nouveaux pathogènes et la propagation rapide de l'épidémie, le suivi et la surveillance de la transmission de la maladie devient particulièrement importants. Face à une telle menace, la société doit se préparer à l'avance pour réagir rapidement et efficacement si une telle épidémie est déclarée. Cela nécessite une mise en place des dispositifs de suivi et de prévention.

Dans ce contexte, nous nous intéressons, dans le présent travail, à l'élaboration d'un Système d'Information Décisionnel Spatio-temporel pour le suivi et la surveillance du phénomène de propagation de l'épidémie de la grippe saisonnière au sein de la population de la ville d'Oran (Algérie). L'objectif de ce système est double : il consiste, d'une part, à comprendre comment l'épidémie se propage par l'utilisation du réseau social Small World (SW) et du modèle à compartiments d'épidémie SEIR (Susceptible-Exposed-Infected-Removed), et d'autre part, à stocker dans un entrepôt les données multiples tout en les analysant par un outil d'analyse en ligne de donnée Spatiale dit SOLAP (Spatial On-Line Analytical Processing).

Mots clés : Modèle à compartiments SEIR (Susceptible-Exposed-Infected-Removed), Réseau social Small World (SW), Modèle hybride SEIR-SW, Entrepôt de données, Spatial OLAP, Simulation, Système d'information à référence spatiale, Analyse du réseau social.

Abstract

Today, infectious diseases represent a major public health problem. With the increase of bacterial resistance, the emergence of new pathogens and the rapid spread of epidemic, monitoring and surveillance of disease transmission becomes important. In the face of such a threat, the society must prepare in advance to respond quickly and effectively if an outbreak is declared. This requires setting up monitoring mechanisms and prevention.

In this context, we are particularly interested by development a Spatiotemporal decision support system for monitoring and preventing the phenomenon of seasonal influenza epidemic spread in the population of Oran (city at Algeria). The objective of this system is twofold : on one hand, to understand how epidemic is spreading through the social network by using SEIR (Susceptible-Exposed-Infected-Removed) compartmental model within Small World network, and on the other hand, to store multiple data in data warehouse and analyzing it by a specific online analysis tool Spatial OLAP (Spatial on-line Analytical Processing).

Keywords : Compartmental SEIR (Susceptible-Exposed-Infected-Removed) model, Small World social network (SW), SEIR-SW hybrid model, Data warehouse, Spatial OLAP, Simulation, Spatial Reference System, Social network analysis.

Acronymes

Acronyme	Définition
AD	Aide à la Décision
ARS	Analyse de Réseau Social
ED	Entrepôts de données
OLAP	On-Line Analytical Processing
RS	Réseau Social
SEIR	Susceptible, Exposed, Infected, Removed
SW	Small World
SEIR-SW	Susceptible Exposed Infected Removed within Small World network
SID	Système d'Information Décisionnel
SIG	Système d'Information Géographique
SOLAP	Spatial On-Line Analytical Processing
SYDSEP	SYstème Décisionnel Spatiotemporel pour l'EPidémiologie

Introduction Générale

Dans la présente introduction, nous exposons le contexte de ce travail de thèse montrant la complexité de la gestion de suivi, d'alerte et de prévention d'épidémie ainsi que les objectifs d'un système d'information dédié à cette gestion.

Nous explicitons, également, la problématique abordée ainsi que les contributions liées aux choix d'approches et d'outils adaptés à la présente étude.

Mise en contexte

Les maladies infectieuses préoccupent de plus en plus les autorités et les responsables de la santé publique. En effet, face à l'augmentation des résistances bactériennes, l'émergence de nouveaux pathogènes et la propagation rapide de l'épidémie, la surveillance et la prévention de la transmission de la maladie devient, particulièrement, importante et indispensable. Malgré une surveillance permanente et continue des maladies infectieuses, il est à constater que leurs étiologie¹ reste encore largement méconnue.

Trois pandémies de grippe ont été recensées au XXème siècle :

- (i) la grippe espagnole de 1918-1919 à virus H1N1 qui fit entre 40 et 100 millions de morts ;
- (ii) la grippe asiatique en 1957 à virus H2N2 qui fit entre 1 et 4 millions de morts ;
- (iii) la grippe de Hong Kong en 1968 à virus H3N2 qui fit entre 1 et 2 millions de morts.

En effet, cette maladie est la cause d'une morbidité et d'une mortalité importante jusqu'à présent car selon OMS², l'estimation de la mortalité par la grippe saisonnière est évaluée à presque 250 000 à 500 000 et entre 3 et 5 millions de cas graves par an dans le monde.

En épidémiologie, les maladies infectieuses possèdent deux concepts : **l'infectiosité** et **la contagiosité** [Guégan et Choisy, 2008]. Dans le premier processus, un individu contracte une maladie infectieuse lorsqu'il rentre en contact avec un foyer qui représente le site principal de la maladie à partir duquel elle se propage et qui peut être de plusieurs natures. Dans le deuxième processus, on recherche les modes et les voies de transmission de la maladie d'un individu vers un autre selon les modes de transmission : *Direct* (du réservoir du germe à l'hôte) ou *Indirect* (par l'intermédiaire d'un vecteur ou d'un support qui véhicule le germe).

Certaines problématiques liées à la propagation des maladies transmissibles sont de nature complexe et font intervenir de nombreux facteurs issus de domaines différents, de nature qualitative et/ou quantitative. Ces facteurs sont considérés comme les principales causes

1. L'étude des causes et des facteurs d'une maladie

2. Organisation Mondiale de Santé : <http://www.who.int>

de la propagation de ces maladies au sein de la population humaine. Ils sont multiples et liés à plusieurs facteurs qu'il convient d'analyser dans une dimension globale, à savoir : **les caractéristiques et la dynamique de la maladie** (mode de transmission, durée d'infections, etc.), **la structure sociale des individus** comme la concentration de la population dans les zones urbaines et les rencontres, **les facteurs socio-économiques** (démographie inclue), et **l'environnement** (climat, pollution, etc.).

A ce titre, l'évolution des maladies infectieuses causée par ces facteurs a bien entendu des conséquences sur la croissance économique, le développement, la sécurité et la durabilité. Par conséquent, plusieurs équipes de recherches à travers le monde ont modélisé l'impact de mesures de contrôle sur une épidémie, ainsi que plusieurs initiatives nationales et internationales visant à rétablir et à améliorer les dispositifs de surveillance et de lutte contre les maladies transmissibles.

En Algérie, et comme tous les pays du monde, les responsables de la santé publique sont soumis à une nécessité d'un suivi et d'une prévention de ce phénomène émergent car il est devenu de plus en plus inquiétant, vue l'augmentation dans le nombre d'incidences dans le pays. L'Institut National de Santé Publique (INSP³) a pour objet de réaliser des travaux d'étude et de recherche en santé publique et en épidémiologie. Il publie périodiquement des rapports et des statistiques relatifs au phénomène de propagation d'épidémie touchant différentes catégories d'âges de la population et les différents espaces géographiques. À ce jour, peu de systèmes informatiques sont mis en place afin d'aider les responsables de la santé publique dans l'analyse de leurs données multidisciplinaires.

D'une manière générale, les objectifs d'une surveillance épidémiologique sont [Tibiche, 2011] :

1. de connaître l'incidence et les caractéristiques d'une maladie infectieuse ;
2. d'étudier la dynamique de diffusion sociale, temporelle et spatiale d'une maladie et d'en prédire l'extension ;
3. de disposer de système et d'indicateurs d'alerte d'épidémie afin d'intervenir lors d'une épidémie pour interrompre la chaîne de transmission ;
4. de connaître les facteurs de risque des infections afin de proposer les mesures de prévention et des recommandations les plus adaptées ;
5. d'évaluer les actions de prévention.

Les systèmes de surveillance épidémiologique élaborés ont divers objectifs, tels que : le suivi, la surveillance, la prévention, le veille, l'alerte, etc. En effet, les auteurs dans [Toma *et al.*, 1991], considèrent ces systèmes comme des outils d'aide à la décision dans le domaine de la prévention et du contrôle des maladies épidémiologiques. Toutefois, à notre point de vue, ces systèmes souffrent des lacunes liées aux aspects analytiques ainsi que leur limite de traiter plusieurs actions à la fois, comme : le suivi, la prévention et l'alerte. De plus, il est impossible de prédire fermement et par avance qu'une maladie se propagera ou non dans le futur. Dans cette optique, la prévention vise à limiter ou freiner la progression de l'épidémie car on ne pourra jamais stopper d'une manière définitive l'épidémie du fait qu'elle est multifactorielle.

L'épidémie se propage au sein d'une population dans le temps comme dans l'espace. Afin de maîtriser et surveiller sa dynamique d'infections ainsi que sa capacité d'évoluer et « émerger » sous sa forme, il convient de l'analyser dans une dimension spatiotemporelle. Les systèmes d'information à référence spatiale sont, maintenant, très répandus dans le

3. L'Institut National de Santé Publique Algérien placé sous la tutelle du Ministère de la Santé et ayant pour objet de réaliser des travaux d'étude et de recherche en santé publique.

monde scientifique, et sont reconnus comme des outils valables dans toutes les disciplines qui s'intéressent à la surface de la terre et à ses populations. Ces systèmes sont, également, considérés comme une technologie d'intégration qui recouvre plusieurs disciplines et brouille les frontières qui les séparent [Goodchild, 2000].

Problématique et contribution

La problématique abordée dans ce travail se situe aux confins de diverses disciplines dont la principale est l'épidémiologie. Les foyers des maladies transmissibles peuvent constituer une menace pour la santé des individus. En effet, plusieurs équipes de recherches ont investi dans l'étude du mécanisme de propagation d'épidémie afin de proposer un système de surveillance fiable et robuste.

En 1760, Daniel Bernoulli [Bernoulli, 1760] a développé un modèle mathématique d'une épidémie de variole (appelée à l'époque « petite vérole ») pour une nouvelle analyse de la mortalité causée par la petite vérole, et des avantages de l'inoculation pour la prévenir. Son modèle est connu comme étant le premier modèle mathématique de l'histoire de l'épidémiologie. De nombreux travaux ont suivi ce travail afin de comprendre le mécanisme de propagation de la maladie transmissible.

Dans cette optique, la simulation par des modèles mathématiques paraît un bon outil pour comprendre le mécanisme de propagation des maladies infectieuses. Plusieurs modèles existent dans la littérature permettant de modéliser les différents statuts de la maladie durant l'infection, en divisant la population hôte en plusieurs classes représentant les différents états de santé durant une maladie, tels que : SI (Susceptible, Infected), SIR (Susceptible, Infected, Removed), SEIR (Susceptible, Exposed, Infected, Removed), etc.

Dans la présente étude, nous visons les objectifs suivants :

1. la proposition d'un modèle de propagation SEIR-SW pour contribuer, efficacement, à la modélisation et la simulation d'une catégorie de systèmes complexes intégrant à la fois l'aspect biologique ainsi que l'aspect social de la population ;
2. la mise en place d'un SYstème Décisionnel Spatiotemporel pour l'EPidémiologie (SYDSEP) permettant de supporter : la modélisation et la simulation d'un processus de propagation issue du monde réel, l'intégration et le stockage ainsi que l'exploitation de données provenant de sources multiples tout en assurant la gestion d'entrée/sortie du simulateur SEIR-SW.

Dans le modèle de propagation, nous nous intéressons, particulièrement, au modèle à compartiments SEIR, ce choix est motivé par le fait que ce modèle prend en considération la durée de latence de la maladie (classe Exposed). Pendant cette période, l'individu est infecté mais il n'est pas infectieux (il ne peut pas transmettre la maladie aux autres individus). De plus, ce modèle est plus adapté dans le cas de l'immunité permanente (les malades sont immunisés de la souche de la grippe après leurs guérison). Les modèles SEIR modélisent mieux les maladies caractérisées par une période de latence par la présentation du processus complet et simple de la maladie en cas d'une souche, telle que : la grippe, la rougeole, la variole, etc.

Les modèles traditionnels épidémiologiques supposent que la transmission d'une infection dans une population hôte homogène augmente bien avec le nombre d'individus et avec un risque d'infection aléatoire [Anderson *et al.*, 1992][May, 1997]. Par ailleurs, la transmission d'une épidémie se fait par un contact direct entre un susceptible et/ou un infecté comme le cas de la grippe saisonnière.

De plus, la topologie de la structure sociale des individus a un impact sur la transmission de la maladie car les individus avec leurs relations forment un réseau social. Ce dernier n'est pas qu'une simple liste de noms et un ensemble de relations mais plutôt un système où les liens interagissent les uns avec les autres et subissent des transformations et des changements dans le temps et dans l'espace. Par conséquent, l'intégration du modèle à compartiments SEIR avec le modèle réseau social s'avère nécessaire.

Quelques études en épidémiologie se sont focalisées sur la structure du réseau social et les processus de transmission et de diffusion des informations au sein des réseaux sociaux [Zekri et Clerc, 2002][Newman et Girvan, 2004], mais il n'existe que très peu d'études où le phénomène de transmission et les facteurs influençant sont étudiés [Basileu *et al.*, 2010]. Dans le présent travail, nous proposons un modèle hybride (SEIR-SW) basé sur une intégration du modèle SEIR et le modèle réseau social de type petit monde (Small World), ce choix du réseau est dû aux caractéristiques de ce dernier liées à sa structure topologique proche du monde réel et ses propriétés structurelles : petit diamètre et coefficient de clustering élevé [Watts et Strogatz, 1998]. Le modèle hybride SEIR-SW génère des données simulées de l'ensemble de la population ciblée dans le temps et dans l'espace. En outre, ce modèle permettra de générer différents scénarios de propagation d'épidémie par le changement des paramètres initiaux du modèle qui jouent un rôle primordial dans l'évolution de la maladie. Le modèle hybride (SEIR-SW) suggéré a pour objectif principal :

- d'une part, de comprendre comment l'épidémie se propage au sein d'une population ?, et quels sont les facteurs favorisant cette transmission ?,
- et d'autre part, de prévoir l'apparition future des foyers épidémiques naissants afin de les réduire et les éradiquer.

L'estimation des paramètres inconnus du modèle de propagation SEIR-SW est une étape nécessaire dans notre étude. Cette estimation a été réalisée par ajustements successifs sur plusieurs simulations. Les paramètres retenus étant ceux qui minimisent les écarts entre l'épidémie simulée et l'épidémie observée. Par ailleurs, notre étude a permis d'estimer ces paramètres et de dégager d'autres ayant une forte influence sur l'évolution de l'épidémie. Toutefois, les paramètres de SEIR-SW ne sont pas les seuls ayant un impact sur la diffusion de la maladie. En épidémiologie de la grippe, plusieurs facteurs influençant la survenue de la maladie et leur maîtrise facilitent le processus de la prévention. L'objectif de notre système SYDSEP est double : il permet, d'une part :

- d'étudier la distribution géographique dans le temps tel que l'espace de la propagation des maladies infectieuses au sein des populations humaines ;
- de déterminer et d'analyser les facteurs qui favorisent l'apparition et la progression de la maladie dans ces populations afin de freiner l'émergence de la maladie.

Et d'autre part, la compréhension des processus d'infection afin d'identifier des moyens de lutte et des conditions de leur mise en œuvre. En outre, la prévision de l'issue finale de l'épidémie afin de connaître l'ampleur et leur diffusion ou l'impact des procédures de lutte.

Le système SYDSEP est constitué, principalement, de trois sous-systèmes, à savoir :

1. Un sous-système de simulation,
2. Un sous-système d'entreposage,
3. Un sous-système d'aide à la décision.

Contributions originales (approches et outils)

Ce projet de thèse vise à apporter des nouveaux concepts permettant aux décideurs de la santé publique le suivi, l'alerte et la prévention de la propagation d'épidémie afin de réagir rapidement et efficacement devant une survenue d'épidémie.

Contributions originales du simulateur SEIR-SW

L'objectif de la modélisation d'épidémie est de mettre au point une représentation réaliste du phénomène de propagation d'épidémie par : la présentation des différents états d'un individu durant une maladie, l'étude de l'évolution de la maladie dans le temps et dans l'espace et la compréhension des facteurs qui pourraient causer la transmission. Classiquement, les modèles compartimentaux d'épidémie reposent sur les systèmes d'équations différentielles ordinaires (EDO) pour étudier le phénomène complexe de la propagation d'épidémie. Ces modèles considèrent que le degré des nœuds (le nombre de voisins d'un individu dans le réseau social) est fixe, égale à une matrice de valeurs figées ou à une loi de distribution. Cependant, ces modèles ont échoué à représenter les individus distribués d'une façon hétérogène dans l'espace et formant un réseau social afin d'effectuer une analyse de ce réseau. Il est donc important de considérer cette hétérogénéité et le contexte géographique dans lequel la maladie se développe pour préciser le risque épidémiologique. L'objectif ici est de contribuer à élaborer un modèle permettant de modéliser et de comprendre l'évolution de l'épidémie grippale. Ce modèle est basé sur un réseau social où le degré des nœuds est variable pour représenté les relations dans le monde réel. En outre, il permet une analyse de ces réseaux afin de comprendre les facteurs sociaux influençant la propagation d'épidémie.

Plus concrètement, notre démarche consiste à formaliser, dans un programme informatique, le réseau social SW pour la création des individus représenté par des nœuds et des relations de rencontres entre ces individus représentées par des arrêts. Le but d'une telle création consiste à modéliser et analyser les différents types de liens existant entre les individus de la population. Ensuite, l'application de modèle SEIR pour étudier la transmission et l'évolution de la maladie au niveau individuel. Et l'évolution du réseau sociale (graphe sociale) dans le temps. Ce simulateur SEIR-SW a pour objectif d'estimer les paramètres inconnus de la maladie par l'utilisation de la méthode d'ajustement.

Contributions originales de SID

Face aux problèmes de santé, les décideurs manquent d'outils cartographiques simples et puissants pour explorer, synthétiser et analyser des données spatio-temporelles telle que la détermination des zones à risque. En effet, la visualisation des foyers épidémiques sur la carte géographique est souvent gérée par le SIG (Système d'Information Géographique). Toutefois, cette technologie souffre encore de plusieurs lacunes dues en grande partie à un manque de capacités analytiques capables de supporter les problèmes multidimensionnels suite aux progrès des solutions issues du domaine de l'informatique décisionnelle. La solution la plus diffusée pour faire évoluer les SIG vers un vrai outil d'aide à la décision est de les coupler avec le moteur OLAP, car la maîtrise de la complexité des problèmes décisionnels à référence spatiale nécessite l'utilisation de méthodes, de techniques et d'outils d'analyse puissants, qui doivent non seulement gérer mais aussi analyser des données multiples. Des solutions, telles que le SOLAP, permettant de supporter les données spatiales dans l'OLAP, ont été développées par Bédard [Bédard *et al.*, 1997].

L'originalité de ce travail de recherche porte aussi sur l'intégration du processus SW-SEIR dans un SID afin d'améliorer les insuffisances d'étude spatio-temporelle et géographique du processus SW-SEIR et augmenter les performances du SID. A notre connaissance, aucune étude n'avait traité de la propagation d'épidémie dans le contexte de suivi, prévention et alerte à la fois en stockant les données réelles multidisciplinaires et simulées dans un entrepôt de données puis effectuant une analyse multidimensionnelle SOLAP en montrant l'apport de cet outil dans le suivi, la surveillance et la prévention de propagation d'épidémie dans la région d'étude afin de fournir un outil d'aide à la décision spatiotemporel au responsable du programme de lutte contre l'épidémie de la grippe saisonnière. En effet, ce travail de thèse illustre une marche difficile au travers des problèmes de la complexité dans les systèmes vivants. Nous sommes partis d'une problématique de nature complexe et multidisciplinaire. La complexité du processus dynamique de transmission de la maladie nous conduit à la transposer dans un formalisme mathématique lié à la théorie des graphes, qui nous a permis de proposer un nouveau modèle étudiant la propagation des maladies infectieuses. L'utilisation d'entrepôt de données et le moteur SOLAP permettra une analyse multidimensionnelle des données multiples dédiées à l'épidémiosurveillance. La mise en œuvre du modèle proposé nous a conduit à un outil formel capable de répondre à notre problématique.

Plan du document

D'une manière générale, la thèse est composée de deux parties : (i) La modélisation mathématique et les réseaux complexes et (ii) Les systèmes d'Information Décisionnels. Elle est organisée de la manière suivante :

Première partie :

Chapitre I : présente le contexte général de la modélisation mathématique en épidémiologie et au sein des réseaux complexes en particulier. Nous présentons d'abord quelques concepts de base, puis nous décrivons la modélisation mathématique en épidémiologie dans le but d'introduire les notions et le vocabulaire utilisés dans cette thèse. Ensuite, nous présentons un panorama de modèles mathématiques et modèles réseaux complexes existant dans la littérature suivi d'une illustration des propriétés des réseaux complexes.

Chapitre II : étale un état de l'art sur : les modèles mathématiques existant dans la littérature ainsi que les modèles à compartiments traitant l'impact des réseaux complexes sur la propagation des maladies.

Chapitre III : décrit notre première contribution consistant à proposer le modèle SEIR-SW dédié à la modélisation et la simulation du comportement individuel et global du système étudié. Le modèle proposé est composé, essentiellement, de deux modèles, à savoir : le modèle à compartiments et le modèle réseau petit monde "Small World".

Dans ce même chapitre, une étude expérimentale du générateur SEIR-SW est menée afin d'étudier le phénomène de propagation de la grippe saisonnière dans la région d'Oran. Enfin, nous validons notre proposition par des méthodes statistiques.

Deuxième partie :

Chapitre IV : fournit une présentation générale des SID. Nous énonçons, tout d'abord, les notions élémentaires se rapportant aux SID et leurs différentes architectures ainsi que la modélisation multidimensionnelle. Nous présentons, par la suite, la phase de restitution dans le SID par l'exploitation de l'outil OLAP et SOLAP.

Chapitre V : étale un état de l'art sur les systèmes d'épidémiosurveillance et les ED

médicales. Nous présentons une classification des différentes approches et nous exposons un ensemble de travaux intégrant l’aspect social dans les SID médicales.

Chapitre VI : décrit notre seconde contribution visant à intégrer le simulateur SEIR-SW dans un SID. Nous décrivons, succinctement le SID proposé SYDSEP pour la gestion de la santé publique. Nous détaillons ensuite l’architecture multidimensionnelle proposée pour notre entrepôt de données spatiales.

Troisième partie :

La troisième partie (**Annexes**) aborde, sous forme d’annexes, les différents domaines liés aux problématiques abordées :

- **Annexe A** : est dédiée à un glossaire des termes utilisés en épidémiologie.
- **Annexe B** : introduit les notions générales sur les facteurs climatiques.
- **Annexe D** : est consacrée à l’étude menée sur l’analyse multidimensionnelle.

Le présent document s’achève par une **conclusion** où nous récapitulons les apports de la présente étude tout en mettant l’accent sur l’utilité des propositions d’approches et d’outils dans le domaine de la santé publique et ; ouvrons des perspectives à la présente étude.

Première partie

La modélisation mathématique et les réseaux complexes

Chapitre 1

Concepts et définitions de base

Sommaire

1.1	Introduction	9
1.2	La modélisation mathématique	10
1.2.1	Concepts de base	10
1.2.2	La modélisation mathématique en épidémiologie	11
1.2.3	Quelques modèles mathématiques de transmission de maladie	12
1.3	Les réseaux complexes	16
1.3.1	L'épidémiologie par réseaux de contacts	17
1.3.2	L'analyse structurelle des réseaux	20
1.4	Conclusion	23

1.1 Introduction

L'épidémiologie mathématique a pris une importance considérable depuis sa naissance au XVIII^e siècle par le mathématicien Daniel Bernoulli qui a publié un travail sur la maladie de la variole. Il s'agit en fait du premier modèle mathématique inspiré d'une problématique médicale et plus précisément épidémiologique, marquant ainsi la naissance d'un domaine dénommé aujourd'hui « Biomathématique » [Dreyer et Gabriel, 2010].

Récemment, l'application de la modélisation mathématique au sein des réseaux complexes a connu un essor important car la plupart des réseaux du monde réel modélisés sont de type dynamique et complexe. Afin d'étudier un phénomène émergent comme la propagation d'épidémie, nous sommes dans le besoin de comprendre la dynamique du système et les mécanismes de base de son fonctionnement. Cependant, ces deux aspects sont devenus inséparablement associés.

Dans ce chapitre, nous allons explorer les différents concepts de base liés à la modélisation mathématique en épidémiologie. Nous présentons, également, un panorama de classes de modèles épidémiologiques, ainsi qu'une interprétation avec la présentation des principales propriétés de chacun. Nous étudierons, par la suite, les réseaux complexes, nous présentons quelques définitions et préliminaire liés au concept étudié, les caractéristiques, les propriétés et la structure des réseaux. Nous nous intéressons, particulièrement, aux méthodes d'analyse et à leur évolution.

1.2 La modélisation mathématique

Avant d'entamer le contenu de la cette section, nous voudrions commencer par quelques définitions et notations générales.

1.2.1 Concepts de base

1.2.1.1 Que ce qu'un modèle ?

Un modèle est une description de la réalité établie dans un langage précis, mais cette description doit permettre des prévisions ou des explications [Bertrandias et Bertrandias, 1997] :

- *Prévisions* des valeurs de certains variables difficilement accessibles ou des valeurs futures, par exemple au moyen de calculs, de dessins ou du fonctionnement d'un mécanisme analogique ;
- *Explications* de phénomène complexes à partir de phénomènes simplificatrices, de principes généraux ou par coordination de mécanismes simples.

1.2.1.2 Quels sont les objectifs de la modélisation d'épidémie ?

Selon Daley [Daley *et al.*, 2001], la modélisation des épidémies vise trois objectifs principaux :

- Mieux comprendre les mécanismes par lesquels les maladies se propagent, ce qui implique l'importance d'une structure mathématique ;
- Prévoir l'évolution future de l'épidémie ;
- Comprendre comment contrôler la propagation de l'épidémie, par exemple, réduire le nombre de susceptibles dans le modèle par la vaccination.

1.2.1.3 Système dynamique

Rechenmann dans [Rechenmann, 2010] définit un système dynamique comme : « *Un ensemble d'entités en interaction. Du fait même de ces interactions, la valeur de grandeurs attachées à ces entités évolue dans le temps. C'est en étudiant l'évolution de ces valeurs que l'on cherche à comprendre et à prédire le comportement de ces systèmes. Cela repose sur leur représentation mathématique à l'aide d'équations différentielles* ». On peut distinguer trois types de systèmes dynamiques : les systèmes stochastiques (aussi appelés systèmes aléatoires), les systèmes déterministes et les systèmes chaotiques.

1.2.1.4 Système complexe

Un système complexe est un système composé d'un ensemble d'éléments homogènes ou hétérogènes, interagissant les uns avec les autres générant ainsi un ensemble d'interconnexions. L'ensemble de ces interactions génère, sans aucune application d'un principe [Moncion, 2008]. Nous distinguons alors deux niveaux :

- un niveau local ou micro : représentent le niveau des composants, avec des propriétés locales à chacun d'eux ;
- Un niveau global ou macro : représentant l'ensemble du système.

La différence entre les systèmes simples et les systèmes complexes réside dans le résultat des interactions entre les composants du système. Pour étudier le comportement des systèmes complexes, on doit s'intéresser non seulement au comportement de leurs constituants, mais également à la manière dont ils se combinent pour façonner le comportement du système [BenAmor, 2008].

1.2.2 La modélisation mathématique en épidémiologie

La modélisation mathématique en épidémiologie connaît depuis quelques années un essor important dans la littérature, grâce aux nombreux avantages qu'elle apporte, elle peut traiter une variété de problèmes liée au domaine d'épidémiologie, telles que : la prédiction de l'évolution de la maladie à partir des données réelles, l'étude de la dynamique de l'épidémie afin d'identifier les solutions de contrôle les plus efficaces, ainsi, qu'elle permettra un suivi et une surveillance de l'épidémie, l'estimation des paramètres caractérisant la maladie, etc.

Les modèles mathématiques en épidémiologie peuvent être classés en quatre principales catégories : les modèles discrets ou continus, les modèles d'EDO (Equations Différentielles Ordinaires), les modèles d'EDP (Equations aux Dérivées Partielles) et les modèles déterministes ou stochastiques.

La transmission d'une maladie infectieuse passe par plusieurs étapes, ces dernières peuvent être organisées sous forme d'un processus nommé : le processus de contagion ou d'infection.

Dans ce qui suit, nous introduirons le processus de l'infection suivi par une présentation de quelques modèles mathématiques de la propagation d'épidémies.

Le processus dynamique de l'infection

La dynamique de la maladie et les conditions environnementales ont un impact important sur la diffusion de la maladie à large échelle. Un individu est atteint d'une maladie infectieuse lorsqu'il est mis en contact avec un foyer (appelé aussi **agent pathogène**) qui peut être de diverses natures (un individu infecté, un moustique, un puits, etc.). Le processus de contagion fait référence à un événement de transfert de l'infection d'un agent pathogène à un autre.

Dans la Figure 1.1, nous illustrons le contexte dans lequel cet événement prend place :

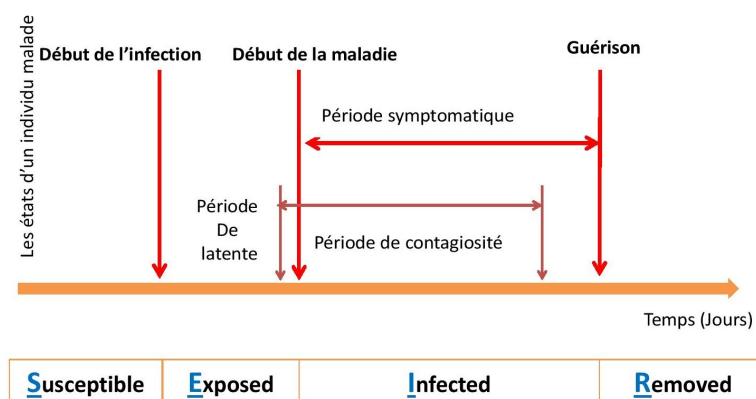


Figure 1.1 – Représentation du processus de contagion

Cette présentation schématique démontre les différentes phases (états de la maladie). Le caractère **contagieux** d'un individu n'est acquis qu'après un temps de **latence** après

l'infection. L'individu **infecté** reste **contagieux** un certain temps : il est alors soit prêt à **contracter** à nouveau la maladie, soit **résistant** à une nouvelle infection, soit mort. Cette dimension temporelle des mécanismes de contagion, s'exprime dans la modélisation des systèmes dynamiques [Guigal, 1995].

1.2.3 Quelques modèles mathématiques de transmission de maladie

Daniel Bernoulli [Bernoulli, 1760] fut l'un des premiers mathématiciens qui a tenté de modéliser les effets de la maladie dans une population. Il a utilisé un modèle déterministe où il a proposé un modèle pour estimer les avantages de l'inoculation variolique dans le but de réduire le taux de mortalité dans la population Française. A côté des modèles déterministes, il existe également d'autres modèles dénommés stochastiques. Dans la littérature associée à la modélisation mathématique d'épidémie, les modèles les plus connus sont les modèles compartimentaux.

Dans ce travail de thèse, nous nous intéressons à ce type de modèles car ils sont très largement utilisés en permettant d'appréhender, à travers leur ajustement aux données disponibles, les facteurs d'hétérogénéité responsables de la propagation observée dans les populations hôtes. Ceci permet de définir les priorités qui doivent prévaloir dans la récolte des données afin d'affiner le paramétrage des modèles [Sauvage et Pontier, 2005]. Nous distinguons deux sortes de systèmes dynamiques couramment utilisées :

- *Déterministes*, basés sur un système d'équations différentielles ordinaires ou partielles ;
- *Stochastiques*, basés sur des équations différentielles stochastiques ou sur une modélisation markovienne par processus de saut.

1.2.3.1 Modèles déterministes

Les modèles déterministes sont des modèles dirigés par des lois mathématiques bien définis, on peut donc prévoir l'évolution du système étudié dans le temps. Dans la littérature mathématique, il existe une diversité de modèles déterministes, nous présentons dans ce qui suit les modèles les plus connus :

(a) **Le modèle simple SI (*Susceptibles, Infectives*) de Hamer**

Le SI (Susceptibles, Infectives) est le premier modèle dynamique simple, il a été développé par W. H. Hamer en 1906. Il suppose au départ qu'il n'y a ni décès ni guérison, et l'individu infecté reste et demeure contagieux jusqu'à la fin de sa vie. Ce modèle comprend donc deux compartiments : (i) les individus susceptibles d'être infectés (S), (ii) les individus infectés (I).

Dans le domaine d'épidémiologie, nous pouvons schématiser le modèle d'épidémie par des boîtes ou des compartiments qui représentent les différents statuts dans lesquels peuvent se trouver les individus d'une population pendant la maladie, les changements d'états (statuts) possibles étant indiqués par des flèches. En effet, le schéma du modèle SI est illustré par la Figure 1.2 :

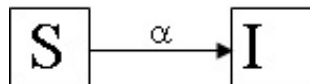


Figure 1.2 – Schéma de transmission du modèle SI

La transmission de l'infection se fait à travers un contact direct entre des susceptibles S avec un ou des infectés I avec un facteur α de proportionnalité (aussi appelé taux

d'infection ou taux de transmission).

Le nombre de nouveaux cas atteints par l'infection pendant l'intervalle de temps dt sera égale à αSI . Cela conduit à deux équations différentielles :

$$\begin{cases} dS/dt = S'(t) = -\alpha.I(t).S(t) \\ dI/dt = I'(t) = \alpha.I(t).S(t) \end{cases} \quad (1.1)$$

Avec N est la population totale, $N = S(t) + I(t)$.

(b) **Le modèle simple SIR** (*S*usceptibles, *I*nfectives, *R*emoved)

Ce modèle a été proposé en 1927 par les deux chercheurs : W.O. Kermack et A.G. McKendrick [Kermack et McKendrick, 1927], leur méthode est toujours considérée comme valide, elle est largement utilisée dans des travaux de recherches d'actualité. A l'époque, l'objectif des deux chercheurs était de comprendre pourquoi la grande pandémie de grippe espagnole de 1918 n'avait pas infecté toute la population. Le modèle est composé principalement de trois compartiments comme le montre la Figure 1.3.

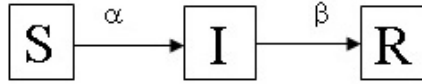


Figure 1.3 – Schéma de transmission du modèle SIR

Trois populations à définir :

$S(t)$: nombre d'individus sains

$I(t)$: nombre d'individus malades

$R(t)$: nombre d'individus retirés qui correspond à la population qui quitte le compartiment des infectés par guérison et immunisés contre la maladie, décès, etc.

À ces trois états différents, nous pouvons associer trois équations d'évolutions propres telles que :

$$\begin{cases} dS/dt = S'(t) = -\alpha.I(t).S(t) \\ dI/dt = I'(t) = \alpha.I(t).S(t) - \beta.I(t) \\ dR/dt = R'(t) = \beta.I(t) \end{cases} \quad (1.2)$$

Avec N : la population totale, $N = S(t) + I(t) + R(t)$ et $S(0) = S_0 > 0$, $I(0) = I_0 > 0$. Ce modèle est généralement utilisé pour décrire des maladies développant une immunité.

(c) **Autres extensions du modèle SIR**

Le modèle SIR est la base de la plupart des modèles étudiant le phénomène de la diffusion de la grippe [Anderson *et al.*, 1992], [Hethcote, 2000]. Il est aujourd'hui régulièrement utilisé comme par exemple dans le cas de la grippe A en 2009. De plus, ce modèle a subi des changements par l'ajout des extensions au modèle standard. Ces extensions de modèle ont tous les mêmes principes de base énoncés précédemment mais comprennent des degrés de complexité variés, nous présentons dans la Table 1.1 quelques extensions avec le schéma correspondant et leur interprétation.

Table 1.1 – Quelques extensions du modèle SIR

Le modèle	Interprétation	Schéma
SIRS	Le modèle SIR avec immunisation temporaire : les individus de la classe R réintègrent après un délai la classe des susceptibles.	<pre> graph LR S[S] -- alpha --> I[I] I -- beta --> R[R] R --> S </pre>
SEIR/ SEIRS	Le modèle SIR avec introduction d'une nouvelle classe Exposé et donc d'un nouveau compartiment qui précède l'apparition des symptômes : SEIR : immunité permanente; SEIRS : immunité temporaire.	<pre> graph LR S[S] -- alpha --> E[E] E -- beta --> I[I] I -- gamma --> R[R] R -.-> S </pre>
MSIR	Le modèle SIR avec immunité passive : un individu est né avec une immunité passive transférée naturellement de la mère.	<pre> graph LR M[M] -- mu --> S[S] S -- alpha --> I[I] I -- beta --> R[R] </pre>
MSEIR/ MSEIRS	Le modèle SIR avec immunité passive et période de latence : Avec le facteur d'immunité passive, on ajoute la période de latence : MSEIR : immunité permanente, MSEIRS : immunité temporaire.	<pre> graph LR M[M] -- mu --> S[S] S -- alpha --> E[E] E -- beta --> I[I] I -- gamma --> R[R] R -.-> S </pre>

1.2.3.2 Modèle stochastique

Dans les modèles stochastiques, les systèmes aléatoires évoluent d'une façon aléatoire (au hasard) dans tout l'espace sans qu'aucune équation qui les dirige et sans qu'aucune prévision ne soit possible dans le temps.

Les deux chercheurs L.Reed et W, H. Frost ont développé le premier modèle mathématique stochastique pour décrire avec précision la propagation des maladies dans les populations [Abbey, 1952]. Ce modèle développé dans les années 1920, est connu sous le nom du "modèle Reed-Frost d'épidémie".

Les modèles stochastiques font intervenir la statistique et les probabilités, ils conduisent à des résultats comportant des informations sur la variabilité des paramètres étudiés. De manière générale, les modèles déterministes sont plus simples que les modèles stochastiques qui tiennent compte du hasard.

1.2.3.3 L'estimation des paramètres

1. L'estimation des paramètres liés au modèle épidémique

Dans le domaine épidémiologique, nous avons souvent besoin de connaître certaines caractéristiques inconnues du modèle étudié. Plusieurs techniques d'estimation numériques de ces derniers ont été proposées, tels que : la vraisemblance, les modèles bayésiens, etc.

Les paramètres du modèle mathématique d'épidémie inconnus telle que le taux de transmission de l'infection ont eu une importance considérable pour la compréhension de la dynamique du système. En effet, l'estimation de certains paramètres est indispensable car la méconnaissance de ces derniers rends très difficile la comparaison des modèles aux données réelles. Ainsi, elle permet de dégager les paramètres clefs de l'épidémie, de tester les politiques de prévention et de faire des prévisions par l'élaboration des scenarios de propagation.

Certaines variables peuvent être mesurées (par exemple : le nombre des infectés), qui seront utilisées dans le modèle pour étudier la dynamique du système afin de pouvoir estimer les autres variables. De plus, pour connaître l'origine de la maladie (par exemple : le nombre d'infectés initial) une fois que l'épidémie soit propagée.

Dans le présent travail de recherche, nous utilisons une méthode d'estimation des états et des variables non mesurés à partir des données du terrain en se basant sur l'algorithme de Levenberg-Marquardt [Kristensen, 2004] tout en déduisant une méthode simple d'estimation des paramètres inconnus. L'idée est de comparer les résultats du modèle de propagation proposé avec les données réelles et de trouver, par conséquent, les valeurs des paramètres qui minimisent l'écart entre les résultats (sorties) du modèle et les données réelles.

L'algorithme de Levenberg-Marquardt : ou algorithme L-M, est largement utilisé pour un objectif d'optimisation. Il fut développé par Kenneth Levenberg [Levenberg, 1944], puis publié par Donald Marquardt [Marquardt, 1963]. Il permet d'obtenir une solution numérique au problème de minimisation d'une fonction non linéaire et dépendant de plusieurs variables. L'algorithme combine les avantages de l'algorithme de Gauss-Newton et l'algorithme du gradient [Kristensen, 2004]. Il est rapide et a une convergence stable. Plus stable que celui de Gauss-Newton, il trouve une solution même s'il est démarré très loin d'un minimum.

La méthode de Levenberg-Marquardt permet de trouver le minimum d'une fonction $F(x)$ exprimé par la somme des carrés des fonctions non linéaires $f_i(x)$, tel que :

$$F(x) = \frac{1}{2} \sum_{i=1}^m [f_i(x)]^2 \quad (1.3)$$

Etant donnée la jacobienne de $f_i(x)$ dénotée $J_i(x)$, la méthode de Levenberg-Marquardt cherche dans la direction donnée par la solution P :

$$(J_k^T J_k + \lambda_k I) P_k = -J_k^T f_k \quad (1.4)$$

Avec : λ_k sont des scalaires non négatifs et I est la matrice d'identité.

Pour certains scalaires Δ relatif à λ_k , le vecteur P_k est la solution du problème de minimisation sous contrainte

$$\frac{\|J_k P + f_k\|_2^2}{2} \text{ soumis à } \|P\| \leq \Delta \text{ [Gill et al., 1981].}$$

2. L'estimation du nombre de reproduction (Seuil épidémique)

A l'origine, le nombre de reproduction de base (\mathcal{R}_0) est issu de la démographie et de l'écologie, il est nouvellement utilisé en épidémiologie. Ce concept est connu comme : « *un concept clé en épidémiologie* ». Nous le définissons heuristiquement comme étant : « *le nombre moyen de nouveaux cas d'infection, engendrés par un individu infecté moyen (au court de sa période d'infectiosité), dans une population entièrement constituée de susceptibles* » [Sallet, 2010].

Il joue un rôle très important pour la prédiction, car il est relié par les trois paramètres qui peuvent diminuer l'évolution d'épidémie : la transmission, le nombre de contact d'un individu et la période d'infectiosité (contagiosité). Le \mathcal{R}_0 apparaît dans la plupart des modèles de propagation d'infections.

Flahault [Flahault *et al.*, 1988] a obtenu le seuil épidémique par le biais de la fonction suivante :

$$\mathcal{R}_0 = \alpha.k.D \quad (1.5)$$

avec D : période de contagiosité.

Si la valeur de \mathcal{R}_0 est supérieure à 1 ($\mathcal{R}_0 > 1$) alors chaque individu infecté est capable de « se reproduire » en infectant plus qu'un seul autre individu, ce qui permettra à la maladie de se répandre dans la population, introduisant une épidémie. Par contre, si la valeur de \mathcal{R}_0 est inférieure à 1 ($\mathcal{R}_0 < 1$), un individu infecté aura (en moyenne) moins d'un descendant. Par conséquent, il n'y a pas d'épidémie. Dans la table suivante (Table1.2), nous résumons quelques estimations du nombre de reproduction.

Table 1.2 – Quelques estimations du nombre de reproduction \mathcal{R}_0

Infection	Lieu	époque	\mathcal{R}_0
Variole (Smallpox)	Pays développés avant	Les compagnes globales	3-5
Diphtérie (Diphtheria)	USA (plusieurs endroits)	1910,1947	4-6
Coqueluche	Angleterre et Pays	1942, 1950	17
(Whooping cough)	de Galles Maryland (USA)	1908, 1917	13
Rougeole	Angleterre et Pays de Galles USA	1956, 1968	13
(Measles)	(divers endroits)	1910, 1930	12-13
Paludisme(Malaria)	/	/	>100

1.3 Les réseaux complexes

Un réseau complexe est un ensemble de nœuds et de liens reliant ces derniers entre eux. Différents types de réseaux sont définis selon la nature des nœuds et des liens.

Dans les réseaux sociaux, *les nœuds*, appelés aussi *les acteurs du réseau* peuvent représenter des individus, des organisations, ou des groupes d'individus, et *les liens* représentent *les interactions ou relations sociales* entre les acteurs du réseau : lien de parenté, collaboration entre entreprises, relations sexuelles [Newman, 2002].

L'analyse des réseaux sociaux permet de décrypter les relations entre les acteurs, de comprendre la position des acteurs au sein du réseau et d'étudier la forme générale de ce réseau. Il s'agit d'une analyse de la structure où l'acteur est étudié par les liens qu'ils

tissent avec d'autres. Autrement dit, l'analyse du réseau permet d'évaluer la structure d'un réseau et d'apporter des éléments explicatifs quant au comportement de l'acteur.

1.3.1 L'épidémiologie par réseaux de contacts

L'essor de la théorie des réseaux (l'utilisation de graphes comme modèles de systèmes réels) à la fin du XX siècle a permis le développement d'une toute nouvelle approche analytique : l'épidémiologie par réseaux de contacts. L'avantage premier de cette méthode est qu'elle considère explicitement la structure sociale¹, appelée *réseau de contacts*, en la modélisant à l'aide d'un réseau complexe. En analysant la topologie de cette structure, il est possible de prévoir et de quantifier l'issue finale d'une épidémie et ainsi, en principe, de tester à l'avance différentes stratégies de prévention ou d'intervention. Nous abordons dans ce qui suit les concepts de base des réseaux à base de graphe, ensuite, les différentes classes de ce dernier :

1.3.1.1 Concepts de base

Habituellement, nous appelons réseau un ensemble d'entités (acteurs, ordinateurs, atomes, villes, etc.) que l'on appelle *sommets* ou *nœuds* reliées entre elles par un ensemble de liens appelées *arcs* si la relation est orientée et *arrêtes* dans le cas contraire. Un *réseau* est défini par un *graphe*. Dans tout ce qui suit, les graphes utilisés sont des graphes non orientés, non pondérés, et simples (les graphes ne contiennent pas de boucle). Un graphe G est un couple (V, E) où :

- (a) $V = v_1, v_2, \dots, v_n$ tel que $n = |V|$ est un ensemble de sommets (ou nœuds) ;
- (b) et $E = e_1, e_2, \dots, e_m$ tel que $m = |E|$ un ensemble de liens (ou arêtes) ;
- (c) Soient v_i et v_j deux nœuds du réseau, les nœuds v_i et v_j sont dits adjacents ou voisins s'il existe une liaison e entre le nœud v_i et le nœud v_j dans G , tel que : $e = \{(v_i, v_j) \in E\}$.

1.3.1.2 Les différentes classes de réseau

Nous commençons cette section par l'étude des classes de réseau complexe les plus célèbres afin de pouvoir identifier les aspects théoriques et méthodologiques de chacun. En effet, la plupart de ces réseaux peuvent être généralisés, nous commençons par le type de graphes le plus exploité : les graphes aléatoires.

1. Les réseaux aléatoires

La naissance de modélisation de réseaux se basant sur les réseaux aléatoires (En anglais : Random Network) était en 1959. Les deux chercheurs Erdős et Rényi publient un article [Erdős et Rényi, 1959] dans lequel ils introduisent ce nouveau concept. Ce graphe est composé de N nœuds, et on relie d'une façon aléatoire chaque couple de nœuds (v_i, v_j) par un lien avec une probabilité p . Les auteurs ont proposé plusieurs versions de leur modèle, le plus simple de ces réseaux est appelé $G_{n,p}$. En effet, chaque nœud a $(N - 1)$ voisins potentiels, et donc le degré moyen de chaque nœud pour N grand est $k = p(N - 1) \approx pN$.

La distribution des liens suit une loi de Poisson de la forme suivante :

$$P(k) = e^{-\lambda_k} \cdot \frac{\lambda_k^k}{k!} \quad (1.6)$$

1. Cette structure joue un rôle primordial dans la dynamique de la propagation des maladies infectieuses en étant le support par lequel les maladies se propagent d'un individu à un autre.

D'où k désigne le nombre de liens.

Une caractéristique intéressante du modèle d'Erdős et Rényi réside dans l'apparition de la phase de transition en fonction de k , ce qui provoque la formation d'une composante géante à partir d'une valeur critique de k . L'inconvénient majeur de ce modèle réside dans la distribution homogène du degré de ses nœuds (selon une loi de Poisson) qui n'est pas la distribution hétérogène observée dans les réseaux du réel (selon une loi de Puissance). Par conséquent, le graphe aléatoire d'Erdős et Rényi n'est pas le bon modèle pour représenter les réseaux réels. La Figure 1.4 illustre un exemple du modèle étudié.

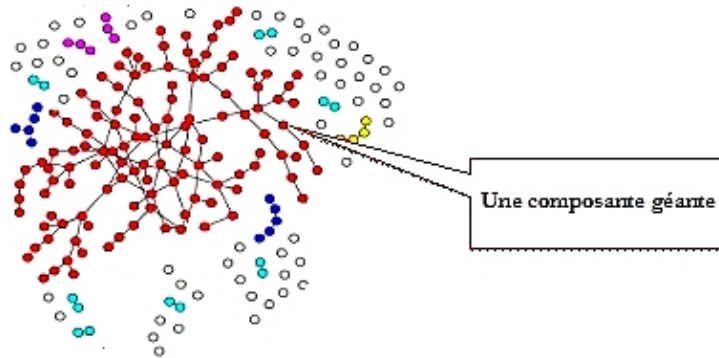


Figure 1.4 – Exemple du graphe aléatoire d'Erdős et Rényi

Afin de pallier aux limites du modèle d'Erdős et Rényi, Molloy et Reed [Molloy et Reed, 1995] proposent une généralisation de ces graphes aléatoires pour les rendre plus réaliste. Leur idée se base essentiellement sur le changement du degré de distribution des nœuds. Cette distribution de degré ne doit pas suivre forcément une loi de Poisson mais elle peut, par contre, suivre n'importe quelle loi.

Plusieurs travaux et extensions de ce modèle ont été développés en se référant aux différentes lois qui existent. Cependant, la faiblesse majeure de ce modèle réside dans le fait de ne pas résoudre le problème des forts coefficients de clustering.

2. Les réseaux réguliers

Les réseaux réguliers sont des réseaux très ordonnés et souvent utilisés dans des modèles d'automates cellulaires. Dans un réseau régulier, chaque nœud a exactement le même nombre de liens. Autrement dit, tous les nœuds ont un degré de quatre et, par conséquent, la distribution du degré présente une fonction Delta Dirac à quatre. En effet, pour affirmer une structure régulière, soit x le nombre de liens d'un nœud v_i , le nombre x doit être identique pour tous les nœuds v_i du réseau. Ainsi, $\forall v_i \in V, k_{v_i} = 2 * x$. Bien que ce type de structure s'observe en réalité très peu dans la nature, elle est en revanche souvent utilisée comme base pour la formation de réseaux plus réalistes [Stattner, 2012]

3. Les réseaux petit-monde

Le concept petit monde (En anglais : Small World) est connu après l'expérience du petit monde réalisée par le psycho-sociologue Stanley Milgram en 1967 a conduit au concept de « six degrés de séparation ».

Milgram [Milgram, 1963] crée une expérience pour vérifier l'effet petit monde par une expérience : il a écrit 160 lettres qu'il a donné au hasard à des gens d'Omaha, Nebraska (Etats Unis), il leur demandé de les faire parvenir à un courtier de New-York, en interdisant d'utiliser la poste, juste passer la lettre de main en main, à des

connaissances bien choisies pour se rapprocher du courtier. À la fin, 42 lettres sont arrivées à destination.

Ce modèle s'approche plus de la réalité. C'est un réseau au sein duquel les connexions fortes avec certains points coexistent avec des points plus éloignés. La propagation de l'information est plus rapide, tout en conservant certaines propriétés des réseaux classiques. De plus, entre toute paire de sommets, il existe un chemin très court (polylog) qui peut être trouvé facilement (navigabilité). Dans ce travail de thèse, nous sommes intéressés par ce modèle.

Plusieurs modèles existent dans la littérature utilisant l'architecture petit monde (Small World), nous présentons quelques-uns :

- Le modèle de Watts et Strogatz : Watts et Strogatz [Watts et Strogatz, 1998], ont proposé un modèle qui intègre à la fois un petit diamètre et un fort coefficient de regroupement (clustering). Dans ce modèle, l'effet petit monde de Watts et Strogatz (1998), signifie que la plupart des nœuds sont connectés par un plus court chemin à travers le réseau, et qu'il y a un effet de regroupement (clustering) signifiant qu'il y a une grande probabilité pour que deux nœuds soient connectés directement à un autre s'ils ont un nœud voisin en commun. Le réseau petit monde occupe une place intermédiaire entre le réseau régulier et le réseau aléatoire c.-à-d. il est ni totalement aléatoire ni parfaitement régulier, la probabilité p_w de recablage ou reconnexion (rewiring) joue un rôle important dans le passage d'un type de réseau vers un autre. La Figure 1.5 illustre l'effet du processus de reconnections sur le modèle petit monde :

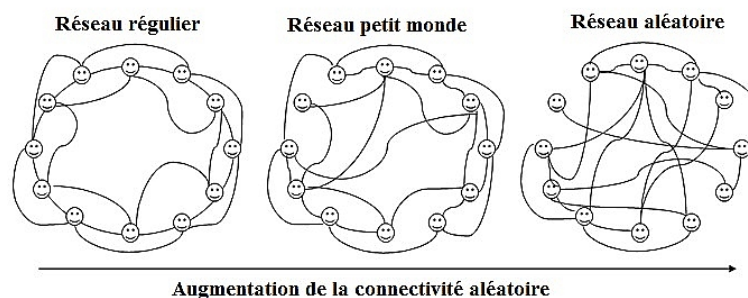


Figure 1.5 – Le modèle petit monde (Small World) proposé par Watts Strogatz

Une autre version du modèle a été proposée et analysée par Newman et Watts [Newman et Watts, 1999]. D'autres modèles se basent sur l'effet petit monde sont cités dans ce qui suit :

- Le modèle de Kleinberg : Jon Kleinberg propose dans [Kleinberg, 2000] un autre modèle de petit monde où les nœuds sont connectés selon une topologie en grille, et un faible nombre de raccourcis longues distances sont ajoutés de manière aléatoire entre certains nœuds (voir Figure 1.6), l'auteur introduit la propriété de petit monde navigable, autrement dit, la propriété dynamique de navigabilité dans un réseau petit monde.

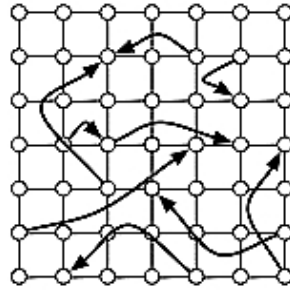


Figure 1.6 – Le réseau de Kleinberg

- Le modèle de Lebhar : Une autre variante du modèle a été développée par Lebhar [Lebhar, 2005], l'auteur a développé un algorithme de routage décentralisé sur le modèle aléatoire de petit monde de Kleinberg plus efficace et introduit ainsi le concept de processus de *petit-mondisation*. Dans [Lebhar, 2005], l'auteur a traité deux problématiques principales de l'effet petit monde navigable :
 - construire de nouveaux algorithmes décentralisés efficaces, dédiés aux grands réseaux ;
 - construire de nouvelles architectures de grands réseaux de petit diamètre, dont les chemins courts peuvent être calculés de façon décentralisée.

Ces modélisations sont intéressantes dans le cas des réseaux informatiques et deviennent moins dans le cas des réseaux réels, elles représentent des caractéristiques importantes que ne présentaient pas les modèles précédents d'Erdos et Rényi, et de Watts et Strogatz.

4. Les graphes sans échelle

En 1999, Albert et Barabasi [Barabási et Albert, 1999] ont publié un modèle dynamique permettant d'obtenir une distribution des degrés suivant une loi de puissance des nœuds entrants et sortants dans les citations d'articles scientifiques. Ils ont introduit la notion d'attachement préférentiel qui correspond à l'ajout des sommets un à un, en reliant chaque nouveau sommet préférentiellement aux sommets existants de plus hauts degrés. Albert et Barabasi sont les premiers introduisant la notion des réseaux « sans échelles » ou « invariant d'échelle ». La distribution des liens des réseaux invariants d'échelle suit une loi de puissance ou la probabilité d'avoir k liens est : $P(k) = k^{-a}$ où a est une constante.

1.3.2 L'analyse structurelle des réseaux

L'analyse des réseaux sociaux est une méthode d'analyse qui permet l'exploitation et la compréhension de la morphologie et la dynamique des réseaux sociaux. Elle permet aussi de mieux comprendre le comportement des acteurs et des communautés, elle se déroule en trois phases : la récupération des données, leur analyse et leur visualisation. Chaque phase a ses propres outils même si certains outils traitent plusieurs phases.

- **La récupération des données** : les données sont souvent recueillies par le biais de questionnaires, d'étude de terrain ou de simulation. Ces données récupérées pour l'analyse sont sous forme d'une liste de nœuds (souvent appelés acteurs ou des entités telles que des personnes, des organisations, etc.) et de liens (des relations binaires tels que les relations sociales, des dépendances ou échange). Ces nœuds et liens peuvent avoir des attributs supplémentaires qui peuvent être de différent type. Les relations constituant un réseau social peuvent être orienté, non orienté, ou mixtes.

- **L'analyse du réseau** : l'objectif de l'analyse de réseau social est d'identifier les acteurs importants, les liens cruciaux, les sous-groupes, les rôles, les caractéristiques du réseau, et ainsi de suite, pour répondre aux questions de fond sur les structures. Selon [Brandes et Wagner, 2004], il existe trois principaux niveaux d'intérêt : l'élément, le groupe et le niveau du réseau.
 - Au *niveau élément*, nous nous intéressons à des propriétés (à la fois absolues et relatives) des acteurs, des liens ou des incidences. Des exemples de ce type d'analyses sont l'identification et le classement goulot d'étranglement structurel d'éléments de réseau.
 - Au *niveau du groupe*, nous nous intéressons à classer les éléments d'un réseau et les propriétés des sous-réseaux. Les exemples sont les classes d'équivalence des acteurs et l'identification de regroupement (cluster).
 - Enfin, au *niveau du réseau*, nous nous intéressons à des propriétés de l'ensemble des réseaux tels que la connectivité ou l'équilibre.
- **Visualisation du réseau** : La représentation visuelle des réseaux sociaux est importante, elle permet de présenter les données du réseau et de visualiser le résultat de l'analyse. Cette représentation est liée à la visualisation de graphes.

Les différentes phases sont illustrées par le diagramme de la Figure 1.7 :

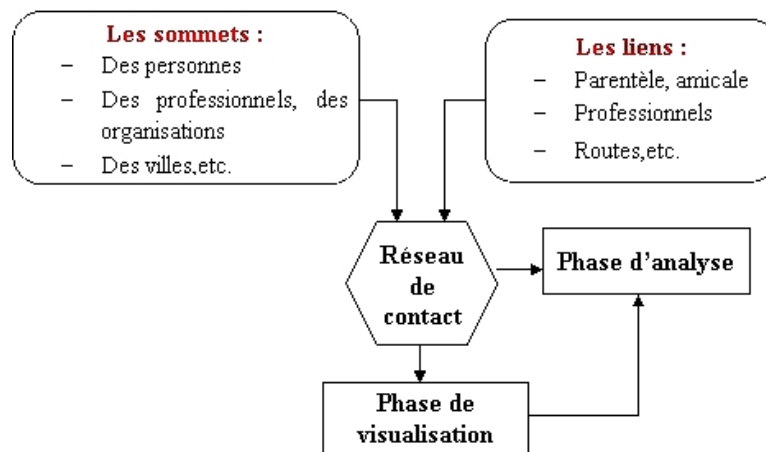


Figure 1.7 – Le diagramme des différentes étapes de l'analyse des réseaux de contact

Dans ce qui suit, nous résumons les mesures visant à caractériser les propriétés structurales de réseaux :

1.3.2.1 Les mesure des réseaux

Les mesures caractérisant les réseaux permettent de décrire la structure de ces derniers. De nombreuses mesures ont été proposées et étudiées afin de comprendre la topologie structurelle du réseau. Dans ce qui suit, nous présentons les principales mesures utilisées dans notre étude, tout en montrant son intérêt pour l'analyse des réseaux.

- **Densité** : la densité d d'un graphe G est la proportion de liens existant dans G par rapport au nombre total de liens possibles : $d(G) = 2m/n(n-1)$. Si m est de l'ordre de n on dit que le graphe est creux (par opposition aux graphes denses). En effet, cette mesure est sensible au nombre de sommets, ainsi la densité égale à 0 correspond au graphe où tous les sommets sont isolés, et égale à 1 au cas d'un

graphe complet. Dans un graphe issu d'observations empiriques, plus le nombre de sommets augmente et plus la densité tend à baisser.

- **Plus court chemin** : c'est la longueur du plus petit chemin reliant deux nœuds dans le réseau. L'un des algorithmes permettant de calculer la distance entre deux nœuds dans un graphe est : L'algorithme de Dijkstra [Dijkstra, 1959]. La distance moyenne entre deux nœuds i.e. la distance moyenne entre deux paires de nœuds permet d'évaluer le temps de transmission nécessaire entre deux individus « quelconques ».
- **Diamètre** : le diamètre d'un réseau est formellement le plus long des plus courts chemins entre deux entités, ou nœuds, du réseau, via ses connexions. Il permet de connaître le temps maximum pour transmettre la maladie.
- **Degré** : le degré $d(i)$ d'un nœud est le nombre d'arrêtes incidentes au nœud i , autrement dit, le nombre de nœuds voisins de i .
- **Distribution de degré** : peut-être calculer comme suit [Ebel *et al.*, 2003] :

$$P(k) = \frac{|\delta(v)|}{N} \quad (1.7)$$

$|\delta(v)|$: désigne le nombre de sommets du réseau G ayant le degré k et N : désigne la taille de G (nombre de nœuds). L'équation 1.7, représente la proportion des sommets de G ayant un degré k . Le degré k_i de nœud i est le nombre de liens reliés au nœud i . La distribution des degrés permet la compréhension de la répartition de la connectivité et la structure du réseau.

- **Coefficient de clustering (coefficient de regroupement)** : est la probabilité que deux voisins d'un nœud soient aussi voisins entre eux. Il peut être interprété comme la probabilité pour que deux plus proches voisins de i soient connectés les uns aux autres.

Le coefficient de clustering moyen du graphe G est la moyenne du coefficient de clustering de tous ses sommets (nœuds). Dans la littérature, il existe deux définitions du coefficient de clustering : coefficient de clustering global (aussi appelée transitive) et coefficient de clustering locale en moyenne [Watts et Strogatz, 1998] :

- Le coefficient de clustering global : est défini comme :

$$CC = \frac{3 * \text{nombre de triangles}}{\text{nombre de triplets connexes}} \quad (1.8)$$

Avec :

Un triangle est un sous-graphe complet à trois nœuds ;

Un triplet connexe est un ensemble de trois sommets avec au moins deux liens entre eux.

- Coefficient de clustering local moyen : Pour un nœud i , nous pouvons définir le coefficient local comme suit :

$$CC_i = \frac{2 | (i, j) \in E \wedge (j, i) \in E |}{k_i * (k_i - 1)}; \quad (1.9)$$

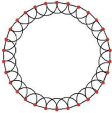



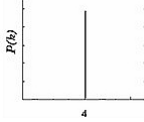
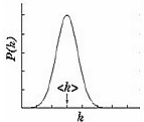
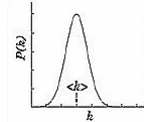
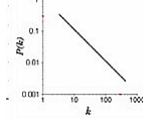
Où N_i est le voisinage du nœud i , de cardinal k_i . Cette mesure peut être vue comme la probabilité que deux nœuds se connectent au hasard indépendamment et uniformément entre eux. Plus le coefficient est grand, plus le voisinage est proche d'une

clique², ce qui est connu par la cliquicité³ (cliqueness). En considérant la moyenne de ces coefficients locaux, nous obtenons le coefficient local moyen :

$$CC = \frac{1}{|V|} \sum_{i=1}^{|V|} CC_i \quad (1.10)$$

Dans la Table 1.3, nous présentons un récapitulatif des différentes structures citées au-dessus avec certaines mesures présentées par [Börner *et al.*, 2007].

Table 1.3 – Présentation des différentes structures avec leurs mesures

Structure	Régulier	Aléatoire	Petit monde	Sans échelle
Représentation				
Nœuds	42	42	42	42
Liens	84	84	84	81
Distance moyenne	5,6	2,75	2,9	2,6
Coefficient de clustering	0,6	0,13	0,16	0,22
Degré moyen	4	4	4	3,86
Distribution des degrés				

La structure utilisée pour le réseau petit-monde présente une courbe de distribution des degrés en forme de cloche. Cependant, la propriété petit-monde ne détermine par une forme particulière pour la courbe de distribution. En effet, il est important de préciser que les propriétés petit-monde et scale-free ne sont pas exclusives. Certains réseaux du monde réel peuvent donc être à la fois petit-monde et présenter une distribution des degrés qui suit une loi de puissance [Stattner, 2012].

1.4 Conclusion

L'application de concept réseau social dans les modèles compartiment d'épidémie est actuellement la plus courante pour maîtriser comment une propagation d'épidémie prend place et évolue dans une population. Bien que de nombreux travaux aient été menés pour étudier et comprendre ces phénomènes, nous observons que la plupart font l'hypothèse de structures statiques, ce qui ne reflète pas la réalité des réseaux du monde réel ayant un caractère dynamique [Stattner, 2012].

Dans ce chapitre, nous avons présenté l'un des principaux aspects de recherche constituant l'un des contextes des travaux présentés dans cette thèse. Nous avons abordé la

2. Une clique d'un graphe est un sous ensemble de sommets tous en relation deux à deux.

3. C'est une mesure de la création de cliques dans un réseau. Elle caractérise le fait que des nœuds adjacents à un nœud donné dans un graphe sont adjacents entre eux.

notion de modélisation mathématique en épidémiologie avec une présentation des modèles les plus célèbres de diffusion de la maladie. Ensuite, nous avons expliqué le concept de réseau social ainsi que leurs caractéristiques. Ces deux aspects représentent l'objectif du modèle proposé dans le chapitre .4. En effet, dans cette thèse, nous nous intéressons à mettre en œuvre des outils théoriques adaptées à l'étude de la propagation d'épidémie au sein d'un réseau social d'individu.

Le Chapitre suivant est dédié à un état de l'art sur les différents modèles d'épidémie.

Chapitre 2

Etat de l'art sur les modèles d'épidémie

Sommaire

2.1	Introduction	25
2.2	Principe de la modélisation mathématique en épidémiologie	25
2.3	Les modèles mathématiques d'épidémie simple	26
2.4	Les modèles compartimentaux déterministes	27
2.4.1	Les modèles compartimentaux déterministes simples	27
2.4.2	Les modèles compartimentaux déterministes spatiaux	28
2.5	Modèles compartimentaux stochastiques	30
2.6	Les modèles mathématiques d'épidémie complexes	31
2.6.1	Les modèles à base d'agent / individus-centrés	31
2.6.2	Les modèles de réseau	32
2.6.3	Les modèles de métapopulation	34
2.6.4	Modèles à base des structures de contact	35
2.7	Conclusion	36

2.1 Introduction

En épidémiologie, les modèles mathématiques sont très fréquemment utilisés. Ces modèles répondent à des questions de plus en plus complexes qui peuvent avoir des conséquences importantes sur notre société. Dans ce chapitre, nous nous intéressons aux travaux menés sur les modèles mathématiques pour la compréhension des phénomènes du monde réel. Nous explorons les différentes classes de modèles associées aux mécanismes de propagation d'épidémie. En effet, nous distinguons quatre grands groupes de familles d'approches dans ce domaine : les modèles mathématiques d'épidémie simple, les modèles déterministes, les modèles stochastiques et les modèles complexes.

2.2 Principe de la modélisation mathématique en épidémiologie

Les modèles mathématiques supposent une connaissance à priori des lois déterminant le phénomène étudié. Le principe de ces modèles réside tout d'abord, dans la transcription des différents liens existant entre la transmission de la maladie et les autres facteurs

(les facteurs de risque, l'environnement, etc.) sous forme d'un modèle (équations mathématiques, modèle d'individu, etc.). Ensuite, des données expérimentales ou des observations de terrain peuvent être utilisées pour estimer certains paramètres du modèle ou pour tester la validité des lois établies. Enfin, ces modèles peuvent servir à quantifier l'effet de la variation de certains facteurs ou de mesures de contrôle sur le niveau de transmission. Dans la plupart de ces modèles, la population est répartie en plusieurs catégories appelées **compartiment** en fonction de leur état de santé (susceptible à se faire infecter, infecté non-contagieux, infecté contagieux, immunisé, décédé, etc.).

Dans ce travail de thèse, nous nous intéressons aux modèles compartimentaux car ils sont très largement utilisés. Aujourd'hui, deux formulations des modèles compartimentaux sont couramment utilisées :

- 1) **déterministe**, basée sur un système d'équations différentielles ordinaires ou partielles ;
- 2) **stochastique**, basée sur des équations différentielles stochastiques ou sur une modélisation markovienne par processus de saut.

Dans ce qui suit, nous détaillerons les différents groupes d'approches qui existent dans la littérature des modèles mathématiques. Nous présentons les travaux les plus pertinents et qui répondent au mieux à la problématique étudiée. La liste des références citées en dessous n'est certainement pas exhaustive et les applications concrètes des modèles mathématiques restent toujours possibles.

2.3 Les modèles mathématiques d'épidémie simple

Le premier modèle mathématique de l'histoire de l'épidémiologie a été publié par Daniel Bernoulli en 1760 [Bernoulli, 1760]. Bernoulli énonça un modèle pour réduire le risque de décès par la variole. En 1906, Hamer formula et analysa un modèle de temps discret dans une tentative de comprendre la récurrence de l'épidémie de la rougeole [Hamer, 1906]. Son modèle est considéré comme étant le premier à supposer que l'incidence (nombre de nouveaux cas par unité de temps) dépend de la densité des sujets susceptibles et infectieux.

En 1927, l'une des généralisations importantes a été réalisée pour les modèles épidémiques par Kermack et McKendrick [Kermack et McKendrick, 1927]. Les deux chercheurs ont développé le modèle SIR (Susceptible- Infected- Removed) dans un travail célèbre qui a fourni la base de la plupart des recherches dans le domaine de la modélisation des épidémies et qui a donné naissance à plusieurs modèles en épidémiologie.

D'une manière générale, la simulation de ces modèles est basée sur une résolution numérique d'équations différentielles ordinaires (EDO) avec prise ou pas en considération de la phase de latence de la maladie :

Pas de phase de latente :

$$\begin{cases} dS/dt = S'(t) = -\alpha.I(t).S(t) \\ dI/dt = I'(t) = \alpha.I(t).S(t) - \gamma.I(t) \\ dR/dt = R'(t) = \gamma.I(t) \end{cases} \quad (2.1)$$

Avec une phase de latente :

$$\begin{cases} dS/dt = S'(t) = -\alpha.I(t).S(t) \\ dE/dt = I'(t) = \alpha.I(t).S(t) - \beta.E(t) \\ dI/dt = I'(t) = \beta.E(t) - \gamma.I(t) \\ dR/dt = R'(t) = \gamma.I(t) \end{cases} \quad (2.2)$$

Ces modèles décrivent la dynamique des susceptibles, $S(t)$, qui deviennent infectés avec un taux α . Des exposés $E(t)$ (avec une phase de latente), des infectés $I(t)$ et des retirés $R(t)$ dans le temps. Dans la Figure 2.1 (2.1a, 2.1b), nous illustrons l'évolution des variables d'états : $S(t), E(t), I(t)$ et $R(t)$ dans le temps pour les deux modèles.

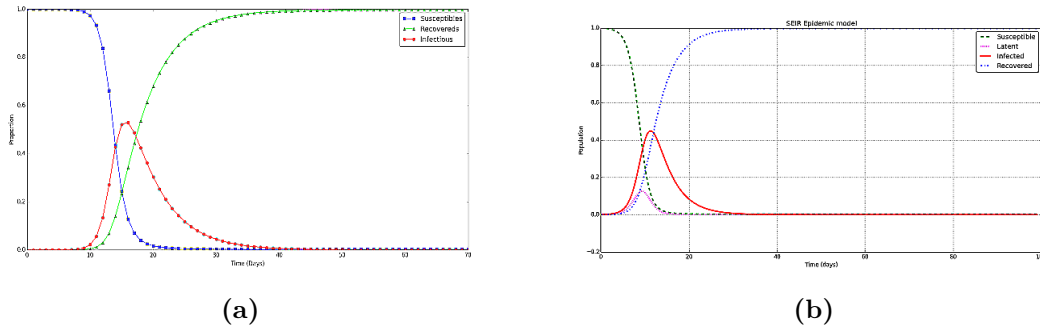


Figure 2.1 — L'évolution des variables d'états dans le temps pendant une épidémie : (a) : le modèle SIR (Pas de phase de latente), (b) : le modèle SEIR (avec le temps de latente)

2.4 Les modèles compartimentaux déterministes

Les modèles déterministes [Anderson et May, 1979], [Hethcote, 2000] permettent d'appréhender, à travers leur ajustement aux données disponibles, les facteurs d'hétérogénéité responsables de la propagation observée dans les populations hôtes. Ceci permet de définir les priorités qui doivent prévaloir dans la récolte des données afin d'affiner le paramétrage des modèles.

De plus, le choix des compartiments à inclure dans le modèle dépend des caractéristiques particulières de la maladie en cours de la modélisation et de l'objectif du modèle. Nous distinguons deux groupes de modèles déterministes : des modèles simples et des modèles spatiaux.

2.4.1 Les modèles compartimentaux déterministes simples

Depuis le milieu du XXe siècle, les modèles mathématiques d'épidémiologie ont connu une croissance exponentielle. Une grande variété de modèles mathématiques a été formulée, analysée et appliquée aux maladies infectieuses. Des revues de littérature [Becker, 1979] [Castillo-Chavez, 1989] ont abordé les différents travaux de recherche et la croissance rapide de la modélisation en épidémiologie.

Les premiers modèles mathématiques étaient souvent déterministe et les problématiques qui ont été abordées sont par exemple :

- Quelle est la possibilité qu'une épidémie infecte une grande partie de la population ? ;
- Combien de personnes seront infectées si l'épidémie s'étend ? ;
- Quel est l'impact de la vaccination pour une population donnée avant le début de la maladie ?, Quel est le seuil épidémique ?

Le modèle SIR est la base de la plupart des modèles utilisés pour l'étude de la diffusion d'épidémie, comme par exemple, dans le cas de la grippe « A » en 2009. Afin d'améliorer la performance de ce modèle, des travaux ont rajouté l'étude des facteurs liés à la maladie, nous citons par exemple :

Un modèle avec une perte d'immunité contre la maladie, tel que le modèle SIRS (Susceptible-Infective-Recovered- Susceptible) [Wen et Yang, 2008] [Hooten *et al.*, 2010], d'autres avec aucune protection immunitaire contre la maladie, tel que le modèle SIS (Susceptible-Infective- Susceptible) [Garba *et al.*, 2013], etc.

Travaux connexes

Dans ce contexte, les auteurs dans [Casagrandi *et al.*, 2006] ont développé un modèle simple d'EDO pour étudier les conséquences épidémiologiques du mécanisme des virus de la grippe A.

Par l'amélioration du modèle classique SIR, ils ont introduit une quatrième classe (C) pour les individus ayant une immunité croisée . Le modèle de SIRC affirme que la prévalence du virus est maximale pour une valeur intermédiaire de \mathcal{R}_0 . Via une analyse de bifurcation du modèle, les auteurs ont discuté l'effet de la saisonnalité sur les régimes épidémiologiques.

utilisent les modèles d'épidémie SIR et SEIR avec un taux d'incidence non linéaire, en intégrant l'aspect du retard dans les modèles des EDO. En particulier, ils considèrent deux modèles d'EDO avec retard dans lesquels les retards sont causés : (1) par la latence de l'infection dans un vecteur, et (2) par la période de latence dans un hôte infecté. En construisant la fonction de Lyapunov appropriée et utilisant le principe d'invariance de Lyapunov-LaSalle, ils ont prouvé la stabilité globale de l'équilibre endémique et l'équilibre en absence de maladie en cas de retard dans le temps pour chaque modèle. Les résultats ont montré que les propriétés globales des équilibres ne dépendent que du nombre de reproduction de base (\mathcal{R}_0).

L'un des problèmes majeurs des modèles épidémiques simples est qu'ils conduisent à des schémas épidémiques irréalistes. Ils considèrent que tous les individus de la population peuvent être infectés et qu'aucun afflux de nouveaux sujets susceptibles par le biais de la naissance ou de la migration. En effet, des modèles avec extensions rajoutées sont mis en place par plusieurs chercheurs afin d'étudier l'impact de certains facteurs, tels que : médicaux, sociaux et démographiques sur la propagation d'épidémie [Capasso, 1993].

Des versions améliorées du travail de [Huang *et al.*, 2010], ont exploité l'effet de la structure par sexe et contact [Sharomi et Gumel, 2011], la structure d'âge où la sensibilité à la maladie dépend de l'âge d'individu [Melnik et Korobeinikov, 2013], et la démographie (naissance et le décès) [Wang *et al.*, 2014].

Discussion

L'utilisation des modèles mathématiques d'épidémie simples ne se limitent pas aux modèles cités en dessus mais ils incorporent des approches spatiales ou spatiotemporelles. Par exemple, il est possible d'utiliser des approches déterministes dans un modèle basé sur l'individu mobile [Reluga *et al.*, 2006].

2.4.2 Les modèles compartimentaux déterministes spatiaux

Une variété de modèles mathématiques compartimentaux déterministes existent dans la littérature, chaque modèle utilise ses propre paramètres dans une situation bien particulière. En effet, des équipes de recherche ont utilisé les modèles compartimentaux déterministes simples, d'autres équipes ont rajouté a ces modèles standards des **structures spatiales** pour :

- Reconstituer rétrospectivement les dynamiques de la pandémie de Hong Kong en 1968 à travers un réseau mondial de 52 villes [Rvachev et Longini, 1985] [Longini, 1988] ;
- Modéliser l'épidémie de grippe française de 1984/1985 [Flahault *et al.*, 1988] ;
- Simuler la diffusion de la grippe dans 9 villes européennes en prenant en compte les transports aériens [Flahault *et al.*, 1994]

Ces modèles spatialisés restent rares car ils exigent d'avoir une connaissance très fine des phénomènes étudiés [Reluga *et al.*, 2006].

Travaux connexes

L'utilisation des EDO de temps discrètes dans un espace d'état continu pour étudier la propagation mondiale de la grippe a été utilisée par Rvachev et Longini [Rvachev et Longini, 1985]. Les auteurs ont démontré par la modélisation rétrospective que la pandémie de 1968-1969 est diffusée à travers un réseau de villes mondiales interconnectées par des lignes de voyages aériens. Les paramètres du modèle ont été estimés à partir des données d'incidence de Hong Kong (la première ville avait atteint la nouvelle souche de la grippe (A2 / Hong Kong / 1968)). Les rapports de morbidité et de mortalité obtenus de l'Organisation Mondiale de la Santé (OMS) ont été utilisés pour valider le modèle.

Les données du transport comprenaient le nombre journalier en moyenne annuelle des passagers aériens voyageant entre 52 villes à travers le monde. Le modèle Rvachev et Longini a ensuite été appliqué pour prévoir la propagation d'épidémie dans des pays et par d'autres modèles de transport. Une version améliorée du modèle de Rvachev et Longini a été proposé par Grais dans [Grais *et al.*, 2003]. Dans ce travail, les auteurs examinent les dynamiques spatiales et temporelles de la propagation de l'épidémie entre 1968- 1969 pour la pandémie de l'an 2000. Les résultats de leurs simulations peuvent donner un aperçu sur la façon de la propagation d'une prochaine souche pandémique de la grippe dans le monde.

Dans [Gaudart, 2007], l'auteur a étudié la distribution spatiotemporelle de la maladie du paludisme au Mali pour l'élaboration de la carte à risque épidémique. Il a, également, proposé une approche statistique et une modélisation déterministe. Plusieurs outils d'investigations sont utilisés pour cette étude : l'analyse spatiale et spatiotemporelle pour décrire l'hétérogénéité spatiale, le développement d'une méthode par arbre de régression oblique (SpODT) découpant la région en zones de risques différents, la recherche de clusters spatiotemporels et la modélisation déterministe pour le modèle adapté à la région d'étude tenant compte de la pluviométrie simulée par 4 méthodes (distribution empirique, chaîne de Markov cachées, non linéaire, non paramétrique).

Discussion

A travers ces travaux, nous constatons que le principal avantage de modèles déterministes réside dans leur analyse simple. Ils peuvent être plus complexes dans certains cas, mais toujours possible de les analyser. Par contre, pour gérer mathématiquement un modèle épidémique stochastique, il devrait être simple d'où ce qui n'est pas tout à fait réaliste.

2.5 Modèles compartimentaux stochastiques

Les modèles stochastiques sont beaucoup plus conformes. Ils peuvent intégrer l'ensemble des sources d'hétérogénéité souhaitées (hétérogénéité spatiale, temporelle, comportementale ou autre). Ils tiennent compte de la nature aléatoire de tout événement de transmission et de développement d'une infection [Sauvage et Pontier, 2005]. Contrairement aux modèles déterministes qui visent principalement à étudier une taille de population importante, l'utilisation des modèles stochastiques est préférable lorsqu'on étudie une petite population. Dans ce qui suit, nous présentons certains modèles stochastiques :

Travaux connexes

Dans [Legrand *et al.*, 2004], les auteurs proposent un modèle continu pour le suivi d'épidémie de variole dans une grande ville (2 millions d'habitants) afin d'évaluer les paramètres clés pour le contrôle de cette épidémie. Ils ont construit un modèle stochastique pour simuler l'évolution d'une épidémie contrôlée par la vaccination et l'isolement de cas. Une analyse de sensibilité a été établie.

Trois paramètres clés ont un impact sur l'évolution épidémique :

- le nombre de reproduction de base \mathcal{R}_0 ;
- le temps d'intervention ;
- la proportion de contacts vaccinés et en relation directe avec le vecteur.

Wu et al. [Wu *et al.*, 2006] ont également utilisé un modèle stochastique, centré sur l'individu pour étudier à Hong Kong, la réduction attendue du taux d'attaque lors de la première vague pandémique. A l'échelle mondiale, Gattepaille et Jay [Gattepaille et Jay, 2007] ont proposé un modèle de propagation d'épidémie proche de la réalité. Ce modèle est discret, à la fois dans le temps et dans l'espace, et se base sur un automate cellulaire, réalisant un maillage du planisphère.

Plusieurs modèles ont également examiné l'impact de la structure par âge d'une population, à titre d'exemples : [Hethcote, 1988], [Cauchemez *et al.*, 2008], [Flasche *et al.*, 2011], car les humains ont tendance à interagir avec des personnes du même âge beaucoup plus qu'interagir avec ceux des âges différents.

Discussion

La plupart des recherches citées qu'elles soient basées sur les modèles déterministes ou stochastiques, elles sont fondées sur des modèles épidémiques simples permettant d'analyser et d'étudier l'évolution de l'épidémie dans le temps, ou d'identifier les caractéristiques générales de l'épidémie.

Néanmoins, la structure topologique complexe du réseau de contacts n'est pas prise en considération. De plus, la majorité de ces travaux se basent essentiellement sur une modélisation mathématique par équations. Ce formalisme se heurte, toutefois, à plusieurs types de problèmes [Pavé, 1994] :

- Le calcul effectué est un calcul du comportement moyen du système, le cas où le nombre de certaines molécules est faible est mal pris en compte par cette modélisation ;
- Les réseaux de taille réelle – de l'ordre de la centaine de réactions – sont difficiles à modéliser par un jeu d'équations différentielles ;

- L'introduction de nouvelles populations et l'amélioration du modèle nécessitent la modification de la plupart des équations du modèle ;
- La modélisation par équation différentielle nécessite un haut niveau d'abstraction.

2.6 Les modèles mathématiques d'épidémie complexes

A l'heure actuelle, de nombreux chercheurs et spécialistes en épidémiologie mettent en place des approches et des outils permettant de simuler les modèles de propagation des maladies infectieuses dans le but de réduire la diffusion de celles-ci. Ces approches et outils sont, de plus en plus, améliorés afin de satisfaire les décideurs de la santé publique. Dernièrement, des modèles mathématiques d'épidémie complexes sont apparus, nous les classons dans quatre grandes catégories : *les modèles à base d'agent / individus-centrés*, *les modèles de réseau*, *les modèles de métapopulation*, *les structures de contact*.

2.6.1 Les modèles à base d'agent / individus-centrés

Les modèles de simulation individus-centrés (Automates Cellulaires et Systèmes Multi-Agents) sont considérés comme des moyens efficaces pour représenter une population hétérogène dans son environnement. Le processus de transmission des maladies infectieuses dans ces modèles est géré par le comportement de chaque individu dans le système étudié. De plus, ils permettent de décrire avec plus de précision les phénomènes considérés, sans être limités par une croissance exponentielle du nombre de compartiments et d'équations [Opatowski *et al.*, 2013].

Travaux connexes

Dans [Fuentes et Kuperman, 1999], les auteurs ont focalisé leur travail sur les maladies transmissibles par contact de personne à personne par l'utilisation d'un automate cellulaire. Ils ont analysé la propagation d'un hôte infectieux par les deux modèles d'épidémies SIS et SIR. Cependant, la matrice d'individus qui modélise les contacts est figée et ne décrit pas un réseau social réel. Eubank et al. proposent le système EpiSims (The Epidemiological Simulation System) pour modéliser la propagation de la maladie dans les zones urbaines [Eubank *et al.*, 2004]. Ce système est fondé sur la prise en compte de comportements individuels, les interactions entre individus. La richesse de ce travail est qu'il tire parti d'une autre application, TransSims [Barrett *et al.*, 2001] dont l'objectif est de simuler les déplacements des individus dans la ville. Fergusson et al. [Fergusson *et al.*, 2005] d'une part et Longini et al. [Longini *et al.*, 2005] d'autre part ont étudié la possibilité de stopper un foyer naissant de pandémie en prenant la Thaïlande comme pays d'étude. Les modèles utilisés étaient de type stochastique, basés sur l'approche individus-centrés, permettant ainsi de prendre en compte les contacts familiaux, sociaux et professionnels des individus de la population.

Dans [Basileu *et al.*, 2010], les auteurs proposent un modèle de diffusion spatiale à base d'agents hybrides basé sur le modèle SEIR. Le modèle permet de déterminer, en chaque zone, le nombre d'agents touchés par la maladie au jour le jour. Via ce modèle, le décideur peut ainsi connaître, en chaque unité de son territoire, l'ampleur de la pandémie. La transition d'une phase à l'autre durant une infection est modélisée de façon markovienne.

Récemment, les auteurs dans [Langlois *et al.*, 2013] ont développé la plateforme MA-Géo (Modélisation Agent Géographique) dans le but de la simulation multi-agents dédiée aux problématiques spatialisées et multi-niveaux, suffisamment visuelle et intuitive. Le

diagramme conceptuel intègre les Agents, les Organisations, les Comportements et les relations du méta-modèle AOC (Agents-Organisations-Comportements), qui permet de créer facilement des modèles complexes et multi-niveaux. Dans ce travail, les auteurs présentent la plateforme MAGéo à travers l'exemple du modèle SIR spatialisé.

Discussion

Les modèles à base d'agents / individus-centrés s'avèrent très utiles pour la modélisation de la transmission des maladies transmissibles. Cependant, leur amélioration s'accompagne d'une augmentation conséquente du nombre de variables et donc de la complexité des modèles. De tels modèles n'ont d'intérêt que dans le cadre d'une bonne connaissance des variables utilisées. Ainsi, la complexification des modèles ne peut se faire qu'en parallèle d'un recueil de données conséquent, dans le cadre d'une collaboration entre modélisateurs, médecins et épidémiologistes [Opatowski *et al.*, 2013]

La Table 2.1 récapitule les avantages et les inconvénients de la modélisation par équations et la modélisation par agent proposé dans [Pavé, 1994] :

Table 2.1 – Comparaison entre la modélisation par équations et par agent

Modélisation par équations	Modélisation par agent
Synthétique	De nombreux paramètres
Résolution mathématique	Important besoin de calcul
Formalisée	Non formalisée
Loin de la biologie, abstraite	Proche de la biologie
Peu modulaire et peu incrémental	Modulaire et incrémentale
Description au niveau des populations	Description au niveau de l'entité (cellule, molécule, ...)

2.6.2 Les modèles de réseau

La théorie des réseaux est présente en mathématiques depuis Euler 1736, elle a été utilisée pour résoudre le problème des 7 ponts de Königsberg [Euler, 1741]. Pour résoudre ce problème, L. Euler schématisa la situation sous forme d'un graphe dans lequel les terres accessibles étaient représentées par un nœud, et les ponts par des connexions entre ces nœuds.

En effet, un réseau complexe est un ensemble de nœuds et de liens reliant ces derniers entre eux. Différents types de réseaux sont définis selon la nature des nœuds et des liens [Diestel, 2005].

Dans le contexte d'épidémiologie, l'un des premiers modèles est basé sur un formalisme issu de la théorie des réseaux a été proposé par M.E.J. Newman en 2002 [Newman, 2002]. Ce modèle s'est depuis imposé comme la référence dans le domaine et est le fondement de la plupart des nouveaux développements [Marder, 2007] [Meyers *et al.*, 2006] [Noël *et al.*, 2009].

Une variété de travaux ont suivi ce travail, afin de comprendre le phénomène de la propagation d'une maladie à travers ces réseaux : les réseaux réguliers « Regular Lattice », aléatoires « Random Networks », petit monde « Small-World » et Sans échelle « Scale-Free » illustrée par la Table 2.2.

Table 2.2 – La classification des différents types de réseaux

Type de réseau	Propriétés	Références
Régulier	<ul style="list-style-type: none"> — Connectivité de voisinage (treillis généralement réguliers) ; — Fort clustering ; — Trop de courts chemins. 	[Harris, 1974], [Anderson <i>et al.</i> , 1992], [Filipe et Maule, 2004]
Petit monde	<ul style="list-style-type: none"> — Réseau régulier reconnecté avec des raccourcis (les plus courts circuits) ; — Fort clustering ; — Peu de courts chemins. 	[Watts et Strogatz, 1998], [Zekri et Clerc, 2002], [Kleinberg, 2000], [Vazquez, 2006]
Aléatoire	<ul style="list-style-type: none"> — Nœuds connectés avec une probabilité P ; — Faible clustering ; — Trop de courts chemins. 	[Erdős et Rényi, 1959], [Sander <i>et al.</i> , 2002], [Volz, 2008], [Miller, 2011]
Sans échelle	<ul style="list-style-type: none"> — La distribution des degrés suit une loi de puissance ; — Le nombre de nœuds de degré k est proportionnel à k^α. 	[Barabási et Albert, 1999], [Pastor-Satorras et Vespignani, 2001], [Barthélemy <i>et al.</i> , 2005]

Dans ce qui suit nous présentons quelques travaux de la modélisation mathématique couplée aux modèles réseaux sociaux d'épidémie :

Travaux connexes

Saramäki and Kaski [Saramäki et Kaski, 2005] ont proposé un modèle mathématique de type SIR pour modéliser le processus aléatoire de propagation de la grippe basé sur un réseau dynamique petit monde. Par ailleurs, dans leur modèle l'état exposé qui caractérise cette maladie n'était pas prise en considération ce qui rend leur modèle irréaliste. Dans [Carrat *et al.*, 2006], les auteurs proposent un modèle qui comporte deux niveaux : le niveau individuel et le niveau de communauté, dans lequel les auteurs réalisent un réseau petit monde et modélisent les rencontres entre les individus par les graphes aléatoires Barabási-Albert.

Afin de prédire la propagation d'une souche pandémique du virus de la grippe en Italie et l'impact des mesures de contrôle, les auteurs dans [Rizzo *et al.*, 2008], ont développé un modèle déterministe SEIR avec une composante de simulation stochastique. Les résultats de leurs simulations ont montré que des mesures telles que l'augmentation de la distance sociale, la fermeture des écoles, pourrait être utile pour retarder le pic épidémique et offrant ainsi plus de temps pour les vaccins qui seront produits d'où l'aspect spatial était absent dans ce travail.

De leur part, [Balcan *et al.*, 2010] ont focalisé leurs étude sur les déplacements des individus par voies aériens. Ils modélisent à travers un réseau leurs déplacements d'où les nœuds représentent les aéroports et les arêtes sont les vols. Les résultats de leurs études épidémiologiques ont été comparés à des données de propagation d'une souche de grippe saisonnière. Dans [Liccardo et Fierro, 2013], les auteurs ont construit un modèle SIR stochastique pour étudier la diffusion de la grippe dans un réseau de D -dimension représentant le contact entre les individus. Ce modèle a été validé par une comparaison avec les données épidémiologiques de différentes classes d'âges d'Italie pour la grippe A (H1N1). Récemment, [González-Parra *et al.*, 2015] proposent l'utilisation d'un modèle de réseau aléatoire basé sur le modèle SEIR pour simuler et comprendre le processus de propagation d'épidémie de la grippe A (H1N1) dans une région à Venezuela. Leurs résultats confirment que le modèle proposé est plus performant que le modèle d'épidémie traditionnel basé sur les EDO.

Dans les travaux cités au-dessus, l'aspect multidimensionnel est évidemment absent.

2.6.3 Les modèles de métapopulation

Le concept de métapopulation a été introduit en 1968 par Richard Levins [Levins, 1969], pour décrire un modèle de dynamique de population d'insectes nuisibles à l'agriculture (domaine d'écologie). Son idée a été reprise et développée pour le domaine épidémiologique.

Dans le monde réel, les populations humaines ne sont pas réparties uniformément dans l'espace, elles sont souvent agrégées en sous-populations plus ou moins interconnectées. En épidémiologie, on parle de métapopulation pour désigner un ensemble de population de sous-populations. En effet, un modèle de métapopulation se compose d'un ensemble de sous-populations d'une même espèce ayant chacun sa propre dynamique, et étant connectée à d'autres sous-populations par des interactions.

Dans cette approche, la propagation de la maladie se produit à travers des processus de mobilité ou de la migration d'individus entre les sous-populations basée sur un formalisme EDO. Ces caractéristiques suggèrent que la modélisation de la métapopulation devrait fournir une approche appropriée pour la modélisation de la propagation d'épidémie au niveau mondial ou régional par l'intermédiaire, par exemple de : déplacement, transport aérien, etc. Dans la section suivante, nous détaillons quelques travaux liés à ce concept.

Travaux connexes

Colizza *et al.* dans [Colizza *et al.*, 2007] ont étudié un modèle épidémiologique stochastique de métapopulation de la grippe, à partir des données obtenues des compagnies aériennes sur les flux entre les zones urbaines. Ce travail a introduit une nouvelle piste de recherche pour un autre travail amélioré cité dans [Colizza et Vespignani, 2008]. Dans cette étude, les auteurs ont présenté une analyse approfondie du comportement des maladies infectieuses dans les modèles de métapopulation caractérisés par des modèles de connectivité et de mobilité hétérogènes. Ils ont dérivé les équations déterministes de réaction-diffusion décrivant l'évolution de l'épidémie dans les systèmes de métapopulation. L'hétérogénéité du réseau est explicitement prise en compte par les sous-populations de partitionnement en fonction des variables de degrés.

La modélisation des maladies infectieuses par l'étude de scénarios réalistes qui intègrent l'hétérogénéité de la population, les structures sociales, et les processus de mobilité au niveau individuel est introduite par les auteurs dans [Meloni *et al.*, 2011]. Les auteurs ont formulé et analysé un modèle de métapopulation qui intègre plusieurs scénarios de chan-

gements de comportement dans des schémas de mobilité des individus. Ils ont considéré qu'une population d'individus est divisée en sous-populations connectées les unes aux autres suivant un réseau hétérogène.

Au sein de chaque sous-population, les individus sont classés en fonction de leur état de santé dynamique en tant que : Susceptible (S), Infecté (I) et Retiré (R). En absence de changements de comportement (flèches bleues présenté par la Figure 2.2), les individus se déplacent d'une sous-population à une autre à un taux λ de suivant le plus court chemin reliant les deux sous-populations. Les flèches discontinues représentent le deuxième mécanisme de la réaction comportementale dans lequel les gens se déplacent en évitant les endroits à forte prévalence.

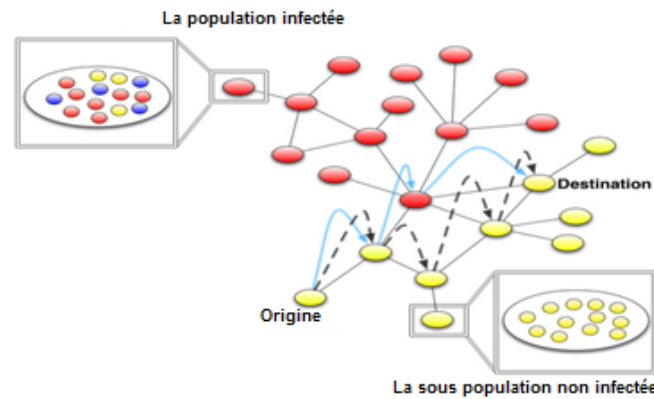


Figure 2.2 – Représentation schématique du modèle de métapopulation [Meloni *et al.*, 2011]

Récemment, Njagarah et Nyabadza dans [Njagarah et Nyabadza, 2014] établissent un cadre de couplage entre un modèle mathématique déterministe SIR basé sur les EDO et un modèle de métapopulation pour décrire la propagation d'épidémie entre deux communautés reliées par des mouvements migratoires (réseau de transport). Les auteurs ont calculé le seuil d'épidémie en considérant les cas suivants : **les communautés sont isolées et les communautés sont reliées**. Leurs résultats ont montré l'intérêt de leur système dans le cas de la maladie du choléra, ils ont démontré qu'il existe une probabilité de la maladie dévastatrice de la communauté relativement aux meilleures installations.

Discussion : des structures de métapopulation plus complexes (si l'on considère stochastiques, des modèles spatiaux et / ou saisonnière) présentent une variété de phénomènes, relativement connus et compris isolément, mais qui génèrent une interaction complexe de forces antagonistes [Grenfell et Harwood, 1997] [Keeling *et al.*, 2004]. Le comportement global du système dépend donc, fortement, de leurs forces relatives, et des valeurs des paramètres et les conditions initiales du modèle.

2.6.4 Modèles à base des structures de contact

Parmi les facteurs qui jouent un rôle crucial dans la transmission des maladies infectieuses au sein d'une population nous citons , le type et la fréquence des contacts entre les individus infectieux et susceptibles en fonction de la contagiosité et du mode de transmission de/des agent (s) causal. En outre, le modèle de mélange des hôtes a un effet important sur la manière de transmission de la maladie. La modélisation des caractéristiques de transmission est donc fortement influencée par les hypothèses autour de l'homogénéité ou de l'hétérogénéité du mélange.

Le mélange homogène d'individus suppose que le contact entre les différents individus se produise au hasard avec une probabilité égale (par exemple, chaque individu est

également susceptible de prendre contact avec un autre individu de même âge ou d'âge différent). Le mélange hétérogène suppose le mélange non-aléatoire en cas où certains individus ou groupes d'individus sont les plus susceptibles d'être en contact avec des personnes infectées que d'autres [Brauer, 2008].

Travaux connexes

Le couplage du concept de l'hétérogénéité avec la structure de contact dans les modèles épidémiologiques permet également une meilleure précision dans les prédictions avec des dynamiques observées à des échelles relativement petites, comme le montre une analyse d'une épidémie de dengue à l'est d'Island [Favier *et al.*, 2005]. Les auteurs ont proposé un modèle d'individu SEIR, tout en prenant en considération dans leur formulation de problème sous EDO le phénomène de naissance et de décès. Dans ce modèle, comme dans leur simulations, le contact entre les personnes est modélisé par deux niveaux : (i) un niveau local (contacts au sein d'un ménage), et (ii) au niveau global (possibilité de contacts entre les personnes appartenant à différents ménages). Les auteurs ont conclu que le nombre de contacts entre l'hôte et le vecteur est très dépendant de facteurs socio-économiques et culturels, tels que l'hygiène et le contrôle des vecteurs. Récemment, les auteurs dans [Apolloni *et al.*, 2014] ont décrit une population structurée spatialement avec un mélange hétérogène de population avec un comportement de voyage par un modèle stochastique d'épidémie de métapopulation multi-hôte. Différentes partitions de la population, des modèles de mélange et des structures de mobilité sont considérées. En effet, deux couches sont à exploiter : spatiale et sociale, illustrées par la Figure 2.3.

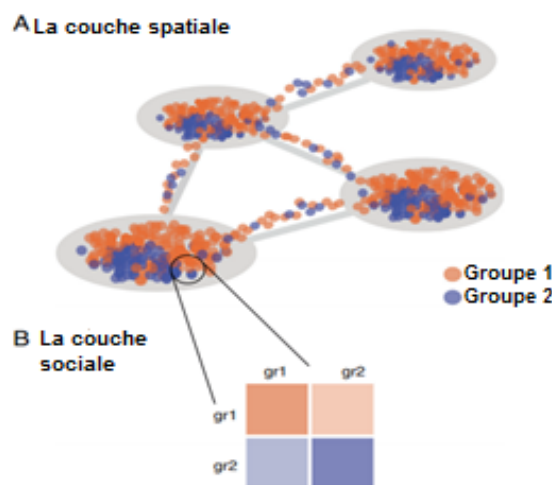


Figure 2.3 – Le schéma du modèle. (A) La couche spatiale, basée sur l'approche de la métapopulation, décrit la structure de l'espace et la mobilité des individus. (B) La couche sociale décrit la structure de contact au sein de chaque sous-population

2.7 Conclusion

Dans ce chapitre, nous avons présenté, brièvement, les principaux travaux liés à la problématique de la modélisation mathématique de propagation d'épidémie. Nous avons classé les approches de la littérature en quatre classes selon les méthodes utilisées. Nous avons, également, discuté dans quelques sections la faiblesse de l'approche abordée.

Nous signalons ici, que nous avons eu des difficultés pour effectuer ce classement, surtout que dans la plupart des travaux récents, les modèles proposés combinent les différents aspects, par exemple : un modèle à la fois déterministe, stochastique, spatial et complexe.

Dans le cadre de notre thèse, notre proposition consiste à adapter, d'une part : un modèle de compartiments à base d'individus pour modéliser et simuler la propagation d'épidémie. Et d'autre part, un modèle complexe pour générer les relations sociales entre les individus de la population cible. Nous mettons, également, en place de nouveaux algorithmes capables de créer un réseau social de type petit monde et d'appliquer le processus de diffusion d'épidémie. La réalisation de ces objectifs fait l'objet du chapitre suivant.

Première contribution

Chapitre

3

Modèle de propagation d'épidémie proposé

Sommaire

3.1	Introduction	39
3.2	Les objectifs du modèle de propagation SEIR-SW	40
3.3	L'identification du problème	40
3.3.1	Comprendre la dynamique de la maladie	40
3.3.2	Comprendre la répartition spatiale et la dynamique de la population	41
3.3.3	Définition des hypothèses	41
3.4	La description du modèle de propagation SEIR-SW	42
3.4.1	Le cadre générale de la démarche de modélisation	42
3.4.2	Le modèle SEIR-SW	42
3.5	Simulation et résultats expérimentaux	50
3.5.1	Jeux de données	50
3.5.2	L'estimation des paramètres du modèle SEIR-SW	50
3.5.3	Résultats de la simulation	52
3.5.4	Estimation du nombre de reproduction de base : \mathcal{R}_0	53
3.5.5	La validation du modèle SEIR-SW	54
3.5.6	L'analyse du réseau social petit monde (SW)	55
3.6	Conclusion	58

3.1 Introduction

Malgré les progrès en termes de traitement et de prévention, les maladies infectieuses continuent de se propager au sein de la population humaine et il n'y a que très peu de telles maladies qui sont éradiquées. Cela mène à une nécessité d'un suivi et d'un contrôle de la propagation des maladies infectieuses afin de mieux les comprendre.

Dans ce chapitre, nous nous intéressons à la modélisation du phénomène de propagation de l'épidémie de la grippe au sein de la population d'Oran (Algérie). Nous mettons en évidence le phénomène de transmission et la complexité de la dynamique de la maladie d'une part, et la complexité structurelle du réseau sociale et son influence sur la dynamique du système d'autre part. A cet effet, notre première contribution, dans la présente étude, consiste à proposer, un modèle de propagation SEIR-SW capable de comprendre le mécanisme de diffusion étudié.

3.2 Les objectifs du modèle de propagation SEIR-SW

L'objectif principal de notre première contribution, décrite en détails dans les sections suivantes, est de mettre en place un nouveau modèle de simulation hybride par l'intégration de deux modèles à savoir :

- le modèle à compartiments Susceptible, Exposed, Infected, and Removed (SEIR) ;
- le modèle social Small World (SW).

Cette hybridation contribue, efficacement, à la modélisation et la simulation d'une catégorie de systèmes complexes intégrant à la fois l'aspect biologique et l'aspect social de la population. En outre, elle permet de fournir un simulateur (générateur) basé sur le modèle de propagation permettant de générer des données simulées.

D'autres objectifs plus spécifiques, sont envisagés par la mise en place du modèle SEIR-SW [Younsi *et al.*, 2014a] [Younsi *et al.*, 2015b] :

- Le développement d'algorithmes dédiés à la surveillance épidémiologique ;
- Le contrôle de l'évolution d'épidémie dans le temps ;
- La compréhension du mécanisme de propagation d'épidémie ;
- L'identification des facteurs favorisant la transmission de la maladie ;
- L'analyse du réseau social afin de comprendre la topologie structurale de la population ;
- La prévision de l'évolution future de la maladie.

La réalisation de tels objectifs testent essentiellement l'adéquation du modèle au système étudié.

3.3 L'identification du problème

Avant de décrire le modèle SEIR-SW proposé, il est incontournable d'identifier les différents aspects liés au problème étudié. Ces aspects sont décrits dans ce qui suit :

3.3.1 Comprendre la dynamique de la maladie

D'une manière générale, le système biologique peut être modélisé par un ensemble d'éléments interconnectés par des liens représentant le mode de transmission.

Dans ce travail de thèse, nous nous intéressons à la maladie de la grippe saisonnière dont la saisonnalité est très importante dans les pays tempérés. Le phénomène de transmission de cette maladie peut s'observer aussi bien au niveau macroscopique qu'au niveau microscopique.

Au niveau macroscopique : la transmission des agents infectieux résulte de l'interaction entre l'agent pathogène lui-même, l'environnement et l'hôte. On distingue trois modes de transmission : direct, indirect (avec l'intervention d'un élément intermédiaire) et la transmission par l'air (aérienne). Le processus de contagion représenté dans (CHAPITRE 1, FIGURE 2) illustre les différents états de l'individu durant l'infection. Afin de comprendre le mécanisme de propagation de la grippe il est nécessaire de modéliser ce processus.

Au niveau microscopique : les agents infectieux responsables des infections sont des micro-organismes : parasites, champignons, bactéries, virus, etc.

Il existe trois types de grippe saisonnière : A, B et C. Les virus grippaux de type A se divisent en sous-types selon des différentes types de protéines de surface du virus. Parmi

les nombreux sous-types des virus A, on retrouve : « grippe porcine » A(H1N1), « grippe aviaire » A(H5N1).

Le virus (A) a un large spectre d'hôte puisqu'il peut infecter aussi bien les humains que les animaux.

Par contre, le virus de la grippe de type B n'infecte que l'homme, ce qui explique son risque d'infection minimal ainsi que la fréquence d'apparition de nouvelles souches pour l'homme. Les cas de grippe de type C surviennent beaucoup moins fréquemment que ceux des types A et B.

3.3.2 Comprendre la réparation spatiale et la dynamique de la population

La répartition spatiale de la population dans un territoire donné est très inégale, il existe des régions où la densité de la population est très élevée (les zones urbaines) par rapport aux autres (zones rurales). Cela est généralement expliqué par les conditions socio-économiques très favorables des populations urbaines.

Dans le cadre d'épidémiologie, les régions denses constituent des zones à risque d'épidémie. Ces régions sont caractérisées par une population importante avec des interactions hétérogène (familiales, amicales, professionnels, etc.). De plus, il existe plusieurs lieux de rencontres facilitant la propagation de la maladie, tels que : le transport, les rencontres (école, travail, etc.). En revanche, ces zones sont aussi caractérisées par l'accessibilité rapide aux établissements de santé (hôpitaux, cliniques, etc.).

D'autres facteurs touchant l'aspect dynamique de la population, tels que la mobilité des individus et les processus de vie, de mort et de migration constituant des modèles fondamentaux en dynamique des populations.

Dans ce travail, les effets démographiques sont négligés pour le modèle SEIR-SW car l'épidémie étudiée (grippe saisonnière) était pour une courte période, c.-à-d. que les effets démographiques n'influent pas significativement sur l'évolution de la maladie.

3.3.3 Définition des hypothèses

Le virus de grippe change de forme constamment. C'est pourquoi on peut attraper une nouvelle grippe chaque année, ce qui explique la saisonnalité de ce phénomène.

Dans ce travail de thèse, pour la modélisation des épidémies ne seront considérés que les cas où l'infection se propage de manière directe entre individus infectieux (infectés) et individus susceptibles (pas de vecteurs d'épidémie).

En effet, nous optons pour le modèle SEIR car durant la saison 2009-2010, un seul type de virus grippal a circulé : les virus de type A. Aucune souche de type B n'a été détectée durant la période de surveillance de la grippe saisonnière selon le bilan annuel [INSP, 2010]. En effet, nous supposons que les immunisés ne redeviennent pas susceptibles car ils sont immunisés de cette souche après leur guérison.

Afin d'élaborer notre modèle SEIR-SW, nous avons dû poser quelques hypothèses :

H1 : La taille de la population est égale à N , supposée fixe ;

H2 : La variable de temps t est de type discret, tel que $t \in T$ ou T est la durée totale ;

H3 : La période de temps Δt peut représenter des jours ou des semaines ;

H4 : A chaque instant t , la population est partitionnée en quatre classes aléatoires : $S(t)$: ensemble d'individus susceptibles, $E(t)$: ensemble d'individus exposés, $I(t)$: ensemble d'individus infectés et $R(t)$: ensemble d'individus retirés ;

H5 : Nous admettons que chaque individu susceptible dans une période Δt soit exposé, infecté puis guéri ;

H6 : Un individu infecté ne peut plus redevenir susceptible.

3.4 La description du modèle de propagation SEIR-SW

Avant de décrire le modèle de propagation SEIR-SW proposé, il convient de décrire le cadre général de la démarche classique de modélisation adopté par notre étude :

3.4.1 Le cadre générale de la démarche de modélisation

D'après [BenAmor, 2008], le modèle est construit comme étant une abstraction du système étudié. Il fournit des résultats à partir d'une analyse quantitative (analyse mathématique, simulation informatique). Parallèlement, des informations de même nature sont récoltées sur le système étudié. Ces données sont comparées aux résultats de la simulation, pour aboutir à la validation du modèle en cas d'adéquation numérique (données quantitatives). Le Schéma général de la démarche classique de modélisation est illustré par la Figure 3.1.

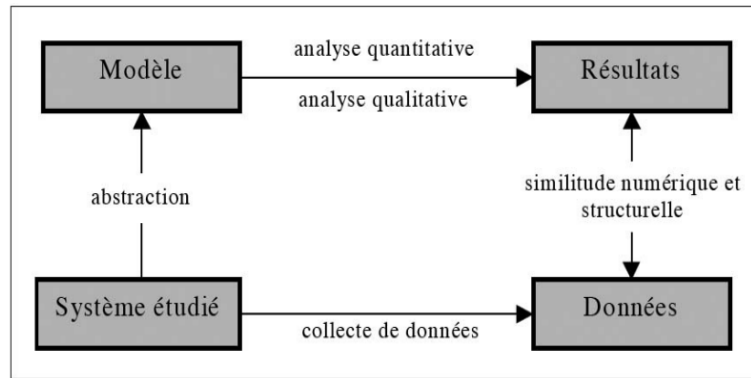


Figure 3.1 – Schéma général de la démarche classique de modélisation

Dans le contexte de notre étude, nous avons adopté la démarche proposée par Ben Amor dans [BenAmor, 2008] en tenant compte de la dynamique de la propagation d'épidémie. Nous partons tout d'abord, d'un ensemble d'hypothèses (posées en section 3.3.3) relativement au système réel que nous jugeons essentielles pour notre modélisation.

L'analyse quantitative du modèle permet de tracer, par exemple, l'évolution de la maladie dans le temps, qui peut être comparée avec des données de terrain (similitude numérique). La validation est une phase essentielle pour que le modèle puisse être utilisé dans une perspective de prévision.

L'élaboration et l'utilisation d'un modèle pour étudier le phénomène de propagation d'épidémie passe par plusieurs étapes dont il est nécessaire de caractériser les principales :

3.4.2 Le modèle SEIR-SW

Dans cette section, nous présentons, en détails le modèle SEIR-SW que nous proposons pour modéliser le phénomène de propagation de la grippe dans la région d'Oran. Ce modèle est conçu par assemblage de deux principaux composants : **le modèle réseau social "petit-monde"** et **le modèle à compartiments d'épidémie SEIR**. La Figure 3.2 illustre le modèle SEIR-SW proposé :

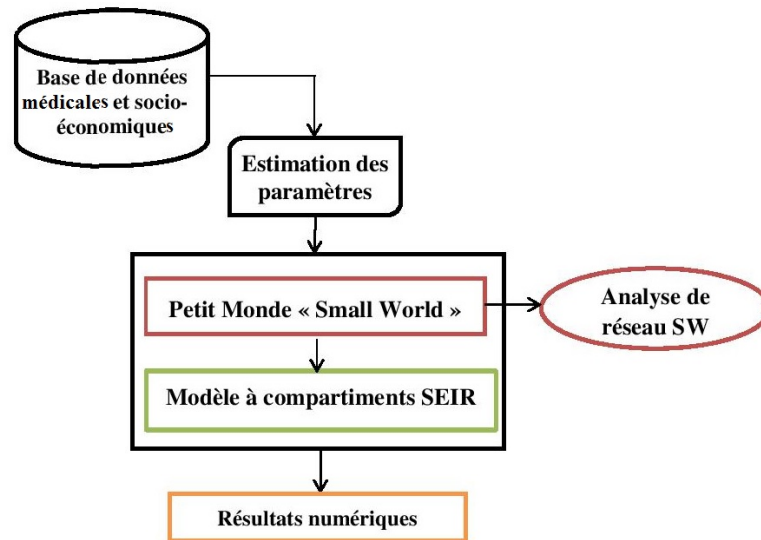


Figure 3.2 – Le modèle SEIR-SW proposé

3.4.2.1 Le modèle réseau social “petit-monde” (Small World)

Les réseaux complexes sont présents dans de nombreux domaines aussi divers les uns que les autres : biologie, sociologie, psychologie, informatique, etc.

Nous utilisons les réseaux complexes afin de tenir compte des interactions permettant la transmission des maladies infectieuses entre les individus. De tels réseaux permettent de prédire l'issue d'une épidémie et d'aider, par conséquent, à tester et à améliorer les politiques de la santé publique.

Un réseau social peut être défini comme : « *un ensemble de personnes ou de groupes de personnes possédant des schémas de contacts ou d'interactions entre eux* », [Degenne et Forsé, 2004]. C'est à partir de ce type de réseaux que la modélisation du monde réel a été introduite de façon empirique grâce à l'expérience de Milgram [Milgram, 1963].

Justification du choix de SW

L'effet petit-monde est aujourd'hui observé sur de nombreux réseaux issus du monde réel [Younsi *et al.*, 2015b]. Plusieurs travaux s'intéressent à l'étude des propriétés structurales pour mettre en évidence des caractéristiques des nœuds ou de la structure. Dans les graphes SW, deux mesures sont à identifier :

1. **la longueur moyenne du plus court chemin** reliant deux nœuds du graphe (average path length), permettant de mesurer l'accessibilité globale du graphe ;
2. **la cohésion (coefficient de clustering)**, permettant de mesurer la probabilité pour que deux nœuds connectés à un troisième soient connectés entre eux, ce qui implique la constitution des cliques au sein du réseau.

La constitution du modèle SW commence par un réseau régulier en anneau, et ensuite un nombre de nœuds supplémentaires est ajouté au milieu du réseau qui est connecté à un grand nombre de sites (nœuds) choisi aléatoirement sur le réseau régulier principal.

Principe de SW

Dans la présente étude, nous avons choisi le modèle petit monde (SW), car ce réseau est le plus proche à la représentation réelle des contacts entre les individus ou les groupes d'individus. Il combine les deux caractéristiques d'un réseau réel : le fort coefficient de clustering et le petit diamètre [Younsi *et al.*, 2014a]. Nous optons aussi aux modèles SW statiques au lieu de dynamique, car ces modèles sont simples et destinés aux épidémies de courte durée. Cela a été prouvé par [Ochab et Góra, 2011] où l'auteur a fait une comparaison entre les deux catégories du modèle SW pour étudier le phénomène de percolation et déterminer le seuil d'épidémie. L'auteur a trouvé que l'écart entre les deux catégories est petit.

Plusieurs études ont utilisé des EDO pour étudier l'effet du petit monde dans les modèles à compartiments d'épidémie. Dans ce travail, nous nous intéressons par une modélisation plus réaliste en utilisant la théorie des graphes. Nous allons développer les méthodes décrites dans cette section pour ajuster le modèle SEIR-SW aux données de la grippe obtenues des établissements de santé de la ville d'Oran, en Algérie. Tous d'abord, nous introduisons les concepts de bases :

Concept de bases

Afin de modéliser les contacts sociaux entre les populations humaines, nous construisons un réseau de SW en représentant les individus par des nœuds et les relations sociales par des liens.

Formellement, un réseau est décrit par un graphe $G = (V, E)$ constitué d'un ensemble de sommets V (appelés aussi nœuds) où $V = \{i \in \text{noeuds}\}$ et d'un ensemble E de liens (appelés aussi arêtes) où $E = \{(i, j)/i, j \in V\}$ d'où i : Nœud source et j : Nœud cible.

Le degré d'un nœud i , représente les voisins du nœud dans un graphe non orienté dénoté $k(i)$, est le nombre de liens entre le nœud i et les autres nœuds du graphe, d'où $k(i) = \{j/(i, j) \in E\}$.

On dénote $T = \{t_0, t_1, \dots, t_m\}$, la séquence de temps pendant laquelle la maladie transmise, et $G = \{G_{t_0}, G_{t_1}, \dots, G_{t_m}\}$ la séquence de réseaux, où chaque élément représente l'état du réseau à l'instant t_l . V est un ensemble d'individus, et E_{ij} , est l'ensemble des arêtes présentes dans le réseau au moment t_l .

La génération du réseau SW : le processus proposé

La construction du réseau SW passe par deux principales étapes [Younsi *et al.*, 2015a] :

1. la construction d'un réseau en anneau régulier ;
2. la reconnexion des arêtes du graphe (réseau).

Étape 1 : Construire un réseau régulier en anneau

Nous commençons par la création d'un réseau régulier en anneau $G(t)$ avec N nœuds (le nombre total de nœuds représentant la population totale) et k arêtes par sommet (k désigne le degré d'un nœud ou le nombre de voisins).

Nous supposons que les réseaux de contact à construire seront des graphes non orientés de telle sorte que le lien entre deux nœuds (source, cible) est le même que (cible, source).

Nous mettons en place un algorithme permettant de construire le réseau régulier. Le processus que suit l'algorithme est le suivant :

1. Dans le réseau initial, le nombre total d'arêtes pour chaque nœud est donné par le paramètre k . Afin de respecter les caractéristiques de SW citées dans [Watts et Strogatz, 1998], k doit être inférieure à la valeur de $\frac{(N-1)}{2}$, cette condition est vérifiée par la ligne 2 dans l'Algorithme 1.
2. Ensuite, nous connectons chaque nœud à son k voisin de chaque côté, cette étape est assurée par les lignes (de 4 à 11).
3. Pour construire le réseau en anneau, nous devons connecter les derniers nœuds dans la liste aux premiers nœuds de la même liste. En effet, lorsque l'indice du nœud cible est supérieur strictement au nombre de nœuds totaux ($d > N$), nous connectons le nœud source $[i]$ par le nœud cible $[d - N]$.

Dans ce qui suit, nous proposons un algorithme permettant de construire un réseau régulier en anneau.

Algorithm 1: Construire le réseau régulier en anneau

```

Data: N, k
Result: G(t)
1 BEGIN
2 for ( $k < (N - 1)/2$ ) do
3   for  $i = 1$  to  $N$  do
4     for  $j = 1$  to  $k$  do
5        $d \leftarrow (i + j)$ ;
6       if ( $d \leq N$ ) then
7         Connecter le Nœud $[i]$  avec le Nœud $[d]$  ;
8       else
9         Connecter le Nœud $[i]$  Connecter le Nœud $[d - N]$  ;
10      end
11    end
12  end
13  Avoir  $G(t)$ ;
14 end
15 END

```

Étape 2 : Reconnecter les arêtes avec une probabilité aléatoire (P_w)

Le processus que suit l'algorithme est le suivant :

1. Dans le graphe $G(t)$, quelques arêtes sont reconnectées aléatoirement avec une probabilité de reconnexion (P_w) en remplaçant (Nœud $[i]$, Nœud $[j]$) avec (Nœud $[i]$, Nœud $[m]$) , nous évitons l'auto-boucles et lien dupliqué. Cette construction permet de basculer entre le réseau régulier ($P_w = 0$), réseau aléatoire ($P_w = 1$), et le réseau intermédiaire ($0 < P_w < 1$) comme le réseau SW (CHAPITRE 1, FIGURE 3).
2. Dans le processus de reconnexion, les arêtes ne sont pas rajoutées ou supprimées mais plutôt remplacées (lignes 8 à 12, dans l'Algorithme 2). Par conséquent, le degré moyen d'un nœud ainsi que le nombre total des arêtes sont inchangés. k n'aura pas une valeur fixe comme dans le réseau régulier, mais il aura une valeur variable pour chaque nœud (avoir une nouvelle répartition de k). Ces arêtes représentent les interactions aléatoires entre les groupes éloignés (exemple : rencontrer des gens dans le transport urbains).

Dans ce qui suit, nous formalisons l'étape 2 sous forme d'algorithme décrivant les étapes de reconnexion des arêtes.

Algorithm 2: Reconnecter les arêtes aléatoirement avec une probabilité (P_w)

Data: $G(t)$, k , p_w
Result: $G(t+1)$

```

1 BEGIN
2 for  $i = 1$  to  $N$  do
3   for  $j = (i + 1)$  to  $i + k/2$  do
4     if  $(j > N)$  then
5        $j \leftarrow j - N$ ;
6     end
7      $C =$  une variable aléatoire uniforme entre 0 and 1;
8     if  $(P_w > C)$  then
9       Noeud[m] est choisi uniformément et aléatoirement de l'ensemble de noeuds , excluant (i)
        et les noeuds adjacents de/vers (i) autre que (m);
10      Déconnecter : Noeud [i] et Noeud[j];
11      Connecter : Noeud [i] et Noeud [m] ;
12    end
13     $t \leftarrow t + 1$ ;
14  end
15 end
16  $G(t)$  évolue à  $G(t+1)$ 
17 END

```

Résultats d'exécution de l'algorithme SW

L'exécution de l'algorithme SW fourni deux listes de nœuds : **des nœuds de source** et **des nœuds cibles** .

Dans la littérature relative à la théorie des graphes, la représentation des nœuds adjacents existent sous deux formes : **la matrice d'adjacence** et **la liste d'adjacence**.

Dans le présent travail, nous optons pour la liste d'adjacence pour représenter notre graphe SW afin de l'analyser et le visualiser.

Ce choix est motivé par des considérations d'implémentation de la méthode : une matrice d'adjacence d'un graphe stockée dans un tableau et ayant N nœuds requiert une complexité de l'ordre de $\mathcal{O}(N * N)$.

Par ailleurs, la place requise par une liste d'adjacence est proportionnelle au nombre de sommets (N) et d'arêtes du graphe c.-à-d. qu'elle est plus efficace en espace. La représentation de la liste d'adjacence adoptée est illustrée par la Figure 3.3.

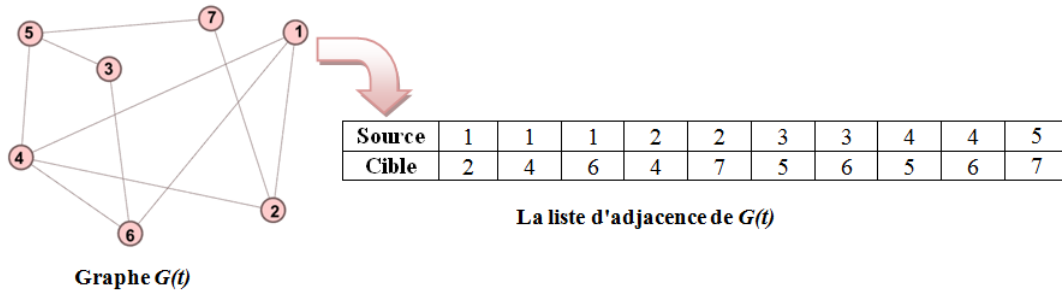


Figure 3.3 – Représentation schématique de la liste d'adjacence et le graphe généré

Nous avons, également, étudié l'effet de la probabilité de reconnexion sur la distribution des degrés. La figure 3.4 montre la distribution des degrés pour le modèle SW en fonction des P_w (probabilité de rewiring), pour $N = 10000$, $k = 6$ et P_w varie entre (0, 1; 0, 3; 0, 6; 1).

Nous rappelons que le degré d'un nœud est le nombre d'arêtes qui lui sont incidents. Les distributions de degrés de nœud obtenues sont en forme de cloche observées dans le graphe d'Erdős–Rényi [Erdős et Rényi, 1959], Watts–Strogatz [Watts et Strogatz, 1998] et Newman–Watts [Newman et Watts, 1999] sont identiques aux distributions de degrés en loi de puissance.

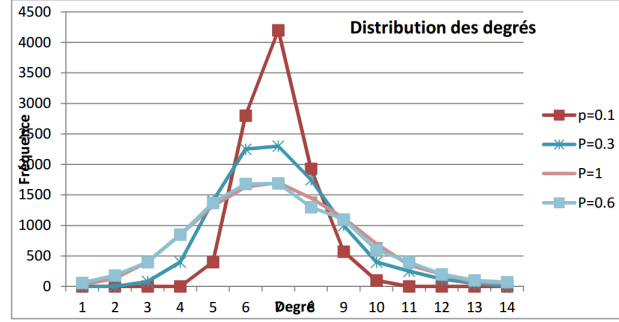


Figure 3.4 – Distribution des degrés dans un graphe SW

3.4.2.2 Le modèle compartimental d'épidémie SEIR

Plusieurs modèles ont reçu une attention considérable dans les disciplines de l'épidémiologie, et en particulier dans la propagation de l'épidémie, nous citons : modèle à compartiments (SI, SIR, SEIR, etc.), automates cellulaires, modèle à base d'agent, etc.

Dans le cadre de notre thèse, nous nous focalisons sur l'un de ces modèles : *le modèle à compartiments SEIR*.

La plupart des modèles étudiés en épidémiologie mathématique classique sont de la classe de modèles à compartiments, car dans la plupart des cas, la population étudiée est divisée en classes ou compartiments définis par le statut épidémiologique [Brauer, 2008].

Le modèle épidémique simple SEIR est un bon modèle pour étudier la propagation de la maladie de la grippe dans une population fixe. Dans ce modèle, la population est composée de quatre classes d'individus (compartiments) : $S(t)$, $E(t)$, $I(t)$ et $R(t)$ d'où le nombre d'individus est variable dans le temps. Ces classes sont utilisées pour représenter l'état de santé des individus comme il est représenté par la Figure 3.5.

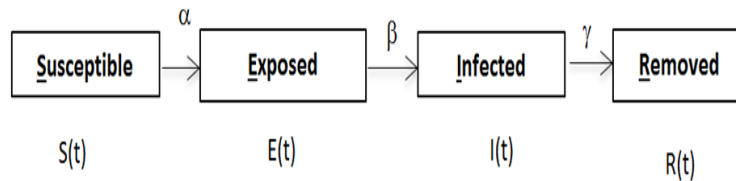


Figure 3.5 – Représentation schématique du modèle de diffusion épidémique SEIR

Dans cette optique, nous sommes, particulièrement, intéressés par la modélisation d'épidémies où il y'a une guérison permanente d'un individu atteint de la maladie, les infectés sont distribués d'une façon hétérogène dans l'espace comme dans le temps.

Chaque individu susceptible se déplace de la classe des susceptibles à la classe de latente avec un taux (α) suite à un contact direct avec une personne infectée, où (α) représente le taux de l'infection d'un seul contact avec une personne infectée.

- Le paramètre (β) est le taux pour lequel ceux de la classe latente passent à la classe infectieuse
- Le paramètre (γ) est le taux de récupération pour lequel les personnes infectées se rétablissent c.-à-d. passer de la classe infectieuse à la classe résistante (retiré).

En effet, la modélisation d'épidémie ne considère que le cas où l'infection se propage de manière directe entre les individus infectés et les individus susceptibles (pas de vecteurs d'épidémie).

Le choix du modèle à compartiments

La simulation signifie le pilotage du modèle d'un système avec des entrées adéquates tout en observant les résultats de sortie correspondants. Les fins de la simulation comprennent non seulement l'apprentissage et la prédiction, mais aussi, ils s'intéressent à l'aspect exploration et découverte de propriétés et de principes organisateurs des phénomènes étudiés [Axelrod, 1997]. Dans la littérature, les deux modèles déterministes et stochastiques ont été développés pour l'épidémie par plusieurs auteurs dont nous citons : [Bailey *et al.*, 1975], [Anderson *et al.*, 1992], et West et Thompson [West et Thompson, 1997]. Ces auteurs ont effectué une étude comparative de ces modèles et ont conclu que : pour une population initiale de susceptibles de grande taille, les résultats obtenus par le modèle stochastique sont approximativement équivalents aux résultats obtenus par le modèle déterministe.

Tous les symboles utilisés dans cette section ont leurs significations indiquées dans la Table 3.1 :

Table 3.1 – Synthèse des paramètres et des variables d'état utilisés par SEIR-SW

Variables d'état	Signification
S	Nombre d'individus susceptibles
E	Nombre d'individus exposés
I	Nombre d'individus infectés
R	Nombre d'individus retirés
$S_0 = S(0)$	Nombre d'individus susceptibles à $t=0$
$E_0 = E(0)$	Nombre d'individus exposés à $t=0$
$I_0 = I(0)$	Nombre d'individus infectés à $t=0$
$R_0 = R(0)$	Nombre d'individus retirés à $t=0$
Paramètres	Signification
k	Degré de distribution
α	Taux de transmission
β/β^{-1}	Taux de latence / Période de latence
γ/γ^{-1}	Taux de rétablissement / période d'infection

Dans la présente étude, nous considérons une population d'individus fermée et homogène compartimentés à tout moment $t \geq 0$ en quatre catégories : susceptibles, exposés, infectés et retirés. Avec $N = S + E + I + R$ à chaque unité de temps t , où N est la taille totale de la population et avec les conditions initiales : $S(0) = S_0$, $E(0) = E_0$, $I(0) = I_0$ et $R(0) = R_0$, respectivement.

Le processus d'application du modèle SEIR

Au cours de la maladie, les individus peuvent avoir quatre états : $V = \{Noeud.S, Noeud.E, Noeud.I, Noeud.R\}$. Nous définissons un ensemble de voisins infectés du nœud i comme $A_{t_l}^i = \{u \in V / (i, u) \in E_{t_l} \text{ et } \text{état}(u) = Noeud.I\}$

Soit $S(t) = \{Noeud.S \in V\}$ au temps t et $I(t) = \{Noeud.I \in V\}$ est le nombre d'individus infectés à cet instant. Si $Noeud.I$ a un contact direct avec $Noeud.S$ et $p = 1 - \alpha$ est la probabilité de ne pas transmettre l'infection, alors, $(1 - \alpha)^{k_i}$ est la probabilité de ne pas transmettre l'infection entre $Noeud.I$ et ses k voisins $Noeud.S$. Par conséquent, $1 - (1 - \alpha)^{k_i}$ est la probabilité d'infection. On définit $I(t)$ comme la liste des individus infectés et $G(t)$ comme un réseau SW obtenu par l'Algorithme 2.

L'algorithme détaillé du modèle SEIR-SW est le suivant :

Algorithm 3: L'algorithme du processus SEIR-SW

Data: $G(t)$, k , α , β , γ
Result: $I(t)$

```

1 BEGIN
2 Time :  $t \leftarrow t_0$ ;
3  $I(t) = \emptyset$ ;
4 Initialement : infecter quelques nœuds
5 while  $I(t) \neq \emptyset$  do
6   Noeud.S devient Noeud.E avec la probabilité  $1 - (1 - \alpha)^{k_i}$  with  $k_i = |A_{t_l}^i|$  ;
7   Noeud.E devient Noeud.I après une période  $t \sim E(1/\beta)$ ;
8   Ajouter Noeud.I à  $I(t)$ ;
9   Noeud.I devient Noeud.R après une période  $t \sim E(1/\gamma)$ ;
10   $G$  évolue à  $G(t+1)$ ;
11   $t \leftarrow t + 1$ 
12 end
13 Retourner  $I(t)$ 
14 END
```

Le processus général de déroulement du modèle SEIR-SW est comme suit :

1. Construction d'un réseau régulier en anneau ;
2. Reconnexion des nœuds aléatoirement avec une probabilité (P_w) ;
3. Un nombre d'individus initialement infectés est choisi ;
4. La maladie se propage dans le réseau G avec une probabilité de transmission d'individus infectés aux individus susceptibles s'il existe une relation entre eux ;
5. Ensuite, chaque individu sera infecté après avoir été exposé. Dans de nombreux travaux de modèles épidémiologiques, la période de latence (durée d'exposition) et d'infection sont généralement supposées être soit une distribution exponentielle ou fixe [Beretta et Kuang, 1998], [Beretta et Kuang, 2001].
 Ici, nous considérons que la période de latence et la période infectieuse sont réparties selon la fonction de distribution exponentielle car la période de latence (incubation) et la période infectieuse sont variables d'une personne à une autre ;
6. Chaque personne infectée sera guérie (immunisée contre le virus) après une période d'infection.
 Après la guérison, il ne peut pas transmettre à nouveau l'agent infectieux. Son immunité est permanente et il ne peut pas retourner dans la classe des susceptibles.
7. Le réseau SW $G(t)$ évolue à $G(t + 1)$.

Les étapes de 4 jusqu'à 7 seront répétées jusqu'à aucun individu ne soit infecté.

3.5 Simulation et résultats expérimentaux

3.5.1 Jeux de données

La base de données utilisée dans cette étude concerne la maladie de la grippe saisonnière de l'année 2009 de la région d'Oran. C'est une base de données médicale/socio-économique obtenue de la Direction de la Santé et de la Population (DSP) d'Oran. Elle est composée de 5504 enregistrements (déclarations) pour les 26 communes de la wilaya d'Oran. Les données recueillies pour chaque patient incluent :

- son identifiant,
- la date de déclaration de la maladie (d'Août 09 à Décembre 09),
- l'âge (1 mois -93 ans),
- le sexe (F-M),
- la commune de résidence et
- le centre de santé de déclaration de la maladie (14 centres entre : Etablissements Publics de Santé de Proximité (EPSP), Etablissement Hospitalier de Santé (EHS), Établissement Public Hospitalier (EPH), Etablissement Hospitaliers (EH)).

Pendant cette période, la région d'Oran a connu deux vagues de grippe saisonnière, la première était entre fin Août et fin d'Octobre 2009 (S_{35} - S_{42}) et la deuxième a commencé fin Octobre jusqu'au fin de décembre 2009 (S_{43} - S_{52}). La Figure 3.6 (3.6a, 3.6b) représentent respectivement les courbes d'incidence pour les deux vagues dans la population totale.

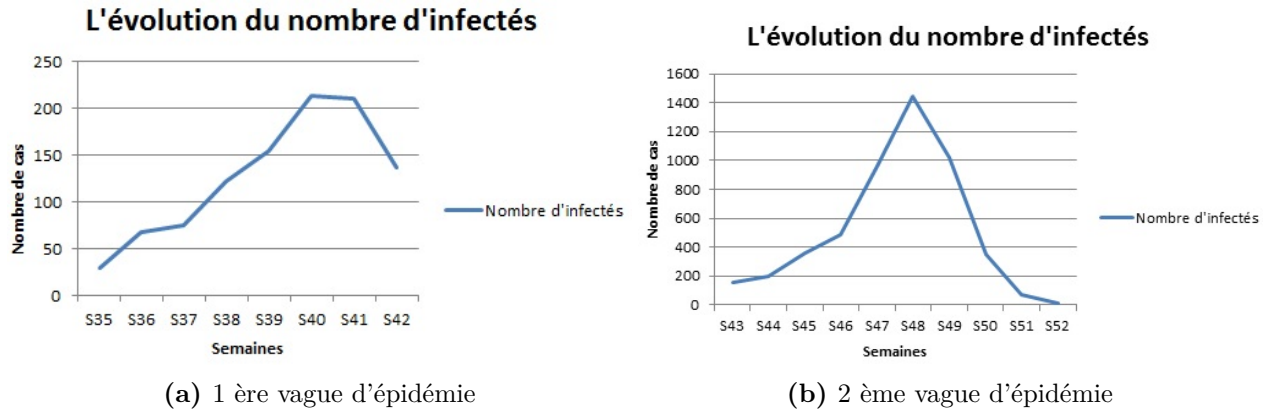


Figure 3.6 – L'évolution du nombre d'infectés de la grippe

3.5.2 L'estimation des paramètres du modèle SEIR-SW

Plusieurs chercheurs ont fourni des estimations fiables de la durée moyenne des périodes de latence et d'infectiosité. Il n'existe pas de mesure directe permettant de disposer de la vitesse de transmission, et très peu d'estimations de la fraction des individus initialement susceptibles dans la population [Cauchemez *et al.*, 2008].

Dans ce qui suit, nous allons estimer les paramètres non disponibles du modèle SEIR-SW, soit à partir des données de terrain (la base de l'épidémie de grippe survenue à Oran en 2009) soit par des méthodes d'ajustements [Younsi *et al.*, 2014a].

3.5.2.1 Le processus d'estimation du modèle SEIR-SW

Nous présentons ici comment estimer les paramètres du modèle SEIR-SW à partir de données réelles. L'idée générale est de comparer les prédictions du simulateur SEIR-SW avec les données et de trouver les valeurs des paramètres qui minimisent la différence entre les prédictions du modèle et les données.

Nous décrivons, dans ce qui suit, les différentes étapes d'estimation de paramètres inconnus pour appliquer SEIR-SW :

1. Tout d'abord, nous avons choisi des valeurs initiales pour les paramètres de l'épidémie (k_0 , P_w , α_0 , β_0 , et γ_0) ;
2. Afin d'appliquer la méthode d'ajustement, nous avons transformé les courbes représentées par les Figures (3.6a et 3.6b) sous forme d'équations non linéaires avec l'utilisation de logiciel Origine Lab¹ ;
3. Finalement, nous avons comparé les résultats (sorties) du modèle SEIR-SW avec les données réelles pour déterminer le nouvel ensemble de valeurs de paramètres de SEIR-SW pour un meilleur ajustement aux données réelles. Pour se faire, nous avons appliqué l'algorithme de Levenberg-Marquardt qui permet d'obtenir efficacement le minimum, la solution déterminée par cet algorithme correspond au nouvel ensemble de paramètres optimaux (k^* , α^* , β^* , γ^*). Le critère utilisé est celui du 2%

Le processus d'estimation de paramètres est représenté par la Figure 3.7 :

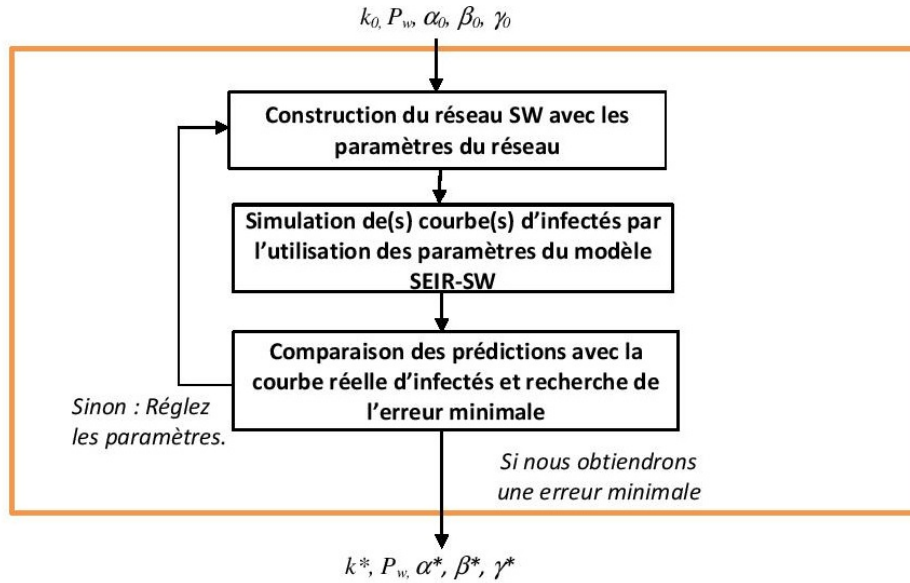


Figure 3.7 – Représentation schématique décrivant le processus de l'algorithme d'estimation pour une vague d'épidémie

3.5.2.2 Présentation des paramètres du modèle SEIR-SW

Dans le modèle SEIR-SW, nous distinguons différents paramètres obtenus à partir de différentes sources :

- les paramètres α et γ : peuvent être obtenus à partir des caractéristiques du virus de la grippe [Cori *et al.*, 2012].

1. Logiciel de traitement et d'analyse de données scientifiques.

- la probabilité de reconnexion $P_w \approx 0,10$, selon les recommandations de [Neuman et Mizruchi, 2010] et [Watts et Strogatz, 1998].
- les autres paramètres doivent être estimés à partir des méthodes d'estimation ou à partir de la base de données réelle.

Les paramètres et les variables d'état exploités par SEIR-SW sont donnés par la Table 3.2. Ces valeurs nous permettent de modéliser la propagation de la grippe pour une courte durée de temps.

Table 3.2 – Les valeurs des paramètres et des variables d'état utilisées par SEIR-SW

	Paramètres	Vague 1	Vague 2	source
Paramètres relatifs à la population	S_0	1350	5020	Estimés
	I_0	8	157	Base de Données
Paramètres relatifs au réseau SW	k	8	10	Estimé
	P_w	0,1	0,1	[Neuman et Mizruchi, 2010]
Paramètres relatifs à la maladie	α	0,16	0,13	Estimés
	β	2,2	2	[Cori <i>et al.</i> , 2012]
	γ	4,5	4	[Cori <i>et al.</i> , 2012]

3.5.3 Résultats de la simulation

Pour valider notre modèle, nous avons effectué des expériences de simulation avec les paramètres illustrés par le tableau ci-dessus (Table 3.2). Les résultats obtenus sont représentés par la Figure 3.8 (3.8a, 3.8b). Nous comparons l'évolution des courbes d'incidence obtenues par SEIR-SW et les données de la grippe pour les deux vagues épidémiques. Nous remarquons que l'écart entre les deux courbes est très petit, cela signifie que les paramètres choisis sont proches des paramètres réels et de la structure sociale d'individus.

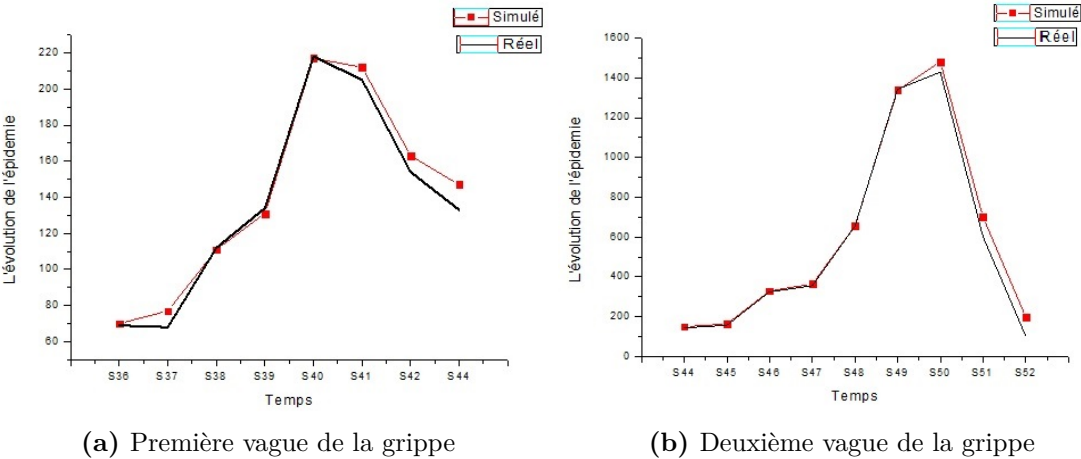


Figure 3.8 – L'ajustement des données réelles et simulées

Discussions

La Figure 3.8 montre les courbes d'infection résultant de l'application de l'algorithme SEIR-SW, par rapport à la courbe d'infection des données réelles. Ces courbes qui représentent l'évolution du nombre d'infectés $I(t)$ dans le temps pour les deux vagues sont différents.

Nous remarquons, que dans la deuxième vague, le nombre d'infectés est plus important que dans la première, ceci est expliqué par le nombre élevé des susceptibles ainsi que le degré par rapport à la première vague.

En outre, le nombre initial des personnes infectées était important ($I_0 = 157$) avec un nombre de contact au moyen ($k = 10$) par personne. Tous ces facteurs favorisent la transmission de l'épidémie grippale.

Pour la première vague, la population infectée continue à augmenter jusqu'à la semaine (S_{40}), puis elle diminue.

Pour la seconde vague, le nombre de personnes infectées va d'abord augmenter avec la propagation de l'épidémie, et ensuite il va diminuer après que l'épidémie soit mise sous contrôle après la semaine (S_{49}).

Les simulations effectuées nous permettent de constater que les quatre paramètres (α, β, γ et k) ont une influence significative sur la propagation de l'épidémie ainsi qu'une incidence sur la diffusion de l'épidémie et sur ses effets et exigeront donc la mise en place des contrôles et des mesures efficaces.

Par conséquent, toute modification de ces valeurs, changera complètement les résultats de la simulation. Les paramètres (α, β, γ) sont liés aux caractéristiques de la maladie de la grippe et, k , est liée aux facteurs sociaux. Lorsque l'une et / ou toutes les valeurs des paramètres augmente l'épidémie se propage rapidement et vice versa.

Les résultats obtenus par le simulateur SEIR-SW confirment que le couplage des modèles classiques à compartiment avec le modèle de réseau de contact est efficace. L'intérêt de ce couplage est d'exprimer de façon réaliste et satisfaisante le système de propagation d'épidémie. Ce dernier est un système multi-échelle car nous avons à la fois :

- **une population** représentée par un réseau de contact composé d'un ensemble de nœuds (individus) interconnectés par des liens (relations divers),
- **un territoire** sur lequel les personnes habitent et la simulation est appliquée, et
- **un processus de diffusion** tenant compte les différents états de santé de chaque individu dans le temps.

En effet, afin de mieux comprendre l'influence de ces facteurs, il est nécessaire d'étudier le nombre de reproduction de base (\mathcal{R}_0).

3.5.4 Estimation du nombre de reproduction de base : \mathcal{R}_0

La dynamique de la transmission de personne à personne qui concerne la transmission vectorielle peut être résumée par deux principaux paramètres : le taux de reproduction et l'intervalle sériel ou intergénérationnel ² [Anderson *et al.*, 1992].

Un paramètre d'intérêt du modèle SEIR-SW est le taux de reproduction \mathcal{R}_0 , qui s'interprète comme le nombre moyen de nouveaux cas générés par un sujet infectieux dans une population susceptible. Ce paramètre est très important dans l'étude de la propagation des épidémies, il permet d'évaluer la virulence d'une maladie infectieuse, par exemple, le suivi de l'évolution de la virulence en cas de maximisation de la valeur de \mathcal{R}_0 et, par conséquent, de prévoir l'évolution de l'épidémie dans le temps. Plus la valeur de $\mathcal{R}_0 > 1$ est grande, plus l'épidémie est importante et peut s'étendre.

Dans le cas d'épidémie de la grippe saisonnière, ce paramètre est mesuré entre $\mathcal{R}_0 = 1,5$ et $\mathcal{R}_0 = 2$

Pour une maladie infectieuse avec une période d'infection moyenne égale à $\frac{1}{\gamma}$, un taux de transmission égale à α et un degré de nœud moyen égale à k . Le nombre de reproduction

2. Intervalle de temps moyen séparant deux cas successifs d'une même chaîne de transmission : $\approx 2,5$ jours pour les gripes saisonnières

de base \mathcal{R}_0 pour le modèle SEIR-SW est donné par :

$$\mathcal{R}_0 = \alpha * k * \frac{1}{\gamma} \quad (3.1)$$

Dans notre étude, nous avons calculé le \mathcal{R}_0 pour les deux vagues. La valeur de $\mathcal{R}_0 = 0,28$ pour la première vague et $\mathcal{R}_0 = 0,325$ pour la deuxième vague.

Cela signifie que dans les deux vagues d'épidémie de la grippe, l'état d'équilibre est endémique.

3.5.5 La validation du modèle SEIR-SW

Une analyse statistique a été appliquée pour valider le modèle SEIR-SW proposé. C'est une étape importante dans tout processus de construction de modèle. Nous avons utilisé le test de Kolmogorov–Smirnov et l'analyse de corrélation linéaire pour comparer les données observées (réelles) et les simulations numériques (les sorties du modèle SEIR-SW).

3.5.5.1 Corrélation linéaire

La corrélation est largement utilisée dans les sciences comme une mesure du degré de dépendance linéaire entre deux variables. Dans ce travail, la corrélation linéaire est utilisée pour mesurer le degré de liaison entre les données simulées obtenues par SEIR-SW et les données observées (réelles). La Table 3.3 présente les résultats obtenus.

Table 3.3 – La corrélation entre les données des simulations numériques de SEIR-SW et les données observées

	Vague 1	Vague 2
Corrélation	0,887	0,80

Des corrélations positives et significatives ($p < 0,05$) ont été enregistrées entre les données simulées et les données réelles. Donc, les données simulées sont liées significativement aux données réelles.

En effet, nous constatons que le modèle SEIR-SW modélise avec précision le comportement dynamique des données observées utilisées dans notre étude.

3.5.5.2 Le test de Kolmogorov–Smirnov

Définition du test de Kolmogorov–Smirnov

Le test de Kolmogorov–Smirnov est un test d'adéquation non paramétrique utilisé pour déterminer si un échantillon suit une loi donnée connue par sa fonction de répartition continue $F(x)$, ou encore si deux échantillons suivent une même loi.

En effet, c'est un test d'ajustement visant à vérifier si les données observées sont compatibles avec un modèle théorique donné.

Soit x_1, x_2, \dots, x_n , une suite de n réalisations de la variable aléatoire x indépendantes et identiquement distribuées (i.i.d) de la fonction de répartition F .

On suppose les hypothèses suivantes :

H_0 : x suit la loi de F contre H_1 : x suit une autre loi c.-à-d. $F = F_0$ contre $F \neq F_0$.

Le test de Kolmogorov–Smirnov est défini par la statistique de test :

$$D_{max} = \sup |F_n(x) - F_0(x)| \quad / x \in \mathbb{R} \quad (3.2)$$

Il consiste à rejeter l'hypothèse H_0 si $D_{max} \geq D_n(n, a)$. Autrement dit, plus la statistique de test D est grande et plus on a de chance de rejeter H_0 . Ainsi, le test de Kolmogorov–Smirnov pour une taille de l'échantillon (n) à un niveau de signification (a) possède une valeur critique de $D_n(n, a)$.

Application du test de Kolmogorov–Smirnov

Dans notre étude, nous allons appliqué le test de Kolmogorov–Smirnov pour les données simulées et les données réelles, pour les deux vagues. Ensuite, nous allons testé l'hypothèse pour que les distributions des deux ensembles de données soit identiques, c.-à-d., les deux fonctions ont les mêmes fonctions de distribution, ce qui confirme l'hypothèse nulle (H_0 ³). La statistique de test, ici désigné D_{max} , était l'écart maximal entre le cumulatif des proportions des deux fonctions. La Table 3.4 représente les résultats du test de Kolmogorov–Smirnov pour les deux vagues de grippe.

Table 3.4 – Illustration des résultats du test de Kolmogorov–Smirnov

Vagues d'épidémie de la grippe saisonnière	Valeurs de test de Kolmogorov–Smirnov (D_{max})	n	a	$D_n(n, a)$
Vague 1	0,0126	8	0,05	0,454
Vague 2	0,0293	9	0,05	0,43

La taille d'échantillon représente le nombre de semaines pour les deux vagues. Les valeurs de $D_n(n, a)$ sont obtenues de la Table de la loi de Kolmogorov–Smirnov.

Pour la première vague, nous avons trouvé :

$$D_{max} = 0,0126$$

$$D_n(8, 0.05) = 0,454$$

En effet, $D_{max} < 0,454$, on accepte l'hypothèse nulle (H_0) à un niveau de 5 % de confiance. Les deux distributions sont conformes.

Pour la deuxième vague, nous avons trouvé :

$$D_{max} = 0,029261391$$

$$D_n(8, 0.05) = 0,43$$

En effet, $D_{max} < 0,43$, soutenant à nouveau l'hypothèse nulle (H_0) à un niveau de 5% de confiance et les deux distributions sont conformes.

Nous constatons que nous acceptons à chaque fois l'hypothèse (H_0), ce qui est en adéquation avec la distribution de l'échantillon.

Dans la section suivante, nous allons aborder la partie d'ARS "petit-monde (SW)", nous étudions les principales caractéristiques topologiques de ce réseau.

3.5.6 L'analyse du réseau social petit monde (SW)

Les réseaux du monde réel sont souvent de natures complexes et diverses : professionnelles, parentales, amicales, etc. Plusieurs méthodes et algorithmes visent à analyser les réseaux sociaux par l'analyse de ses propriétés structurelles. Dans cette section, nous analysons certaines des principales propriétés topologiques de notre réseau SW, y compris la distribution de degré, le coefficient de clustering, et le diamètre du réseau pour les deux vagues d'épidémies. Ces réseaux ont été obtenus à partir des simulations numériques du

3. Les deux échantillons proviennent de la même distribution

modèle SEIR-SW. La Table 3.5 résume les valeurs des propriétés de SW, pour les deux vagues de grippe épidémie.

Table 3.5 – L'analyse du réseau SW pour les deux vagues de la grippe

Propriétés du réseau social SW	Première vague	Deuxième vague
Nombre de nœuds	10 864	51 770
Nombre de liens	8	157
Longueur moyenne de chemin	3.369	3.668
Coefficient de clustering	0,386	0,38
Diamètre	5	5
Distribution de degré moyen	8	10
Densité	0,012	0,004
Modularité (nombre de communautés)	0,723 (17)	0,759 (32)

Discussion

Malgré la non comparabilité de certaines dimensions pour les deux réseaux comme par exemple : la longueur moyenne de chemin, le coefficient de clustering, la modularité, et diamètre, la distribution de degré et la densité de chaque réseau étaient comparables.

En outre, l'augmentation du nombre de nœuds et de liens selon le modèle que nous avons proposé (SW) influence les principales propriétés des réseaux. Autrement dit, nous obtenons le même diamètre (d), le même coefficient de clustering et la même longueur de chemin.

Nous obtenons aussi la même modularité sauf que le nombre de communautés est double, cela est dû à la multiplication du nombre des nœuds par (4) : 1358 pour la première vague à 5177 pour la deuxième vague.

En effet, notre modèle permet de générer des réseaux d'individus infectés à partir des données estimées.

La propagation de la maladie entre les individus dépend, donc, de la structure et de la connexité du RS. Autrement dit, si le RS a une seule composante connexe tout le monde ou une grande partie du RS peut être infectée. Par contre, si le RS est composé de petits groupes d'individus, peu d'individus dans le réseau peuvent être infectés. Nos simulations sont appliquées à la même population (population d'Oran) sur une courte durée de temps (quelques semaines), ce qui signifie que la plupart des caractéristiques des réseaux ont été constante.

Ces résultats ont démontré la nature des réseaux sociaux SW, et ont montré que notre modèle SEIR-SW peut générer des réseaux d'individus infectés en utilisant les données estimées. En outre, ils confirment les résultats de l'étude de [Ochab et Góra, 2011] permettant de supposer que les réseaux SW statiques et dynamiques sont presque identiques pour les épidémies de courte durée.

Pour illustrer la puissance de l'ARS et afin de maîtriser comment les réseaux de contacts des individus changent leurs structures topologiques selon les paramètres du modèle choisis initialement, nous utilisons la plateforme de visualisation Gephi et nous choisissons l'algorithme de spatialisation Force Atlas.

Nous montrons comment certaines mesures (la distribution de degré, la modularité, et le coefficient de clustering) pourraient influencer directement sur l'évolution d'une maladie dans un réseau de contact. Les Figures 3.9 et 3.10 montrent les résultats de l'analyse de SW :

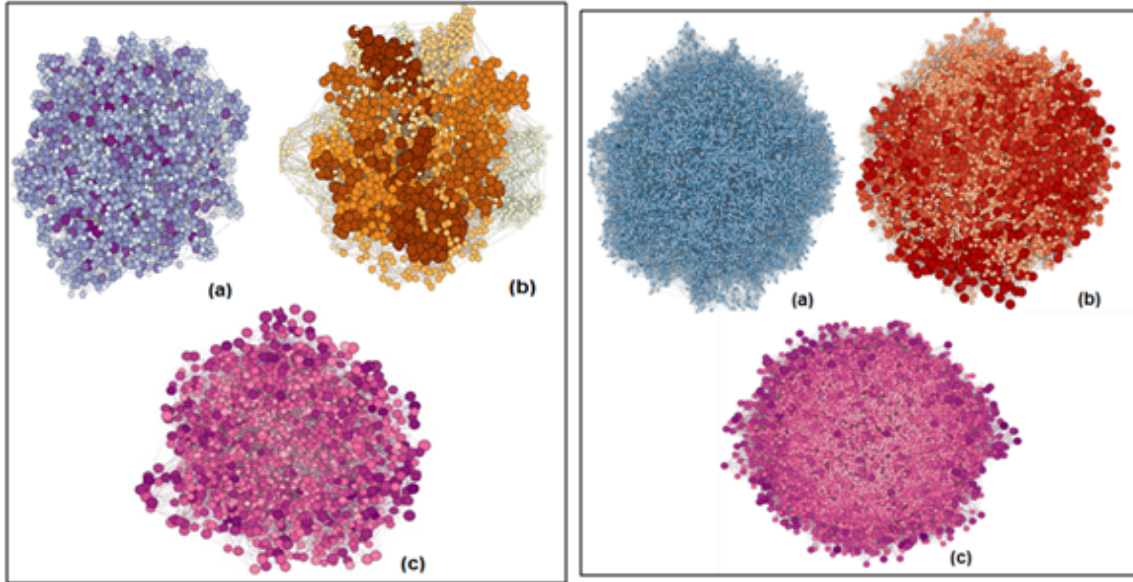


Figure 3.9 – Visualisation du réseau SW pour la première vague de la grippe : (a) degré de distribution, (b) modularité, et (c) coefficient de clustering

Figure 3.10 – Visualisation du réseau SW pour la deuxième vague de la grippe : (a) degré de distribution, (b) modularité, et (c) coefficient de clustering

Nous comparons les deux figures 3.9 et 3.10 :

- Nous remarquons que les deux facteurs **le nombre de personnes susceptibles** et **le nombre de voisins d'un individu (k)** ont un impact significatif sur la topologie du réseau. Quand la valeur de ces deux facteurs augmente, la structure du réseau devient plus complexe, et de nombreuses communautés apparaissent.
- Nous observons, également, que la transmission de la grippe est conforme avec la structure de la communauté de réseau SW, et que l'épidémie se propage rapidement dans les groupes d'individus densément connectés. Par conséquent, au sein d'un réseau de contact, le nombre k doit être maintenu faible, car la maladie se propage plus vite si cette valeur de degré devient importante.

Les décideurs et les professionnels de la santé publique peuvent simuler différents scénarios d'épidémie en utilisant le modèle SEIR-SW. Ce dernier leur permet de comprendre le mécanisme de propagation et de prédire l'issue finale de la maladie en modifiant les paramètres de l'épidémie et les variables d'état.

Remarque

L'épidémie à étudier doit être au sein d'une population hôte caractérisée par la durée de latence, telles que : la rougeole, la tuberculose, etc. Les résultats des simulations peuvent être analysés par une ARS et un seuil épidémique peut être calculé (\mathcal{R}_0) afin d'évaluer différentes stratégies de lutte contre l'épidémie.

3.6 Conclusion

Dans ce chapitre, nous avons détaillé le modèle de propagation SEIR-SW que nous avons proposé pour maîtriser le phénomène de propagation d'épidémie. Ce modèle permet de simuler et d'étudier l'évolution du processus de propagation de l'épidémie, il repose sur une formalisation :

- du réseau d'interaction à travers un graphe,
- de dynamique de la maladie par le processus de contact.

Ainsi, en simulant le comportement individuel et global du système, SEIR-SW permet d'estimer ses paramètres inconnus : à savoir la probabilité de transmission de la grippe saisonnière, le temps de latence et d'infection ainsi que le nombre initial de susceptibles. Il a été montré que les caractéristiques épidémiologiques suivantes ont un impact sur la croissance de l'épidémie : la durée d'incubation, la probabilité de transmission, la durée d'infection, le nombre d'infectés, le nombre de relations, la densité d'un réseau et le nombre de voisins d'un individu.

A la lumière des résultats obtenus, nous constatons qu'il est primordial de s'intéresser à la dynamique de la maladie dans des réseaux issus du monde réel. Notre approche a des implications évidentes pour l'étude de la diffusion des maladies infectieuses. D'une part SEIR-SW peut aider les professionnels de santé à améliorer la compréhension des mécanismes de diffusion au sein des réseaux SW.

D'autre part, il permet de mesurer les propriétés structurelles du réseau de contact. Notre modèle peut, éventuellement, prendre en charge les populations structurées en âge, et en métapopulations en respectant les hypothèses citées précédemment ce qui est impossible avec les modèles se basant sur EDO. Il constitue, également, une plateforme supportant des améliorations telles que l'étude d'une épidémie de plusieurs souches.

Dans la partie suivante, nous allons mettre en place un système d'information décisionnel et spatiotemporel au service de l'épidémiologie en y intégrant le modèle SEIR-SW proposé.

Deuxième partie

**Les systèmes d'Information
Décisionnels**

Chapitre 4

Généralités sur les SID

Sommaire

4.1	Introduction	60
4.2	Qu'est-ce qu'un Système d'Information Décisionnel SID ?	61
4.3	L'architecture d'un Système d'Information Décisionnel (SID)	61
4.4	Entrepôts de données (ED) « Datawarehouse »	62
4.4.1	Définitions	62
4.4.2	Les objectifs d'un ED	62
4.4.3	Les composants d'un ED	63
4.4.4	Les phases de construction d'un ED	64
4.5	Schéma multidimensionnel : le cube Olap (On-Line Analytical Processing)	67
4.5.1	Les Opérations de manipulation d'un cube	67
4.5.2	Les architectures OLAP	69
4.5.3	Outils de visualisation OLAP	70
4.5.4	Spatial OLAP (SOLAP)	72
4.6	Conclusion	73

4.1 Introduction

Dernièrement, la complexité et le volume des connaissances médicales évoluent rapidement, telles que les données médicales, socio-économiques, démographiques et environnementales, qui obligent l'ensemble des responsables en santé publique de gérer toujours plus d'informations pour suivre et surveiller la survenue d'une maladie épidémique. Ils ont besoin d'outils et de modèles pour la mise en place de systèmes décisionnels comportant des données évolutives. Une réponse à cette situation est de proposer un véritable système d'information décisionnel intégrant et exploitant à la fois toutes les données liées à la surveillance et la prévention d'épidémie.

Dans ce chapitre, nous donnons un aperçu sur les notions élémentaires des systèmes décisionnels et sur leur moteur d'exploitation OLAP. Nous commençons par présenter l'architecture et les composantes de base d'un entrepôt de données ainsi que ses étapes de construction. Ensuite, nous basculons vers l'outil d'analyse et d'exploitation OLAP avec ses différentes architectures. Enfin, nous abordons l'aspect spatial dans les systèmes décisionnels.

4.2 Qu'est-ce qu'un Système d'Information Décisionnel SID ?

Un nouveau concept est apparu dans les années 90, c'est l'informatique décisionnelle (Business Intelligence). Il connaît un essor considérable grâce au progrès réalisés en technologie de l'information et le partage d'informations via les réseaux à grande échelle.

Un Système d'Information Décisionnel (SID) est généralement défini comme étant « *un regroupement de données orientées vers certains sujets, intégrées, dépendantes du temps, non volatiles, ayant pour but d'aider les gestionnaires dans leurs prises de décision* » [Inmon, 1996]

En effet, les SID sont des outils du système d'information de gestion orientés vers la production de tableaux de bords et d'outils de pilotage. Ces systèmes visent à faciliter la prise de décision au sein d'une organisation.

4.3 L'architecture d'un Système d'Information Décisionnel (SID)

Dans la littérature, il n'existe pas une architecture standard unique pour un SID, aucun travail jusqu'à présent n'a pu concevoir une architecture idéale. En effet, nous présentons dans la Figure 4.1 une architecture typique d'un SID, elle est constituée de quatre couches logicielles, à savoir : ETL (*Extract, Transform and Load tool*), entrepôt de données, le cube OLAP et les outils de visualisation OLAP.

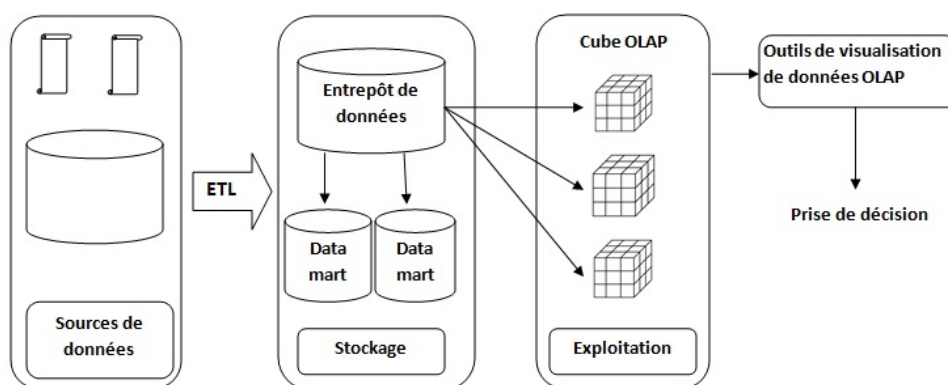


Figure 4.1 – Les principaux composants d'un SID

L'architecture détaillée d'un SID, comme indiquée dans la Figure 1, est composée des éléments suivants :

Les sources de données sont distribuées, variées, et hétérogènes. Elles peuvent être internes ou externes à l'organisme, par exemple : bases de production, rapports, bases des partenaires, etc.

Le processus ETL (Extract, Transform and Load) permet le nettoyage, l'intégration et le chargement périodiques de toutes les données au sein de l'ED nécessaires pour l'analyse.

L'entrepôt de données (datawarehouse) est le lieu de stockage centralisé, il contient les données orientées métier. Une fois ces données stockées dans le Datawarehouse, on va pouvoir créer des magasins de données appelés : **Datamarts**.

Les magasins de données (datamarts) : sont des extraits de l'ED. Les données

sont organisées de manière adéquate pour permettre des analyses rapides à des fins de prise de décision.

Le cube OLAP permet d'accéder rapidement et interactivement à des données stockées via une large variété de vues possible d'informations.

Les outils de visualisation de données permettent de visualiser les données selon des axes d'analyses. L'information est visualisée via des interfaces interactives et fonctionnelles dédiées à des décideurs souvent non informaticiens (directeurs, chefs de services, ...).

4.4 Entrepôts de données (ED) « Datawarehouse »

4.4.1 Définitions

4.4.1.1 Entrepôts de données (ED)

Inmon dans [Inmon, 1996], définit l'entrepôt de données comme étant : « *une collection de données thématiques, intégrées, historisées, et non-volatiles pour supporter le processus de prise de décision d'une organisation* ».

Thématiques (orientées sujet) : les données sont organisées par sujet ou fait.

Intégrées : les données provenant de diverses sources hétérogènes, doivent être uniformes et intégrées dans l'ED.

Historiques : l'évolution des données est essentielle pour la prise de décision, par exemple, en s'appuyant sur les évolutions passées pour prévoir les évolutions futures.

Non-volatiles : les données insérées dans l'ED ne sont jamais modifiées ou effacées c.-à-d. permanentes et ne peuvent pas être modifiées, elles sont conservées pour des analyses futures.

4.4.1.2 Les métadonnées

Sont des informations sur les données indispensables à une exploitation efficace d'un ED. Les métadonnées décrivent le schéma d'un ED et les données individuelles. Dans l'entreposage de données, les métadonnées sont classées d'après leur objet et le public auquel elles s'adressent [Peguiro et Thiery, 2005].

4.4.2 Les objectifs d'un ED

L'objectif principal des entrepôts de données est de supporter la prise de décision en permettant une meilleure exploitation des informations des systèmes opérationnels des entreprises. Néanmoins, il existe d'autres objectifs liés aux gestionnaires et aux décideurs des entreprises, citées dans [Kimball *et al.*, 2002] :

- **L'entrepôt de données doit rendre les données de l'organisation facilement accessibles** : le contenu de l'ED doit être facile à comprendre. Les données doivent être parlantes et leur signification évidente pour le simple utilisateur et pas seulement pour le développeur. Les outils d'accès à l'ED doivent être simples et faciles à utiliser.
- **L'entrepôt de données doit présenter l'information de l'organisation de manière cohérente** : les données de l'ED doivent être crédibles. Elles doivent être assemblées à partir de différents sources de l'organisation et nettoyées ainsi que leurs qualités soient contrôlées. La cohérence implique une qualité élevée, elle suppose aussi que l'on a tenu compte de toutes les données qu'elles sont complètes. Elle exige, en outre, que les définitions communes du contenu de l'ED soient disponibles pour les utilisateurs.

- **L'entrepôt de données doit être adaptable et résistant aux changements :** les modifications de l'ED doivent se faire en douceur, ce qui veut dire qu'elles ne doivent pas invalider les données existantes ou les applications. Si les données descriptives de l'ED sont modifiées, nous devons nous rendre compte convenablement de ces modifications.
- **L'entrepôt de données doit protéger notre richesse informationnelle :** L'ED doit efficacement contrôler l'accès aux informations confidentielles de l'organisation.
- **L'entrepôt de données doit être le socle sur lequel repose l'amélioration des prises de décision :** il doit contenir les données servant à étayer les décisions. En fait, ces dernières sont la valeur ajoutée de l'ED.
- **L'acceptation de l'entrepôt de données par la communauté des utilisateurs est l'une des conditions de sa réussite :** l'acceptation des utilisateurs à l'ED est avant tout liée à sa simplicité d'utilisation.

Contrairement aux systèmes de production, où les données sont structurées par processus fonctionnels, dans les ED les données sont organisées par thèmes. Généralement, l'ED est destiné aux décideurs de l'entreprise non spécialistes de l'informatique. En effet, il doit être équipé d'interfaces graphiques adaptées à ces décideurs.

4.4.3 Les composants d'un ED

Dans un ED, les données sont lues soit à partir de bases de données opérationnelles¹, soit à partir d'autres données internes ou de source de données externes par le système ETL. Ensuite, ce système nettoie et prépare les données pour le traitement, cela peut être un processus complexe. Les données sont stockées dans l'ED pour les exploiter par la suite par les utilisateurs de BI qui accèdent aux données par divers outils. L'ED stocke les bases de données et les métadonnées définissant ses objets. La Figure 4.2 présente les principaux composants d'un ED.

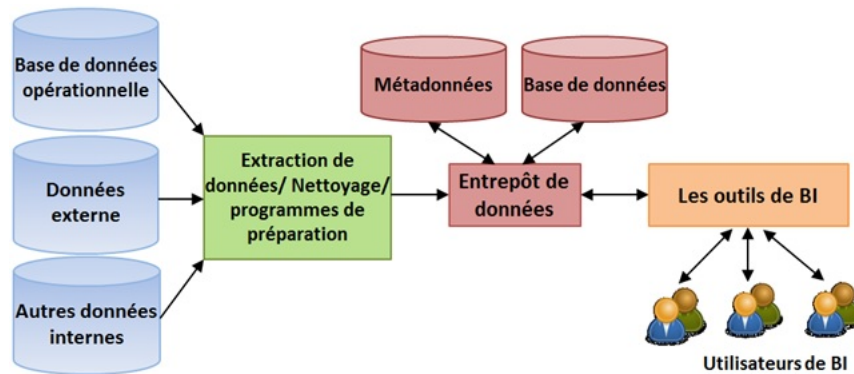


Figure 4.2 – Les principales composantes d'un entrepôt de données [Kroenke et Auer, 2011]

1. Les bases de données qui stockent les données de transaction de l'entreprise au le jour

4.4.4 Les phases de construction d'un ED

La construction d'un ED suit les phases décrites dans ce qui suit :

4.4.4.1 La modélisation d'un ED

Ralph Kimball [Kimball, 1996] a inventé et popularisé une autre technique de modélisation dédiée au décisionnel qu'il appela « *modélisation multidimensionnelle* ».

Contrairement aux modèles relationnels et leurs principes de normalisation ayant leurs limites (notamment en termes de performances d'accès), les modèles multidimensionnels offrent une solution intuitive et complète aux exigences inhérentes de l'informatique décisionnelle.

Vocabulaire associé à la modélisation multidimensionnelle

Dans la modélisation multidimensionnelle, chaque modèle se compose d'une table disposant une clé multiple² appelée *table des faits* et d'un ensemble de tables appelé *table de dimensions*.

1. **La table de faits** : la table de faits, dans un modèle multidimensionnel, stocke les **mesures** de performances résultant des événements de processus de gestion d'une organisation. Elle permet la modélisation des sujets, d'événements ou de phénomènes souhaitant les analyser par l'organisation. Elle est composée de mesures numériques généralement valorisées de manière continue [Kimball, 1996]. Chaque fait est analysé selon des axes d'analyses, nommés *dimensions*.
2. **La table de dimensions** : le sujet analysé, c'est à dire *le fait*, est analysé selon différentes perspectives. Ces perspectives correspondent à une catégorie utilisée pour caractériser les mesures d'activité analysées [Marcel, 1998], nous parlons de dimensions. Les attributs des tables de dimensions jouent un rôle crucial, ils sont textuels et discrets. Les dimensions sont organisées en hiérarchies pour permettre l'analyse des mesures à différents niveaux de détail. La Figure 4.3 représente un modèle multidimensionnel d'une vente, avec CS : Clé de Substitution³ et CE : Clés Étrangères. Exemple de dimension : dimension temps. Pratiquement, cette dimension est présente dans la majorité des entrepôts de données ayant des séries temporelles.

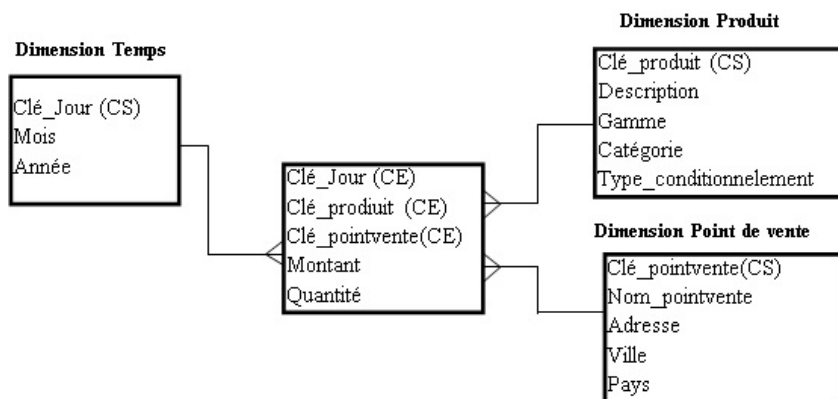


Figure 4.3 – Un exemple de modèle multidimensionnel d'une vente

2. Une clé composée de la concaténation des clés des tables de dimension

3. Clés qui servent à structurer les données et sont gérées par le système de base de données

3. **La hiérarchie de dimension** : une hiérarchie de dimension représente les paramètres d'une dimension selon leur niveau détaillé. Les paramètres sont ordonnés par une relation "*est plus fin*". Un niveau d'hiérarchie peut être inséré soit à côté de la hiérarchie déjà définie soit à l'intérieur. Par exemple pour la dimension temps, la hiérarchie est la suivante : Jour → Mois → Année.
4. **La granularité (Concept de grain de données)** : la granularité constitue l'aspect le plus important dans la conception d'un entrepôt de données. En effet, le concept de la granularité, imprègne toute l'architecture qui entoure l'environnement d'entrepôt de données. La granularité se réfère au niveau de détail auquel l'utilisateur souhaite parvenir. Le niveau le plus détaillé est le plus faible niveau de granularité. Le moins détaillé est le plus élevé niveau de granularité. Les niveaux de granularité sont décrits par un graphe, ou par un ensemble de graphes indiquant comment les membres d'un niveau sont regroupés afin de former un niveau supérieur. Généralement, chaque graphe est associé à une dimension et représente une hiérarchie. En effet, chaque niveau de la hiérarchie correspond à une agrégation des données associées au niveau inférieur.

La modélisation logique des données

A partir des faits et des dimensions, il est possible d'établir une structure de données simple qui correspond au besoin de la modélisation multidimensionnelle. Nous distinguons quatre groupes de modèles :

1. **Le modèle de données en étoile (Schéma en étoile)** : le modèle de données en étoile comme son nom l'indique, est constitué d'une table de faits centrale et plusieurs tables de dimensions dénormalisées. Chaque table de fait correspond à un fait conceptuel et inclut une clé primaire, des clés étrangères liées à des dimensions et une colonne pour chaque mesure du fait. Il existe une table de dimension pour chaque axe d'analyse avec tous les niveaux d'agrégation. Le modèle en étoile est destiné à l'utilisateur et à deux buts principaux : la simplicité et la performance en temps de réponse aux requêtes (nombre de jointure limité).

En revanche, la dénormalisation du modèle introduit de la redondance dans les données, ce qui nécessite plus d'espace de stockage [Kimball, 1996].

Les principaux avantages et inconvénients du modèle sont listés dans ce qui suit [Boussaid, 2015] :

Avantages

- Facilité de navigation ;
- Performances : nombre de jointures limité ; gestion des données creuses ;
- Gestion des agrégats⁴ ;
- Fiabilité des résultats ;

Inconvénients

- Toutes les dimensions ne concernent pas les mesures ;
- Redondances dans les dimensions ;
- Alimentation complexe.

2. **Schéma en flocons** : le modèle flocon en neige est une version normalisée du schéma en étoile. Il est composé d'une table de fait entourée par les différentes dimensions,

4. Les agrégats, ou tables d'agrégats, représentent les enregistrements récapitulatifs

qui sont décomposées en sous hiérarchies (chaque niveau est représenté dans une table différente).

Cette modélisation est plus facile à maintenir permettant d'éviter certaines redondances de données, par contre, elle est plus lente lors de l'interrogation suite à l'augmentation du nombre de jointure.

3. **Schéma mixte** : le schéma mixte ou hybride est la combinaison des deux types de schéma décrits précédemment où seulement les tables de dimension contenant le plus grand volume de données, et le plus de redondance, sont normalisées. Dans cette modélisation, certaines dimensions sont : normalisées, partiellement normalisées ou encore dénormalisées.
4. **Schéma en constellation** : plusieurs modèles en étoile partagent entre eux les mêmes dimensions. En effet, un modèle en constellation (constellation d'étoiles) comprend plusieurs faits et des dimensions communes ou non. Les dimensions de modèle sont plus ou moins normalisées.

Le choix entre la normalisation ou dénormalisation des dimensions est souvent établie en considérant deux critères : le coût de stockage et les performances en temps de réponse aux requêtes attendues [Kimball *et al.*, 2002].

4.4.4.2 L'alimentation d'ED

L'alimentation d'ED se déroule par le processus d'alimentation de données (ETL). En effet, l'outil ETL (Extract, Transform and Load) est un progiciel de type médiateur (middleware) qui permet de réaliser un passage en masse d'information d'une base de données vers une autre. Il est utilisé pour alimenter l'ED à partir de base de données opérationnelles. Ainsi, il permet l'extraction, le nettoyage et l'importation des données des différentes sources et il les charge dans un entrepôt de données (data warehouse) en temps réel. Il prend en charge différentes sources de données, par exemple : les SGBD relationnels, les fichiers XML, fichiers à formats fixes ou avec séparateurs. Plusieurs applications d'ETL existent dans la littérature, nous citons : Apatar, Pentaho Data Integration, Talend Open Studio, Oracle / Warehouse Builder, etc [El Akkaoui et Zimányi, 2009]. Ce processus à étapes multiples et automatisées présenté par la Figure 4.4 est décrit ci-dessous :



Figure 4.4 – Une présentation schématique du processus ETL

- Extraction de données des bases de données opérationnelles (ERP, SGBDR, fichiers en dur, ...);
- Transformation de ces données pour nettoyer, mettre en conformité, standardiser, documenter, corriger et dé-dupliquer;
- Chargement des données dans le système décisionnel : Datawarehouse, Datamarts, ou Cube.

En effet, un système ETL est tout système qui permet [Boussaid, 2015] :

- D’offrir un environnement de développement, des outils de gestion des opérations et de maintenance ;
- De découvrir, analyser et extraire les données à partir de sources hétérogènes ;
- De nettoyer et standardiser les données selon les règles d’affaires établies par l’entreprise ;
- De charger les données dans un entrepôt de données dans et/ou les propager vers les data-marts.

Le chargement de l’entrepôt de données à travers le processus ETL constitue une tâche difficile et qui prend beaucoup de temps pour sa réalisation avec un coût considérable dans les systèmes humains et les ressources financières [Bakari *et al.*, 2012]. En effet, c’est un processus complexe qui nécessite un environnement de développement, des outils de gestion des opérations et de maintenance.

4.4.4.3 L’exploitation de l’ED

C’est la phase d’analyse en ligne chargée de la création du cube à partir de l’ED. Cette étape contient les outils OLAP (On Line Analysis Processing) pour l’Analyse des données, de fouille de données (datamining), de restitution des données sous différentes formes (graphiques ou tableaux) et d’administration.

4.5 Schéma multidimensionnel : le cube Olap (On-Line Analytical Processing)

R. Kimball [Kimball *et al.*, 2005] définit le concept OLAP comme « *Une activité globale de requêtage et de présentation de données textuelles et numériques contenues dans l’entrepôt de données ; Style d’interrogation spécifiquement dimensionnel* ». En effet, l’OLAP est un ensemble de règles permettant de réaliser des analyses de données multidimensionnelles de manière performante, notamment en créant des hyper cubes. Les traitements mis en œuvre par l’OLAP prennent en compte l’aspect semi-automatique du processus décisionnel en proposant des manipulations interactives, tout en répondant aux besoins spécifiques en analyse d’information [Marcel, 1998]. Tous les outils pouvant synthétiser, explorer, confirmer, expliquer, prédire les données sont des outils de restitution.

Les auteurs dans [Agrawal *et al.*, 1997] décrivent des traitements aptes à consolider, visualiser et synthétiser des informations reflétant la vision de l’analyste. Cette vision correspond à une structuration des données selon plusieurs dimensions, pouvant représenter des notions très variées telles que le temps ou la localisation géographique, mais aussi la valeur d’un chiffre d’affaires ou le code identifiant des produits.

Nous parlons alors de la modélisation et de traitements multidimensionnels des données. Contrairement à l’OLTP (processus de transaction en ligne) qui s’applique aux bases de données à usage relationnel, le modèle OLAP s’applique aux bases de données à usage décisionnel.

4.5.1 Les Opérations de manipulation d’un cube

Dans un modèle dimensionnel, on cherche à représenter les données dans un cube (ou hypercube). En effet, un cube est un espace d’analyse à n dimensions (par exemple : temps, lieu, personne, etc.) La navigation dans le cube permet à l’analyste de visualiser

les informations contenues dans le cube et de passer d'un niveau d'agrégat à un autre afin de connaître le détail des données qui ont initialement servi à la constituer.

Les opérations appliquées sur le cube sont algébriques c.-à-d. que le résultat généré est un autre cube, deux types d'opérateurs existent [Gray *et al.*, 1997] :

1. **Les opérations liées à la granularité** : Les opérations appliquées sur la granularité caractérisent la hiérarchie de navigation entre les différents niveaux de détail d'une dimension. Nous citons les deux opérations suivantes :
 - (a) Roll-up ou forage vers le haut : consiste à représenter des données à un niveau de granularité supérieur conformément à la hiérarchie définie sur la dimension ;
 - (b) Drill-down ou forage vers le bas : Drill-down ou forage vers le bas : consiste à représenter les données du cube à un niveau de granularité inférieur, donc sous une forme plus détaillée.

Les modèles multidimensionnels servent aussi de supporter d'autres règles d'algèbres de manipulation OLAP, nous détaillons, dans ce qui suit, quelques opérateurs sur le cube :

2. **Slicing et Dicing (restriction)** : La spécification d'une tranche de cube « Slice » consiste à exprimer une restriction sur une des données de l'un des axes d'analyse. Une opération de type « Dice » consiste à générer une nouvelle table, qui exprime une restriction sur les données d'un indicateur d'analyse.
3. **Pivot (Rotation)** : les opérations de Pivot permettent une réorientation d'une analyse pour fournir une vue alternative des données. L'opération la plus courante consiste à changer l'axe d'analyse (Pivot de dimension).
4. **Drill through (forage)** : consiste à visualiser une même information sous l'angle de plusieurs dimensions.
5. **Le calcul d'agrégats** : lors d'une analyse OLAP, le processus d'analyse agrège les données en fonction des niveaux de granularité sélectionnés à travers une fonction d'agrégation (somme, moyenne, maximum, etc.). Les opérations de forage font un usage intensif de ces fonctions d'agrégation Teste09. Il est parfois nécessaire d'effectuer un regroupement plus large pour connaître d'autres valeurs.
6. **Drill-across (forage latéral)** : cette opération consiste à utiliser simultanément les agrégats de plusieurs modèles en environnement ROLAP seulement.

D'autres opérations de construction, produisant une table multidimensionnelle (TM) à partir d'une base de données multidimensionnelles, et un ensemble d'opérateurs fondamentaux [Ben Messaoud, 2006] :

- L'imbrication (*nest*) : permet d'intégrer, dans les dimensions d'une TM, les données provenant d'une ou de plusieurs dimensions.
- La sélection (*select*) : restreint l'ensemble des valeurs affichées. La restriction peut bien porter sur les valeurs des attributs que sur les mesures du fait.
- L'opérateur de classement (*switch*) : intervertit deux valeurs d'un attribut d'une dimension pour permettre l'ordonnancement des valeurs affichées.
- L'opération de conversion d'un paramètre en mesure (*push*) : transforme un paramètre afin qu'il apparaisse dans la TM comme une mesure. Les valeurs du paramètre sont alors affichées dans les cellules.
- La conversion d'une mesure en paramètre (*pull*) : transforme une mesure en paramètre de la TM. Les valeurs des mesures sont affichées au niveau des en-têtes en ligne ou en colonne. Les opérations d'ajout (*addm*) et de suppression (*delm*) de mesures permettent de modifier l'ensemble des mesures analysées.

La plupart des travaux sur les manipulations OLAP ont étendu les opérateurs de l'algèbre relationnelle pour le modèle en cube. D'autres travaux ont proposé des opérateurs pour spécifier et manipuler un cube prenant en compte les structures multidimensionnelles.

4.5.2 Les architectures OLAP

Dans la littérature des systèmes OLAP, il existe différentes implémentations, nous citons, dans les sections suivantes, les trois implémentations les plus connues à savoir : (a) ROLAP (Relational OLAP) qui utilise une base de données relationnelle ; (b) MOLAP (Multidimensional OLAP) qui repose sur une base de données multidimensionnelle ; et (c) HOLAP (Hybrid OLAP) qui réunit les deux technologies.

4.5.2.1 L'approche ROLAP (Relational OLAP)

Dans une architecture ROLAP illustrée par la Figure 4.5, les données détaillées et agrégées sont stockées dans une base de données relationnelle. Cette dernière est structurée particulièrement selon un : schéma en étoile ou en flocon de neige. Les requêtes OLAP sont traduites en SQL.

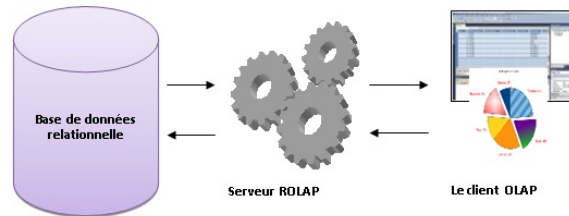


Figure 4.5 – Architecture d'un système ROLAP

Les résultats des requêtes SQL retournés par le SGBR sont ensuite traduits par le moteur ROLAP, dans un langage approprié, pour être visualisés au niveau du client OLAP. Autrement dit, le serveur extrait les données par les requêtes SQL et interprète les données selon une vue multidimensionnelle pour les présenter au module client.

4.5.2.2 L'architecture MOLAP (Multidimensional OLAP)

Dans cette architecture présentée par la Figure 4.6, les données détaillées et agrégées sont stockées dans une base de données multidimensionnelle (hypercube). La modélisation se fait directement à partir du cube. Le serveur OLAP extrait les données de l'hypercube et les présente directement au module client. Ces cubes sont implémentés comme des matrices à plusieurs dimensions.

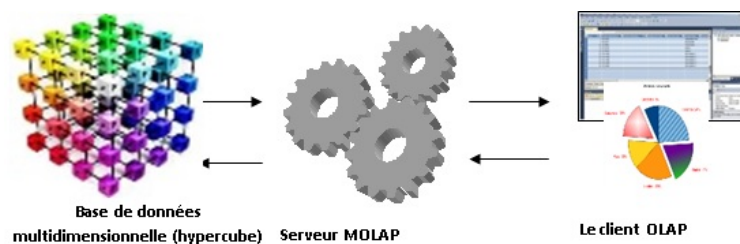


Figure 4.6 – Architecture d'un système MOLAP

4.5.2.3 L'approche HOLAP (Hybride OLAP)

Une architecture qui combine les deux architectures citées au-dessus : MOLAP + ROLAP (Figure 4.7). Les données sont stockées dans des tables relationnelles et les données agrégées sont stockées dans des cubes. Les requêtes vont chercher les données dans les tables et les cubes.

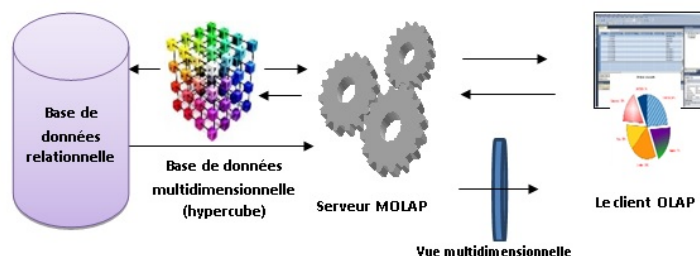


Figure 4.7 – Architecture d'un système HOLAP

La Table 4.1 résume la différence entre les trois stratégies de point de vue stockage des données de base, chargement des données, et des performances des requêtes.

Table 4.1 – Comparaison des systèmes ROLAP, MOLAP et HOLAP [Boulil, 2012]

Caractéristiques	ROLAP	MOLAP	HOLAP
Volumes de données	gros volumes	petits volumes	volumes moyens
Chargement de données	les plus efficaces	les moins efficaces	modérés
Requêtage	les plus lents	les plus rapides	modérés
Passage à l'échelle	les meilleurs	les plus faibles	modérés

4.5.3 Outils de visualisation OLAP

La restitution (la visualisation) est la partie la plus importante pour l'utilisateur. Les outils de visualisation représentent l'interface à travers laquelle les utilisateurs interagissent avec le système d'information décisionnel. Ces outils facilitent l'accès aux fonctionnalités d'interrogation de l'application ainsi que l'affichage des résultats.

De plus, ils assurent des facilités de manipulation. En effet, l'objectif principal du système d'information décisionnel est de fournir les résultats aux requêtes des utilisateurs, dans un temps de réponse raisonnable, et sans connaissance particulière dans le domaine de l'informatique. À travers la technique OLAP, le SID a besoin d'assurer que les vues soient facilement organisées et que les interfaces soient conviviales et intuitives.

Dans ce qui suit, nous décrivons quelques outils de visualisation des données et de navigations :

4.5.3.1 Tableau de bord

Parmi les outils de visualisation les plus usuels dans le domaine de business intelligence (BI), les tableaux de bord de BI (BI dashboard).

Few dans [Few, 2004] a défini le tableau de bord comme étant : “*Un affichage visuel de l'information la plus importante et nécessaire pour atteindre un ou plusieurs objectifs,*

l'information est consolidée et disposée sur un seul écran afin qu'elle puisse être contrôlée à un coup d'œil".

En effet, les tableaux de bord permettent à l'utilisateur d'obtenir un aperçu des données les plus importantes en un seul coup d'œil. Ils peuvent fournir une solution efficace pour une quantité de données énorme et permettent une synthèse rapide des actions en cours. Les tableaux de bord peuvent être classés en fonction de leur rôle. Dans [Few, 2007], l'auteur a identifié les types de tableaux de bord suivants : stratégiques, analytiques, opérationnels, ou informationnels.

4.5.3.2 Reporting

Le "Reporting" est une famille d'outils de BI destinés à assurer la réalisation, la publication et la diffusion de rapports d'activité selon un format prédéterminé. Ils sont essentiellement destinés à faciliter la communication de résultats chiffrés ou d'un suivi d'avancement. La technologie "Reporting" dans BI contient beaucoup plus de fonctionnalités que la répartition des informations. Le BI distingue trois principaux types d'outils de reporting, à savoir : *outils de reporting de production, rédacteurs de rapport de bureau et outils de requêtes gérées*. La Figure 4.8 illustre les principaux composants d'un système de reporting.

Les données provenant de sources de données différentes sont lues et traitées. Comme indiqué sur le schéma, les systèmes de reporting peuvent obtenir des données à partir des bases de données opérationnelles, des entrepôts de données et de magasin de données. Certaines données sont générées au sein de l'organisation, d'autres sont obtenues à partir de sources publiques, et des données peuvent être achetées auprès des services de données.

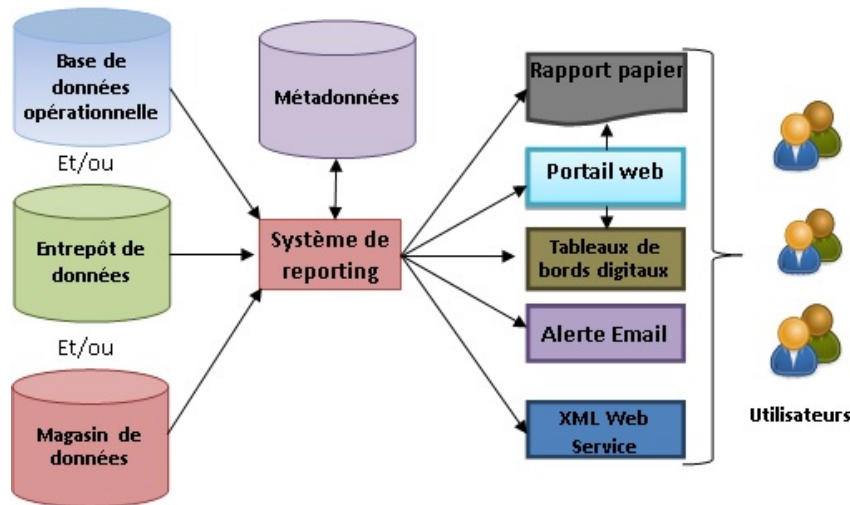


Figure 4.8 – Les composants d'un système de reporting [Kroenke et Auer, 2011]

Le système de reporting maintient une base de données des rapports de métadonnées. Les métadonnées décrivent des rapports, des utilisateurs, des groupes, des rôles, des événements, et d'autres entités impliquées dans l'activité de reporting. Les rapports peuvent être préparés sous forme de : rapport papier, portail web, tableaux de bord digital, email/alerte, XML web service.

4.5.4 Spatial OLAP (SOLAP)

Le SOLAP représente une solution efficace pour l'analyse spatio-multidimensionnelle de données spatiales. En effet, avant d'aborder plus en détail le concept, il paraît incontournable de préciser la nature des problèmes décisionnels à référence spatiale.

Les auteurs dans [Pusceddu et Chakhar, 2005], ont cité quatre problèmes majeurs liés à l'aide à la décision spatiale :

Les problèmes décisionnels spatiaux sont (i) de nature multidimensionnelle, interdisciplinaires et semi-structurés, (ii) impliquant plusieurs personnes et institutions, ayant généralement des préférences et des objectifs conflictuels, (iii) nécessitant la définition de plusieurs critères contradictoires et dont l'importance n'est pas la même, et (iv) demandent une quantité considérable des données quantitatives et qualitatives.

Cependant, l'OLAP est une approche pour répondre rapidement à des requêtes multidimensionnelles. Il est considéré comme un outil d'aide à la décision souvent utilisé pour fournir un accès, d'une manière efficace et intuitif, à un entrepôt de données. Toutefois, cet outil n'est pas tellement robuste pour analyser les données spatiotemporelles. De plus, il ne dispose pas du support cartographique.

Les Systèmes d'Information Géographique (SIG) sont bien adaptés pour l'analyse de données spatiotemporelles, par contre, ils ne sont pas capables d'effectuer une analyse multidimensionnelle. Par conséquent, une nouvelle approche qui combine les fonctionnalités des deux technologies OLAP et SIG est indispensable. Son objectif principal, est de disposer d'un outil d'aide à la décision mieux adapté à l'exploration et l'analyse spatio-temporelle des données. Celui-ci est appelé : Spatial OLAP, ou SOLAP [Bédard *et al.*, 1997].

4.5.4.1 Définition de Spatial OLAP « SOLAP »

Yvan Bédard [Bédard *et al.*, 1997] définit SOLAP comme « *une plate-forme visuelle supportant l'exploration et l'analyse spatio-temporelle faciles et rapides des données selon une approche multidimensionnelle, à plusieurs niveaux d'agrégation, via un affichage cartographique, tabulaire ou un diagramme statistique* ».

En effet, SOLAP désigne à la fois le concept et la technologie permettant l'intégration des outils d'aide à la décision OLAP et des capacités de représentation cartographique des SIG. L'interface utilisateur d'un outil Spatial OLAP fournit des capacités uniques pour explorer les données spatiales d'une manière intuitive et interactive.

La capacité de lier ou de synchroniser plusieurs vues de la même information dans un contexte d'exploration interactive des données apporte de nouvelles possibilités de bénéficier de la recherche actuelle dans la géovisualisation.

4.5.4.2 Les outils SOLAP

Dans cette section, nous allons résumer les outils actuellement disponibles pour SOLAP. Ces outils peuvent être divisés en trois catégories différentes [Bédard *et al.*, 2005] :

- OLAP dominant (Business Objects, Cognos, Knosys) : la composante spatiale est accessoire ;
- SIG dominant : la composante décisionnelle est accessoire ;
- SOLAP intégré : les deux composantes décisionnelle et spatiale sont fortement couplées, elles font appel autant aux fonctions OLAP que SIG (Geo cube, SOVAT).

4.6 Conclusion

L'approche théorique des SID a été présentée dans ce chapitre en introduisant tout d'abord les concepts de base. Les quatre composants sur lesquels un SID repose ont été ensuite successivement abordés : ETL, entrepôt de données, le cube OLAP et les outils de visualisation OLAP. La description des principales fonctionnalités de différentes parties d'un SID a permis d'en mesurer certaines limites dont le besoin est de combiner au sein d'une même plateforme les différents outils. Nous avons, également, exposé toutes les opérations multidimensionnelles proposées sur le cube OLAP. Comme nous avons cité quelques problèmes décisionnels à référence spatiale, et par conséquent, l'utilité du concept SOLAP.

Le chapitre suivant s'appuie largement sur des travaux de recherche réalisés dans le cadre de la surveillance d'épidémie.

Chapitre 5

Etat de l'art sur les Systèmes d'Information Décisionnels

Sommaire

5.1	Introduction	74
5.2	Les Systèmes de surveillance d'épidémies/ Les systèmes d'épidémiosurveillance	75
5.2.1	Des systèmes à l'échelle nationale et mondiale	75
5.2.2	Travaux de recherche connexes	76
5.3	Les systèmes de surveillance à référence spatiale	77
5.3.1	Le SIG comme outil d'aide à la décision	77
5.3.2	Le SIG et les modèles mathématiques	79
5.4	Les SID épidémiologiques	80
5.4.1	Les travaux traitant les entrepôts de données médicales	80
5.4.2	Les travaux traitant les entrepôts de données médicales et l'OLAP	82
5.4.3	Les travaux traitant les entrepôts de données et les réseaux sociaux	83
5.5	Autres stratégies de lutte contre les épidémies	84
5.5.1	Surveillance active ou passive	84
5.5.2	Surveillance exhaustive	84
5.6	Conclusion	85

5.1 Introduction

Le concept de surveillance dans le domaine de la santé publique a connu une évolution rapide et importante ces dernières années, cette évolution a touché deux niveaux : administratifs (acteurs politiques, directeurs, etc.) et/ou médicaux (médecins, épidémiologistes, cliniciens, etc.). Afin de mieux cerner la portée de l'application des approches d'aide à la décision aux problématiques de la santé publique et la nécessité d'améliorer les approches classiques, il convient de faire un bref rappel de l'évolution de ce champ d'études.

Ce chapitre a pour objectif de présenter une revue des travaux existants dans le domaine de la surveillance d'épidémie en général et sur les travaux liés à notre problématique en particulier. Nous classons ces travaux selon leurs approches et leurs objectifs en quatre classes : les systèmes de surveillance, les systèmes utilisant l'information géographique, les systèmes utilisant l'information décisionnelle et les stratégies de lutte.

5.2 Les Systèmes de surveillance d'épidémies/ Les systèmes d'épidémiosurveillance

La surveillance épidémiologique appelée aussi épidémiosurveillance peut être définie comme « *une méthode d'observation fondée sur des enregistrements en continu permettant de suivre l'état de santé ou les facteurs de risques d'une population définie et de déceler l'apparition de processus pathologiques et d'en étudier le développement dans le temps et l'espace en vue de l'adoption de mesures appropriées de lutte* » [Dufour *et al.*, 2006]

Cette définition montre, clairement, que la surveillance épidémiologique a pour finalité de détecter l'apparition de maladies et mettre en place les mesures de contrôle. En effet, les systèmes de surveillance épidémiologique sont des outils d'aide à la décision dans le domaine de la prévention et du contrôle des maladies épidémiologiques [Dufour *et al.*, 2006].

Dans la littérature de la surveillance épidémiologique, il existe plusieurs systèmes d'épidémiosurveillance :

- soit à l'échelle nationale : les réseaux sentinelles (France, Belgique, Etats Unis, le Royaume Uni, la Suède, etc.), l'Institut National de Veille Sanitaire (INVS) ;
- soit à l'échelle mondiale : le système EWARN (Early Warning Alert and Response Network), l'OMS (Organisation Mondiale de Santé), etc.

Chaque système fonctionne selon ses objectifs, à savoir : le calcul des statistiques médicales, la surveillance de l'évolution de la maladie dans le temps et parfois dans l'espace, la publication et l'affichage périodique des statistiques, etc.

5.2.1 Des systèmes à l'échelle nationale et mondiale

L'Organisation Mondiale de la Santé (OMS) a développé, à l'échelle mondiale, un plan stratégique de lutte contre les infections émergentes fondé sur la surveillance, d'alerte et de la réponse, de la recherche appliquée, de la prévention et le contrôle et le renforcement des structures de santé publique.

En France, le système de surveillance épidémiologique des maladies infectieuses à déclaration obligatoire a été mis en place par L'Institut national de Veille Sanitaire (InVS ¹) en 2002. Ce système de surveillance concerne près de 30 Maladies A Déclaration Obligatoire (MADO). Les activités de l'InVS sont transversales, tant pour ce qui concerne les champs thématiques abordés (maladies infectieuses, maladies chroniques, santé au travail, santé environnementale, etc.) que pour les activités menées (veille, vigilance, surveillance, alerte, etc.).

Un autre système a été élaboré pour le suivi et la surveillance des épidémies avec la collaboration mixte d'Inserm, d'Université Pierre et Marie Curie et d'InVS, c'est le réseau sentinelles Français, cette collaboration offre une présentation cartographique de la distribution d'une épidémie dans une région ou dans le pays permettant le suivi et la surveillance en temps réel.

La méthode de surveillance adoptée par ce système se base principalement sur : la surveillance continue et l'enrichissement en continu de la base de données sur l'activité de consultation des médecins généralistes libéraux en France métropolitaine, le veille sanitaire, les enquêtes ponctuelles et les études cliniques.

1. <http://www.invs.sante.fr/>

A l'échelle nationale, la mise en place d'un réseau intranet dénommé "Santé Algérie ²" vise avant tout à offrir des moyens modernes de communication, d'accès et d'échange d'information au profit de l'ensemble des personnels et opérateurs de la santé. Il permet, d'une part, à tous les acteurs de la santé d'accéder à l'information et d'autre part, aux différentes structures de santé de participer à l'organisation, la collecte et l'alimentation en information des banques et bases de données dans tous les domaines de la santé. Il repose, essentiellement, sur une architecture Client/Serveur.

Aux États Unis, il existe le système SEER (Surveillance Epidemiology and End Results), un service de l'Institut National du Cancer des Etats-Unis (NCI), ce système présente une source fiable de renseignements sur l'incidence du cancer et la survie aux États-Unis. Les registres de ce programme recueillent systématiquement des données sur : les caractéristiques démographiques des patients, le site de la tumeur primaire, la morphologie de la tumeur et le stade de diagnostic, premier cours de traitement, et le suivi du statut vital.

Au Canada, le système EWARN ³ a pour but d'identifier les maladies pouvant donner lieu à des épidémies (diarrhée aqueuse aiguë, diarrhée sanglante, dengue, fièvre d'origine inconnue, ictère, rougeole, méningite et paludisme) ainsi que les infections respiratoires aiguës et le tétanos. Ce système a été complété par un système d'alerte immédiate basé sur des appels téléphoniques quotidiens, des messages ou des courriers électroniques notifiant des cas suspects de certaines maladies.

D'autres systèmes ont été développés tel que Lightweight Epidemiological Advanced Detection and Emergency Response System (LEADERS), ou Early Warning Outbreak and Response System (EWORS), ces systèmes sont déployés dans des structures médicales civiles du Sud-Est asiatique [Culpepper et Kelley, 2002].

Le système britannique Prototype for Remote Illness and Symptom Monitor/Realtime Medical Surveillance (PRISM/RMS) repose sur le recueil par un auxiliaire de santé d'un ensemble de 33 symptômes choisis essentiellement pour leurs liens avec les risques biologiques ou chimiques. Les données sont expédiées quotidiennement par satellite à un centre d'analyse en Grande-Bretagne [MacDonald *et al.*, 2004].

Chaque système dans chaque pays est pratiquement différent des autres et chaque système à ses objectifs, sa méthode de travail et ses outils de développement.

Dans le cadre de la mise en place des systèmes et des stratégies de contrôle dédiés à la surveillance d'épidémie, il existe plusieurs travaux pour le développement et la validation d'outils de détections et de prévision des épidémies. Dans ce qui suit, nous présentons les travaux les plus célèbres.

Dans le cadre de la mise en place des systèmes et des stratégies de contrôle dédiés à la surveillance d'épidémie, il existe plusieurs travaux pour le développement et la validation d'outils de détections et de prévision des épidémies. Dans ce qui suit, nous présentons les travaux les plus célèbres.

5.2.2 Travaux de recherche connexes

Turbelin et al, [Turbelin et Boëlle, 2010] ont développé une plateforme Java qui s'exécute en tant que client sur l'ordinateur local. Cette plateforme permet de signaler les cas à un serveur central et de fournir une rétroaction aux serveurs locaux. Une meilleure intégration dans le processus de travail a été signalée, en particulier lorsque le logiciel a été utilisé au moment de la consultation.

2. <http://www.sante.dz/>

3. Système de surveillance/alerte précoce et action

Les systèmes cités précédemment sont des systèmes de surveillances civils, ces systèmes partagent presque les mêmes objectifs et les mêmes conditions d'étude (populations étudiée, régions d'étude, distribution démographique, facteurs de risques, etc.). À côté de cette catégorie de systèmes, il existe une autre catégorie destinée à l'armée. Nous présentons dans ce qui suit un bref panorama des cas d'application :

Le système EpiNATO est le modèle d'une surveillance basée sur le recueil de 25 catégories regroupant les maladies par organes et les blessures par grandes causes [Wilson *et al.*, 2004]. Dans le même contexte, le système Early Notification of Community-Based Epidemics (ESSENCE) utilise les données de consultation de plus de 300 formations de soins militaires aux Etats-Unis.

Dans [Meynard *et al.*, 2004], les auteurs ont montré, de leurs points de vue, que la surveillance épidémiologique en temps réel consiste en une stratégie privilégiant la collecte de données préliminaires (le plus souvent avant qu'un diagnostic formel soit établi), à les transférer immédiatement, à les analyser automatiquement et à fournir spontanément des tableaux de bords sanitaires aux décideurs de tous niveaux.

La comparaison entre les systèmes de surveillance syndromique civils [Yan *et al.*, 2006] et militaires [Meynard *et al.*, 2004] montre de nombreuses différences fondamentales. La surveillance civile est caractérisée par la participation de nombreuses institutions médicales, utilisant différents systèmes d'information et couvrant des populations d'effectifs importants et assez stables.

Un des problèmes principaux de la surveillance civile est d'intégrer de nombreuses structures médicales utilisant des systèmes d'information différents. Cela a stimulé le besoin de standards de communication, de solutions d'interopérabilités, mais sans toutefois subir la pression de la permanence de service [Meynard *et al.*, 2008]. Cette liste de travaux n'est certainement pas exhaustive, il existe pas mal de systèmes d'épidémiosurveillance régional et mondial. Ces systèmes ont connu un succès, ils sont valables et praticables. Cependant, ils présentent quelques lacunes :

- Leurs faiblesses de ne pas supporter l'analyse spatio-temporelle de données multidimensionnelles ;
- Leurs faiblesses de ne pas identifier des facteurs favorisant la transmission des maladies infectieuses ainsi l'analyse des facteurs socio-économique de la population.

Leurs rôles s'inscrivent principalement dans le suivi, la veille et la surveillance, soit dans le temps par la mise en place de systèmes de collecte et de recueil de données en temps réel, soit par un affichage des régions à risque par l'utilisation de SIG.

5.3 Les systèmes de surveillance à référence spatiale

L'application de l'analyse spatiale et de la cartographie en épidémiologie est connue depuis longtemps, mais leur utilisation en santé publique reste limitée suite au manque de données à référence spatiale. Les cartes géographiques sont soit établies à la main soit produites dans des instituts de recherche utilisant un matériel et un logiciel SIG coûteux.

L'analyse de la répartition et de la dynamique spatiale de données sur une maladie est connue par : **l'épidémiologie spatiale** .

5.3.1 Le SIG comme outil d'aide à la décision

Dans une approche fondée sur la surveillance des maladies, les données utilisées peuvent être : *le nombre de cas, la préférence, l'incidence ou la prévalence*, ayant l'avan-

tage de tenir compte de la taille de la population totale et de l'hétérogénéité de sa distribution spatiale.

Les progrès récents des technologies d'information géographique et de cartographie ont permis aux administrateurs de la santé publique d'améliorer leur potentiel de planification, d'analyse et de surveillance. La localisation des données épidémiologiques est l'un des points critiques du recueil des données. Il est généralement très difficile de connaître précisément le lieu d'exposition à cause, entre autres, de la mobilité des individus.

Dans la table 5.1, nous présentons une synthèse des travaux qui traitent les différentes problématiques de la propagation d'épidémie par l'utilisation du SIG ainsi que le positionnement de nos contributions par rapport à ces travaux de recherche :

Table 5.1 – Synthèse des applications des SIG en santé publique

Références	Contributions du travail	Approches et outils	Nos contributions
[Loth <i>et al.</i> , 2010], [Hanafi-Bojd <i>et al.</i> , 2012], [Naqvi <i>et al.</i> , 2015]	La détermination de la répartition géographique et les variations de prévalence et d'incidence des maladies	Le SIG et la statistique spatiale	Oui
[Omumbo <i>et al.</i> , 2005], [Loth <i>et al.</i> , 2010], [Khormi et Kumar, 2011]	L'analyse des tendances spatiales et longitudinales.	Le SIG et analyse spatiale	Non
[Omumbo <i>et al.</i> , 2005], [Ben Alaya <i>et al.</i> , 2008], [Loth <i>et al.</i> , 2010], [Hanafi-Bojd <i>et al.</i> , 2012]	La cartographie des populations à risque.	Le SIG et de la modélisation spatiale	Oui
[Tran <i>et al.</i> , 2004], [Omumbo <i>et al.</i> , 2005], [Loth <i>et al.</i> , 2010], [Khormi et Kumar, 2011], [Hanafi-Bojd <i>et al.</i> , 2012]	La détermination des facteurs de risque.	Le SIG et l'analyse spatiale et typologique	Oui
[Hanafi-Bojd <i>et al.</i> , 2012]	La planification et cibler des interventions	Le SIG (ArcGIS) et la combinaison entre les facteurs environnementaux avec une évaluation multicritères pondérées.	Non
[Khormi et Kumar, 2011]	La prévention des épidémies	Le SIG, le modèle de régression pondérée	Oui
[Li et Mackaness, 2014]	Système d'aide à la décision destiné aux décideurs de la santé publique	Le SIG couplé a une approche multi-agent pour dissimuler la complexité technique vis-à-vis les utilisateurs inexpérimentés.	Oui

Les applications des SIG dans la surveillance d'épidémie ne se limitent pas aux travaux cités ci-dessus. En effet, dans [Gilbert et Pfeiffer, 2012] d'autres travaux de recherche liés à l'épidémiologie spatiale et la grippe aviaire sont abordés. Dans la même optique, les auteurs dans [Carroll *et al.*, 2014] ont détaillé les travaux de surveillance d'épidémie par utilisation des SIG, l'analyse par les réseaux sociaux et d'autres approches.

5.3.2 Le SIG et les modèles mathématiques

L'application du SIG ne se limite pas à l'utilisation des données épidémiologiques, écologiques et spatiotemporelles pour lesquelles l'évaluation et l'analyse des zones et des facteurs à risque épidémique sont primordiales. Il trouve aussi son intérêt dans son couplage avec des modèles statistiques et mathématiques d'épidémie. Malheureusement, peu de travaux ont été consacrés à ce couplage :

Dans le travail de thèse de [Simoes, 2007], l'auteur a proposé un modèle à base d'agent basé, à la fois, sur les automates cellulaires, le réseau aléatoire et le SIG. Le modèle mathématique opté ici était le modèle SEIR. L'objectif de ce travail, est la mise en œuvre d'une plateforme dédiée à la surveillance épidémique pour évaluer certaines stratégies de vaccination spatiales afin d'aider les responsables à définir des politiques de santé.

Dans un travail similaire, dans [Basileu *et al.*, 2010], l'auteur a proposé une plateforme basée sur les réseaux aléatoires, le modèle SEIR, le SIG et les systèmes multi agents afin de représenter les comportements des individus. La plateforme réalisée à des fins d'aide à la décision spatiotemporelle.

Ramírez-Ramírez *et al.*, [Ramírez-Ramírez *et al.*, 2013] ont développé un outil de simulation SIMID (SIMulation of Infectious Diseases) pour la propagation des maladies infectieuses, permettant la visualisation spatio-temporelle de la dynamique des épidémies de grippe. L'outil proposé est basé sur l'utilisation des graphes aléatoires, des SIG et deux modèles mathématiques. Il permet aux professionnels de la santé et aux décideurs une visualisation plus efficace ainsi qu'une évaluation de la dynamique d'épidémie infectieuse dans le temps et dans l'espace.

Dans [Younsi *et al.*, 2014b], nous avons proposé un système décisionnel d'épidémiosurveillance EPI-SDSS, ce système est basé principalement sur le modèle d'épidémie SIR, le modèle réseau social SW et le SIG afin de visualiser les foyers épidémiques de la grippe. EPI-SDSS permet l'affichage des situations épidémiques sous forme de courbes ou de cartes géographiques visualisant la distribution des régions à risque épidémique. Le modèle suggéré par les auteurs est illustré par la Figure 5.1.

Dans une version améliorée d'EPI-SDSS publiée dans [Younsi *et al.*, 2015b], nous avons :

- abordé le modèle à compartiment SEIR au lieu du modèle SIR
- effectué une analyse spatiotemporelle du phénomène de la propagation.

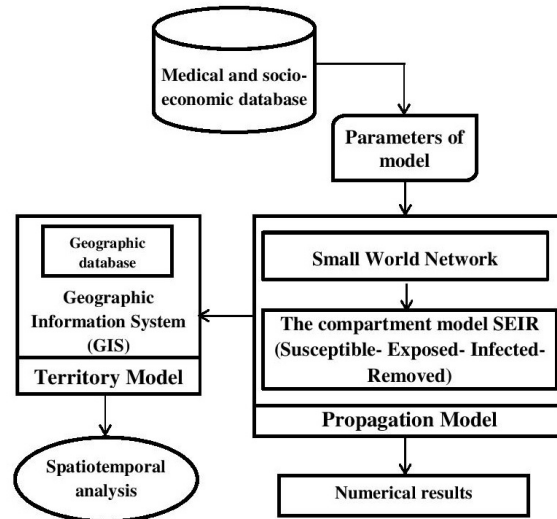


Figure 5.1 – EPI-SDSS : une vue d'ensemble [Younsi *et al.*, 2014b]

Plusieurs travaux ont été consacrés à l'étude du problème de santé publique dans plusieurs contextes : modélisation, conception, stockage, et visualisation comme nous l'avons évoqué ci-dessus. Toutefois, le SIG reste toujours limité dans le traitement des données volumineuses et dans l'analyse spatio-temporelle des données multidimensionnelles.

Dans cette optique, plusieurs chercheurs ont proposé l'utilisation des Entrepôt de données et le couplage des SIG avec l'OnLine Analytical Processing (OLAP) pour améliorer les capacités d'analyse à des fins de décision. Cette catégorie de travaux sera discutée dans la section suivante.

5.4 Les SID épidémiologiques

Les SID sont considérés comme des systèmes d'aide à la décision, ils sont actuellement la meilleure solution permettant l'intégration, le stockage, l'analyse et la restitution d'une importante quantité de données hétérogènes. Dans ce qui suit, nous citons les différents travaux portant sur l'utilisation des ED dans le domaine médical en général et dans le domaine d'épidémiologie en particulier, nous distinguons trois groupes de travaux décrits dans les sections suivantes :

1. les travaux traitant les entrepôts de données médicales,
2. les travaux traitant les entrepôts de données médicales et l'OLAP
3. les travaux traitant les entrepôts de données médicales et les réseaux sociaux,

5.4.1 Les travaux traitant les entrepôts de données médicales

La littérature des ED médicaux est vaste, nous présentons dans cette section les travaux les plus pertinents :

En termes de conception des ED médicales, on retrouve le travail de [Choquet *et al.*, 2008]. Dans ce travail, les auteurs comparent les différentes technologies développées afin d'établir une méthodologie pour choisir l'approche de la mise en œuvre d'un ED en santé.

Un méta-modèle multidimensionnel pour les données médicales a été proposé dans [Midouni *et al.*, 2009]. Les auteurs ont réalisé un outil d'aide à la conception d'un ED

médicales en généralisant le modèle cardiovasculaire du projet MAP⁴. Les auteurs ont modélisé le magasin de données le plus complexe du projet MAP, le magasin cardiovasculaire. Ensuite, ils ont proposé un méta-modèle en généralisant le modèle multidimensionnel du module cardiovasculaire. Enfin, ils ont développé ce méta-modèle en implémentant un prototype Générateur d'Entrepôt de Données Médicales (GEDM).

Dans un travail plus récent, Bouba et al., [Bouba *et al.*, 2013] ont réalisé une conception d'un ED médicales intégrant plusieurs systèmes étudiés (environnement, climat, hydrologie, épidémiologie, socio-économie, etc.) à différentes échelles (temps, saisons, acteurs). Le modèle multidimensionnel proposé est représenté par la Figure 5.2.

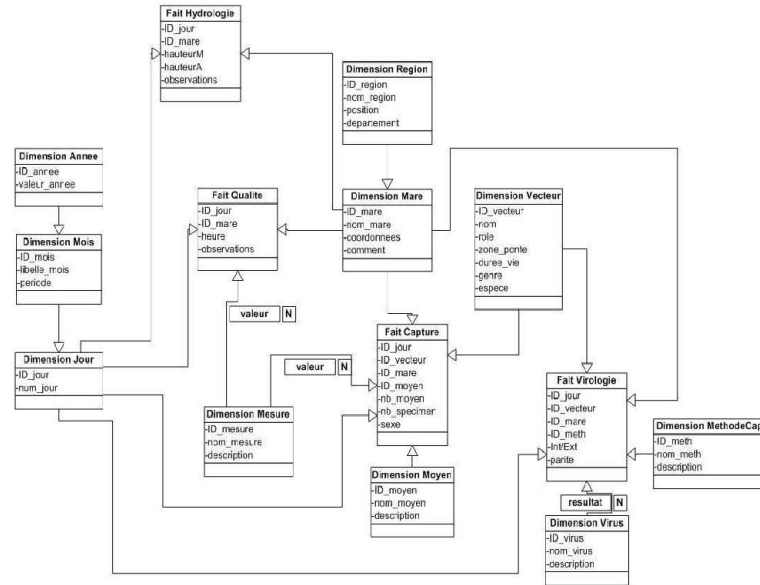


Figure 5.2 – Le modèle multidimensionnel proposé par [Bouba *et al.*, 2013]

Dans la même optique, la proposition d'un modèle géo-décisionnel pour la surveillance épidémiologique de la maladie de la tuberculose a fait l'objet du travail de [Lahleb, 2015] l'objectif principal est de répondre à la problématique de la persistance de la maladie dans certains foyers et zones. L'implémentation de ce modèle a l'ambition d'opter pour des outils spatiaux à savoir les Entrepôts de données spatiaux, SOLAP, etc. nécessitant l'intégration d'un module supplémentaire dédié à la cartographie.

En termes d'applications, nous retrouvons principalement les références suivantes :

- Kerkri el al, [Kerkri *et al.*, 1998] : mise en place d'un ED pour le réseau régional de santé permettant le regroupement des informations médicales, dans le cadre d'un projet de suivi épidémiologique de patients.
- Landais et al.[Landais *et al.*, 2002] : mise en place d'un SID pour le Réseau Epidémiologie et Information en Néphrologie (REIN), consacré à l'insuffisance rénale terminale (IRT).
- [Cuggia *et al.*, 2009] : mise en place d'un ED de documents patient (EDP) alimenté par un SIH (Système d'Information Hospitalier) et développement des outils de recherche d'information.

4. Médecine d'Anticipation Personnalisée

5.4.2 Les travaux traitant les entrepôts de données médicales et l'OLAP

Le projet ICEM/SE a mené au développement d'un outil d'analyse des données spatiales « SOLAP ». Cet outil a le potentiel de contribuer considérablement à l'amélioration de la surveillance en santé environnementale en donnant accès à une vingtaine d'indicateurs de santé et d'indicateurs environnementaux. Le SOLAP facilite l'accès aux données statistiques en permettant de créer facilement des cartes géographiques. Ce projet innove également la combinaison des structures de bases de données multidimensionnelles avec les fonctions SIG, ainsi que l'utilisation du SOLAP [Proulx *et al.*, 2002].

De nombreux travaux ont implémenté l'outil SOLAP soit en tant qu'outil opérationnel soit en tant que prototype préindustriel ayant fait l'objet de nombreuses publications scientifiques dans le domaine de la Géomatique et de l'Ingénierie Informatique. Pour le domaine de la santé, le Ministère de la Santé Canadian a investi dans SOLAP pour améliorer la surveillance de la santé environnementale. Des prototypes ont été réalisés dans le cas du virus du Nil Occidental ou encore sur l'analyse des maladies respiratoires [Rivest *et al.*, 2005]. D'autres études ont été menées sur le bioterrorisme en faisant des analyses multidimensionnelles d'ensembles de données spatio-temporelles épidémiologiques [Berndt *et al.*, 2007].

Les principales études exploitant les techniques de fouille de données (Datamining) sont illustrées par la Table 5.2 :

Table 5.2 – Synthèse des travaux appliquant les ED et la fouille de données en santé publique

Références	Contributions du travail	Approches et outils utilisés
[Mejhed <i>et al.</i> , 2009]	L'élaboration d'un système d'aide à la décision fondée sur le développement et la mise en œuvre d'outils pour analyser les indicateurs épidémiologiques	Nouveaux outils de l'ED et de fouille de données (data mining)
[Zhou <i>et al.</i> , 2010]	La réalisation d'un système basé sur l'ED clinique qui intègre les dossiers de données médical pour l'extraction de connaissances médicales et l'aide à la décision clinique.	Le SID comprend un outil OLAP et un module de fouille de données.
[Santos <i>et al.</i> , 2013]	Mise en place d'un système d'extraction de données automatisé qui permet aux décideurs de la santé publique d'accéder à des informations analytiques concernant les tumeurs cérébrales.	La construction d'un ED et l'utilisation de l'ontologie dans un processus de data mining automatisé.
Suite page suivante ...		

... Suite

Références	Contributions du travail	Approches et outils utilisés
[Zemri <i>et al.</i> , 2015]	Élaboration d'un modèle d'aide à la décision pour la surveillance épidémiologique	L'intégration du SOLAP et l'analyse multicritère. L'utilisation de la fouille de données spatiales

5.4.3 Les travaux traitant les entrepôts de données et les réseaux sociaux

Les travaux les plus pertinents traitant les entrepôts de données et les réseaux sociaux pour la gestion de la santé publique sont résumés dans la Table 5.3 :

Table 5.3 – Synthèse des travaux appliquant les ED et l'analyse des réseaux sociaux en santé publique

Références	Contributions du travail	Approches et outils
[Ortiz-Pelaez <i>et al.</i> , 2006]	Le rôle de la position des fermes dans le réseau et l'impact d'exploitations agricoles en particulier sur la propagation de la maladie.	<ul style="list-style-type: none"> — L'analyse un réseau social composé de 653 nœuds et 797 arrêts extrait d'un ED ; — Le mouvement des animaux constitue le lien entre les membres du réseau et le contact des animaux à travers ces mouvements est le mode de transmission de la maladie. — Trois mesures ont été exploité : l'intermédiarité " Betweenness", le k-voisins et la classification hiérarchique.
[Bouillot, 2011]	La proposition d'une solution pour pouvoir intégrer toutes les données qui peuvent être extraites des tweets ⁵ dans un ED.	<ul style="list-style-type: none"> — L'ED servira de support à une activité d'informatique décisionnelle sur les tweets ; — L'analyse du contenu des messages permettant une analyse multidimensionnelle ; — Une agrégation classique sur le nombre de tweets ; — La construction des dictionnaires spécifiques (SMS, Smileys, Sentiments) pour l'analyse des statuts dans les réseaux sociaux.

Suite page suivante ...

... Suite

Références	Contributions du travail	Approches et outils
[Nagel <i>et al.</i> , 2013]	L'exploration de l'interaction entre l'activité de message du cyberspace, mesurée par des mots clés spécifique des tweets, et l'occurrence de la grippe et la coqueluche à l'échelle mondiale.	<ul style="list-style-type: none"> — L'étude des corrélations entre les tweets agrégées d'une façon hebdomadaires et l'apparition de la maladie ; — L'utilisation de SIG pour visualiser l'évolution des maladies dans le temps et dans l'espace dans la région d'étude.

Peu de travaux ont abordé le couplage des ED et l'analyse des réseaux sociaux dans le cadre de la surveillance d'épidémies. En effet, nos contributions, dans la présente étude, par rapport à ces travaux consistent à concevoir et déployer un ED pour stocker les données multiples : médicales, socioéconomiques, environnementales, et simulées afin de conduire une analyse spatiotemporelle et multidimensionnelle par le SOLAP.

5.5 Autres stratégies de lutte contre les épidémies

5.5.1 Surveillance active ou passive

La surveillance peut être définie comme active ou passive en fonction de la méthode utilisée pour collecter les données. Dans la surveillance passive, on qualifie de passive toute activité de surveillance qui repose sur la déclaration spontanée des cas ou suspicions de cas de la maladie surveillée (par des acteurs sources de données) [Dufour et Hendrikx, 2007].

Dans le cas de la surveillance active, on qualifie d'active toute méthode de surveillance reposant sur la recherche des données par des actions programmées à l'avance et élaborées par l'animateur du réseau de surveillance [Dufour et Hendrikx, 2007].

Une épidémie de rougeole est survenue en 2005–2006 à Mayotte. L'objectif des auteurs dans [d'Ortenzio *et al.*, 2008] était de décrire l'épidémie, d'analyser les caractéristiques épidémiologiques et sociodémographiques des cas et de proposer des recommandations pour la surveillance et la prévention de la rougeole. Une investigation épidémiologique autour des cas a été réalisée et un système de surveillance passif renforcé des cas incidents a été mis en place.

Par ailleurs, dans dans [Nkoghe Mba *et al.*, 2005], un système de surveillance active a été mis en place dans trois districts touchés par l'épidémie (Zadié, Ivindo et Mpassa-Gabon) afin d'organiser la détection des cas et le suivi de leur contacts.

La surveillance active fournit généralement des données de meilleure qualité que la surveillance passive, mais nécessite un surcoût humain et financier dont l'utilité doit être évaluée en fonction de l'objectif.

5.5.2 Surveillance exhaustive

Elle collecte l'ensemble des cas qui surviennent dans une population de référence : surveillance des causes de décès, des maladies à déclaration obligatoire, des maladies professionnelles, des effets indésirables médicamenteux, etc. Les limites de ce type de surveillance peuvent être liées soit à la mauvaise qualité des données, si les données recueillies

sont trop nombreuses pour être validées ou si elles ne pas adaptées aux objectifs de surveillance, soit à leur sous-déclaration. Néanmoins, une surveillance exhaustive est utile pour estimer des tendances d'incidence sur de très longues périodes de temps et pour décrire des phénomènes de santé rares et graves.

5.6 Conclusion

Dans ce chapitre nous avons tenté de présenter les principaux travaux liés à la surveillance épidémiologique. Nous avons classé la surveillance d'épidémie selon les approches utilisées, le contexte et le cas d'étude en quatre classes. Nous avons, également, positionné nos contributions par rapport à ces travaux.

A travers cet état de l'art, nous concluons que l'objectif principal des systèmes de surveillance en épidémiologie utilisant les différentes approches citées au-dessus, est de permettre aux autorités sanitaires et aux épidémiologistes d'améliorer et de faciliter leurs travaux de surveillance et/ou de planification.

Nous constatons, également, que l'utilisation de l'information à référence spatiale est récente dans le domaine d'épidémiosurveillance.

Dans ce contexte, l'intégration des réseaux sociaux dans un ED et leurs analyses dans le domaine de la santé est très récente et peu de travaux ont adopté cette approche. Ce qui nous encourage à opter pour cette intégration.

Dans le chapitre suivant, nous allons mettre le modèle d'épidémie SEIR-SW au service de l'aide à la décision, nous essayons à travers une intégration de ce modèle dans un ED d'exploiter le fonctionnement du SEIR-SW pour une analyse multidimensionnelle.

Deuxième contribution

Chapitre 6

Le système décisionnel proposé SYDSEP

Sommaire

6.1	Introduction	87
6.2	Le système proposé : SYDSEP	88
6.2.1	Les objectifs du système SYDSEP	88
6.2.2	La description du système SYDSEP	88
6.3	La démarche décisionnelle adoptée par SYDSEP	92
6.4	Modélisation UML de SYDSEP	93
6.4.1	Diagramme de cas d'utilisation	93
6.4.2	Diagramme de séquence	93
6.5	Description de l'entrepôt de données	94
6.5.1	Les sources de données	94
6.5.2	Modélisation multidimensionnelle de l'ED spatio-temporel	96
6.6	Mise en œuvre de SYDSEP	99
6.6.1	L'environnement technique	99
6.6.2	Le découpage spatial administratif de la région d'étude	100
6.6.3	SIG et analyse spatiale	100
6.6.4	Création des cubes d'analyse en ligne	101
6.6.5	Suivi et surveillance de la grippe saisonnière	101
6.6.6	La prédiction	104
6.7	Le système SYDSEP : Discussion	107
6.8	Conclusion	108

6.1 Introduction

Dans la partie précédente, nous avons proposé un modèle de simulation SEIR-SW, nous avons décrit ses différents composants, ainsi que la stratégie d'intégration adoptée par ce modèle.

Dans ce chapitre, nous allons présenter le SYstème Décisionnel Spatiotemporel pour l'EPidémiologie (SYDSEP). Ce dernier intègre plusieurs variantes qui contribuent aux mieux pour analyser le contexte d'épidémiologie. Il repose sur une utilisation combinée de système de simulation SEIR-SW, de système d'entrepôt de données et de système d'aide à la décision. Nous allons, tout d'abord, décrire le système intégrant le simulateur SEIR-SW dans un SID. Nous préciserons ensuite les différents types de données utilisées dans notre

système SYDSEP. Ensuite, nous présenterons une modélisation multidimensionnelle de ces données multiples. Enfin, nous terminerons par l'étude de nos analyses dans divers axes spatiaux et temporels par la mise en place du SYDSEP.

6.2 Le système proposé : SYDSEP

Le principal objectif de cette thèse est d'élaborer un système d'épidémiosurveillance. Ce dernier permettra de fournir au personnel de la santé publique un outil souple et simple d'aide à la décision qui l'assiste efficacement dans sa recherche de solutions en cas de crise sanitaire.

6.2.1 Les objectifs du système SYDSEP

La question principale qui se pose : « *Quel système d'épidémiosurveillance faut-il adopté pour supporter : la modélisation et la simulation d'un processus de propagation issue du monde réel, l'intégration et le stockage ainsi que l'exploitation de données provenant de sources multiples, la gestion d'entrée/sortie du simulateur SEIR-SW, et par conséquent, l'analyse des données multiples, spatiotemporelles et volumineux ?* ».

Bien évidemment, d'autres questions sont encore prioritaires :

- Quels sont les foyers d'épidémie et les régions à risque ?
- Est-ce que l'épidémie commence à se propager à partir d'un seul ou de plusieurs sites infectés ?
- Existente-ils d'autres facteurs hormis que ceux liés aux caractéristiques de la maladie et la structure sociale des individus ?

Pour répondre à ces questions, nous focaliserons notre attention sur la solution de Business Intelligence (BI). Cette dernière est très recommandée pour gérer, intégrer et analyser des données volumineuses. En effet, notre objectif réside dans la mise en place d'un système décisionnel spatiotemporel SYDSEP qui permettra aussi de recueillir une foule d'informations sur les risques d'épidémie et sur les dommages associés. Celui-ci est basé sur la modélisation et l'élaboration d'un simulateur hybride (SEIR-SW) et son intégration avec le SID relatif au programme de lutte contre l'épidémie de la grippe saisonnière. Dans le SID, le rôle du système d'entrepôt est le stockage des données réelles et simulées générées par le SEIR-SW pour des fins d'analyses. Ainsi, le SOLAP est primordial pour effectuer une analyse spatio-temporelle du phénomène étudié par l'utilisation des requêtes plus spécialisées.

Nous avons décrit, dans le chapitre précédent, un état de l'art sur les systèmes décisionnels médicaux standards, à référence spatiale, utilisant la fouille de données ou l'analyse des réseaux sociaux. Dans ce chapitre, nous proposons un nouveau système décisionnel hybride en couplant au SID le simulateur SEIR-SW.

6.2.2 La description du système SYDSEP

Dans cette section, nous décrivons le système d'information décisionnel spatio-temporel pour le suivi et la prévention d'épidémie proposé, qui répond aux besoins des responsables de santé publique. Une méthodologie s'appuyant sur différentes approches et techniques pour améliorer les performances de notre SYDSEP sera proposée. En effet, ce système est composé principalement de trois sous systèmes : (1) le système de simulation, (2) le système d'entrepôt, et (3) le système d'aide à la décision. Une vue d'ensemble de notre système est illustrée par la Figure 6.1.

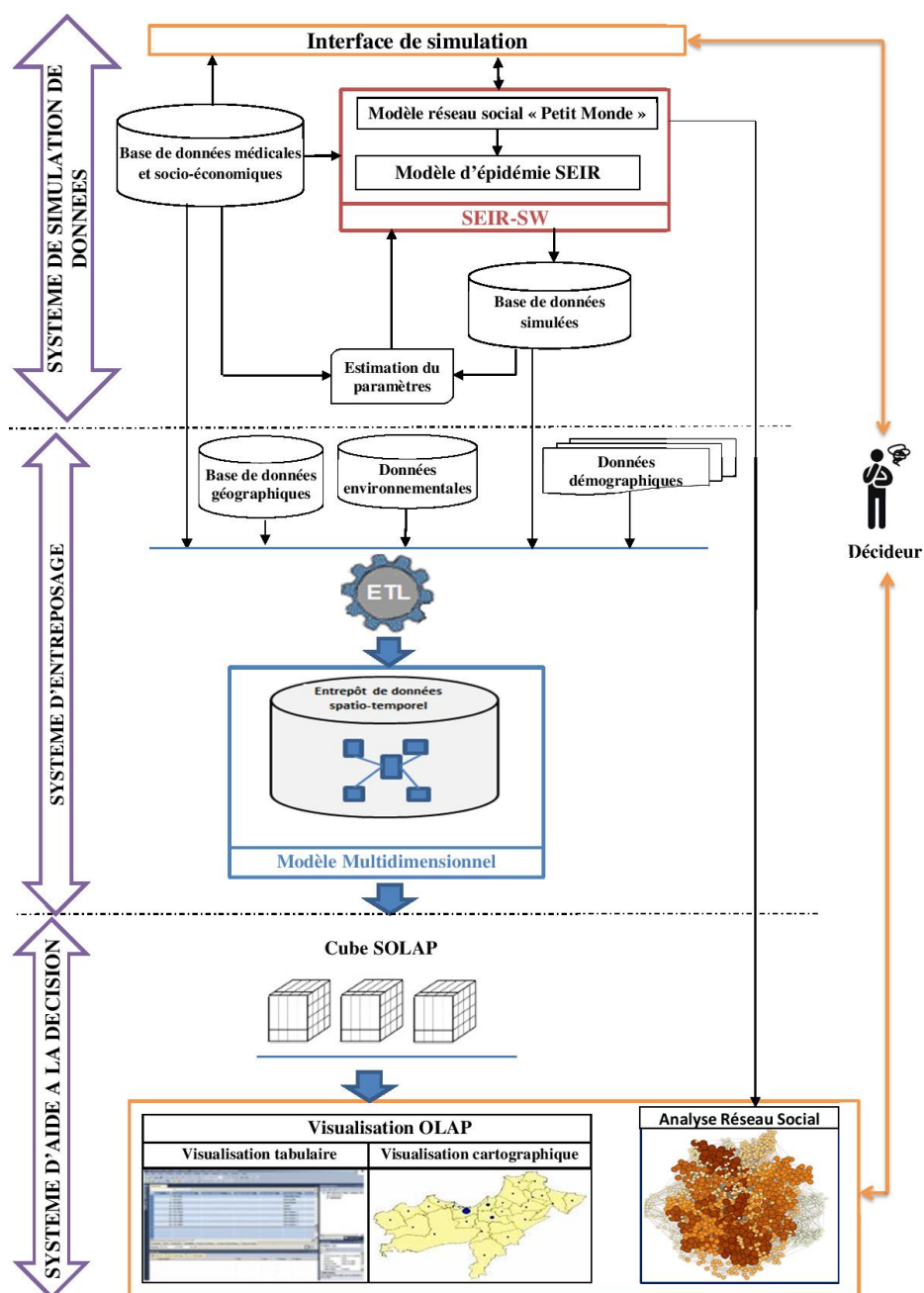


Figure 6.1 – Le système décisionnel SYDSEP proposé

6.2.2.1 Le sous-système de simulation

Dans cette section, nous proposons un nouveau processus qui consiste à incorporer dans le SEIR-SW le contexte décisionnel [Younsi *et al.*, 2014b]. Nous rappelons que le système SEIR-SW est détaillé dans la partie précédente. Le processus de SEIR-SW proposé décrivant l'interaction entre les différents acteurs est illustré par la Figure 6.2.

Opérationnellement, le système SEIR-SW commence par l'identification des paramètres du modèle, à savoir : k , α , β , et γ .

Le décideur (personnel de santé) doit fixer, pour chaque simulation : le taux de transmission, la durée de latence, la durée d'infection et le nombre de contacts par personne. Ensuite, le scientifique (l'homme d'étude capable d'interpréter les résultats de la simulation) utilise les capacités analytiques du SEIR-SW pour générer l'ensemble des données simulées. Ces dernières doivent être confrontées à la réalité. Autrement dit, si ces données sont incohérents c'est parce que les paramètres choisis au préalable ne sont pas correctes, et par conséquent, il faut les modifier.

Si les résultats sont conformes, une présentation des courbes décrivant les états de santé des individus dans le temps et des visualisations du réseau social sont mises en place. Finalement, le décideur analyse l'ensemble des résultats obtenus pour répondre aux questions posées.

Le processus interactif fournit aux décideurs de la santé publique des analyses multiples et interactives ainsi que des simulations. Par conséquent, l'interaction peut être facile et efficace entre les différentes parties prenantes dans le processus décisionnel SEIR-SW.

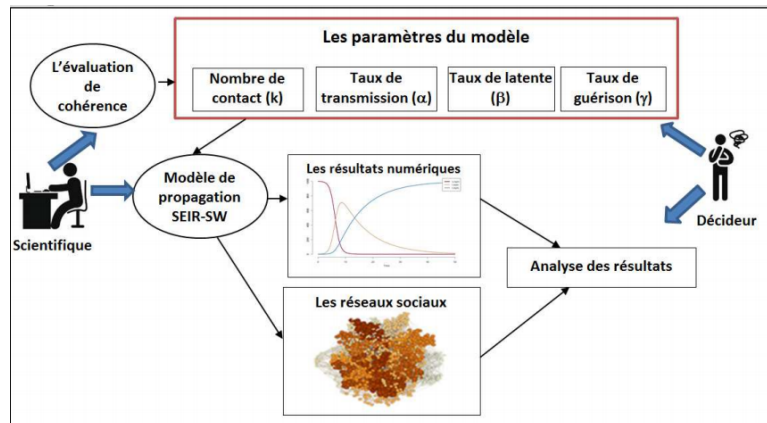


Figure 6.2 – Le processus interactif de SEIR-SW

6.2.2.2 Le sous-système d'entreposage

L'objectif de ce système est de constituer un ED pertinent pour l'application visée. C'est une étape nécessaire dans SYDSEP. Cette étape permettra d'identifier les différentes sources de données et de déterminer le mode d'extraction des données désirées comme nous devons aussi tenir compte de la nature particulière des données. Ensuite, elle passera au stockage de l'ensemble de ces données dans l'ED. Finalement, elle permettra la visualisation des données historiées.

Le processus ETL joue un rôle primordial dans le système d'entreposage, il permettra l'extraction, la transformation et le chargement des données multiples dans l'ED.

Tout d'abord, les données sont extraites de différentes sources de données. Chaque source de données possède ses propres caractéristiques qui doivent être gérées afin d'ex-

traire efficacement les données par le processus ETL.

Durant cette étape, le processus ETL doit prendre en considération les points suivants : (i) l'utilisation de pilotes ODBC/JDBC pour connecter aux bases de données sources, (ii) la compréhension de la structure de données de sources, et (iii) la gestion des données sources de différente nature. Le processus ETL rafraîchit l'ED avec les données modifiées et ajoutées dans les données source. Ce processus est périodique selon le cycle de rafraîchissement et les besoins de l'utilisateur.

Ensuite, les données extraites sont organisées dans une base de données relationnelle temporairement où elles sont transformées et sont nettoyées avant d'être chargées dans l'ED. Dans cette étape, le processus définit la granularité des tables de faits, les tableaux de dimensions, schéma de l'ED, etc. Enfin, le chargement de données multiples dans l'ED implique le chargement des tables de faits et de dimensions avec leur relations. Le processus ETL est illustré par la Figure 6.3.

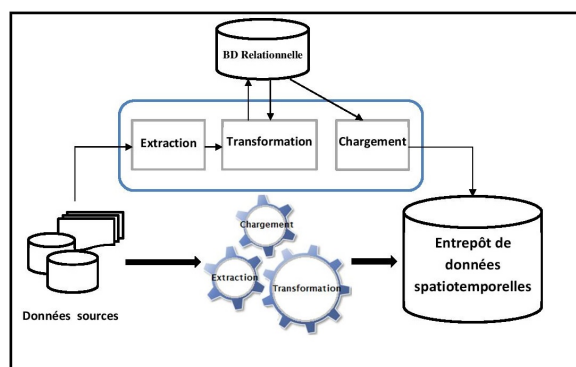


Figure 6.3 – Le processus ETL

La construction d'un entrepôt de données est donc une tâche complexe, elle procède comme suit :

- L'identification des besoins des différents intervenants du système décisionnel ;
- L'analyse des sources de données de manière qu'elles soient de qualité pour leur stockage dans l'ED ;
- L'organisation des données multiples dans l'ED d'où plusieurs problèmes se présentent : la gestion de l'ensemble de données, le rafraîchissement de l'ED, etc.
- L'exploitation des données multiples par des cubes OLAP.

Chaque étape de ce processus présente des problématiques spécifiques.

6.2.2.3 Le sous-système d'aide à la décision

Face aux problèmes de santé, les décideurs manquent d'outils cartographiques simples et puissants pour explorer, synthétiser et analyser des données spatio-temporelles telle que que la détermination des zones à risque. En effet, la visualisation des foyers épidémiques sur des cartes géographiques est gérée par un SIG. Toutefois, cette technologie est incapable de supporter les problèmes multidimensionnels suite aux progrès des solutions issues du domaine de l'informatique décisionnelle. La solution la plus diffusée pour surmonter les insuffisances des SIG, est de les coupler avec le moteur OLAP. Cette intégration permet la maîtrise de la complexité des problèmes décisionnels à référence spatiale qui nécessite l'utilisation de méthodes, de techniques et d'outils d'analyse puissants, afin de pouvoir non seulement gérer mais aussi analyser des données multiples. Des solutions, telles que

le SOLAP, permettant de supporter les données spatiales dans l'OLAP, ont été développées dans [Bédard *et al.*, 1997]. C'est une technologie qui se situe au cœur des systèmes d'information géo-décisionnelle. Le couplage (SIG - OLAP), permet une représentation cartographique des résultats issus des cubes OLAP.

6.3 La démarche décisionnelle adoptée par SYDSEP

La démarche décisionnelle adoptée par le système décisionnel SYDSEP est illustrée par la Figure 6.4.

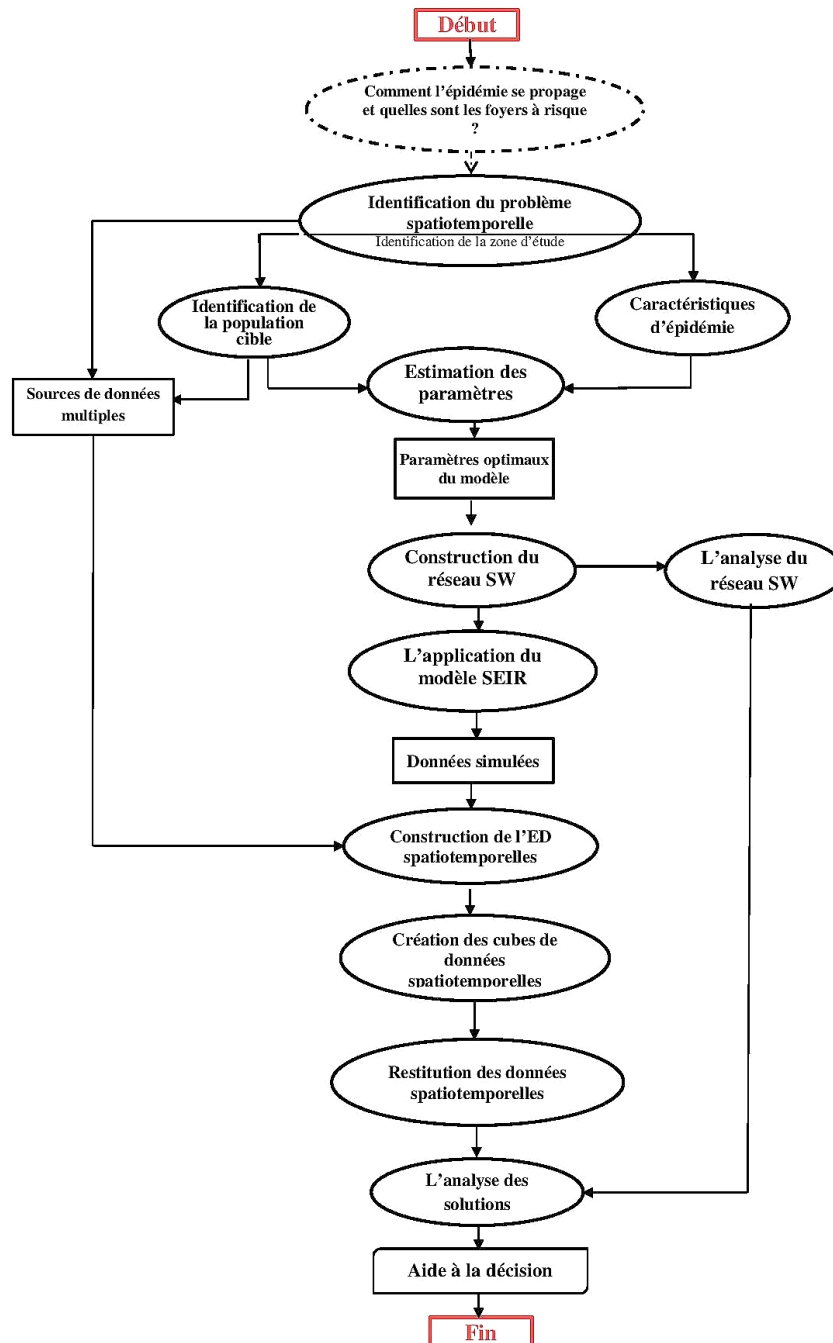


Figure 6.4 – La démarche décisionnelle proposée

6.4 Modélisation UML de SYDSEP

Nous avons opté pour le modèle UML qui permet d’avoir une vue à la fois simplifiée et schématique du système décisionnel proposé. Nous présentons dans les sections suivantes les diagrammes UML (cas d’utilisation et séquences) associés à SYDSEP.

6.4.1 Diagramme de cas d’utilisation

Les différentes fonctionnalités fournies par notre système SYDSEP sont illustrées par le diagramme de cas d’utilisation présenté par la Figure 6.5 :

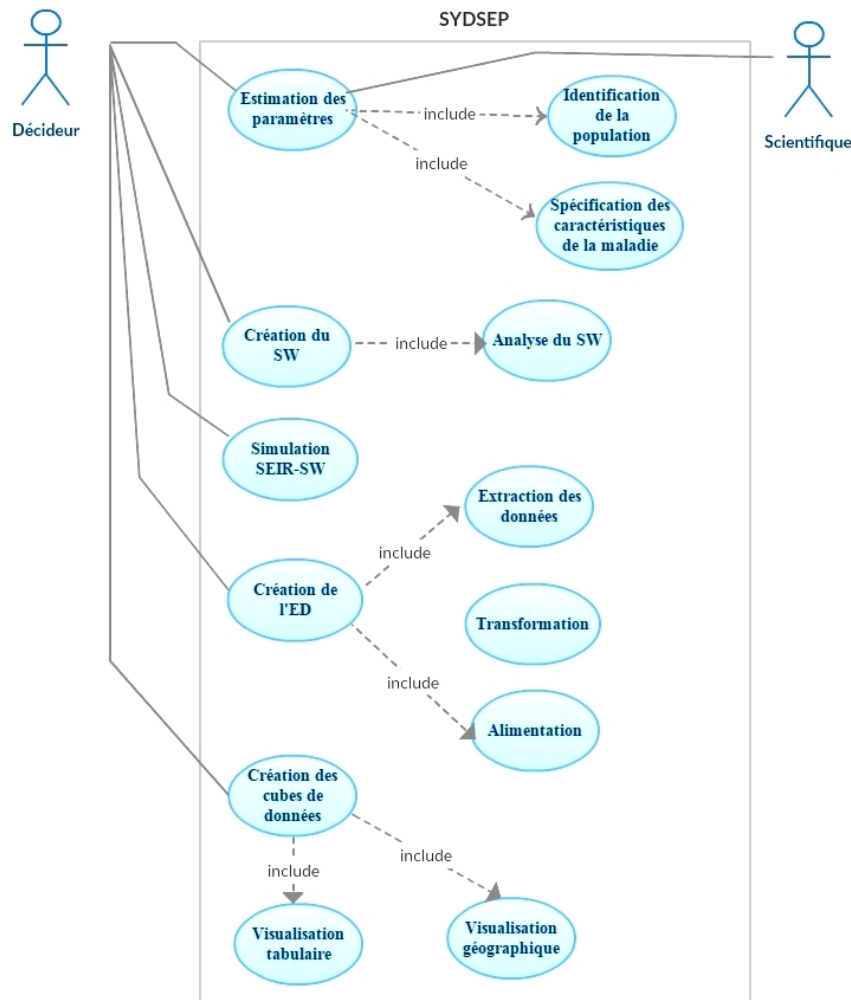


Figure 6.5 – Le diagramme de cas d’utilisation

6.4.2 Diagramme de séquence

Le diagramme de séquence décrivant le système SYDSEP est représenté par la Figure 6.6 :

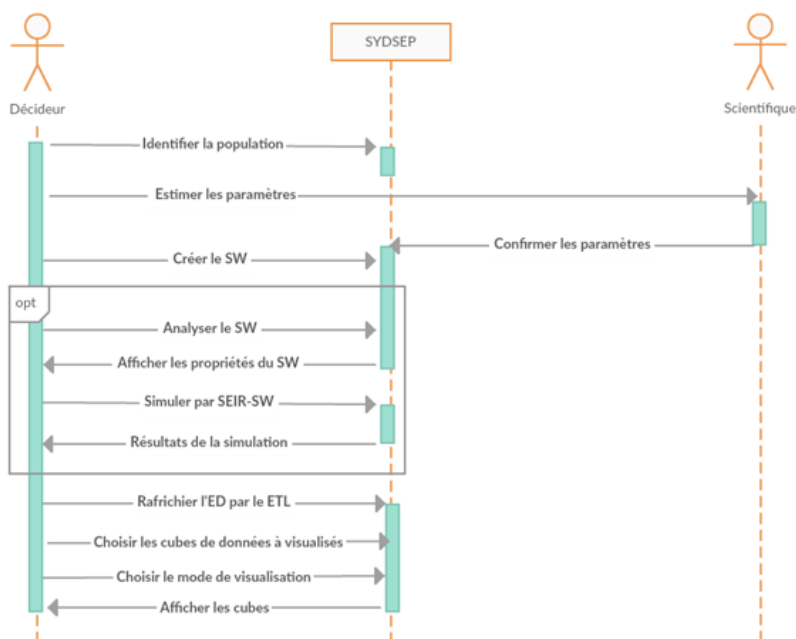


Figure 6.6 – Le diagramme de séquences

6.5 Description de l'entrepôt de données

Dans cette section, nous nous intéressons à la conception de l'entrepôt de données médicales et géo-spatiales. Les données alimentant l'entrepôt ont un impact sur la propagation de la grippe par leurs liens à la maladie, à la population et à la région concernée par l'étude. Elles proviennent de sources multiples, elles sont diverses et souvent hétérogènes et de nature complexe. Ces données peuvent être de type environnemental, démographique, géomatique, socio-économique et médical.

6.5.1 Les sources de données

Afin de réaliser un système de suivi et de prévention des épidémies, nous avons choisi d'étudier et de modéliser le processus de diffusion au travers d'outils informatiques bien adaptés. Pour cela, il était indispensable de :

- mettre en place une stratégie d'acquisition de l'information auprès des différents services fournissant les données (quelles données recueillir ? et sous quelle forme ?) ;
- organiser la gestion de ces données ;
- les intégrer dans le SYDSEP et les exploiter afin d'apporter aux décideurs un outil d'aide à la décision efficace.

Il est à noter qu'il existe différents services et établissements (organismes) nationaux exhaustifs rassemblant des données diverses pouvant être utilisées dans le SYDSEP.

Dans ce qui suit, nous présentons les différents types de données exploitées par notre ED hormis les données simulées :

Environnementales

Les épidémies saisonnières de grippe surviennent généralement, dans l'hémisphère nord, entre octobre et mars. Les basses températures, la faible humidité relative ou absolue, le chauffage des maisons, semblent expliquer la saisonnalité hivernale de la grippe

[Barreca et Shimshack, 2012]. Ces conditions participent directement au déclenchement épidémique de la grippe. En effet, dans ce travail, nous nous intéressons plus aux facteurs climatiques qu'aux autres facteurs environnementaux (Voir Annexe B). Ces données sont fournies par l'ONM (Office National de la Météorologie) de la ville d'Oran ;

Démographiques

Dans le cadre de l'épidémiologie, l'étude de la démographie est intéressante pour mieux connaître la population étudiée. L'évolution du nombre de malades dépend de nombreux paramètres liés à la démographie : sexe, âge, distribution de la population, etc. Les données de la démographie utilisées sont fournies par la DPAT/APC (Direction de la Planification et de l'Aménagement du Territoire d'Oran) ;

Socio-Economiques

Données combinant deux facteurs : social et économique de la population. L'impact socio-économique de la grippe est non négligeable. La détermination des âges ou des groupes à risque sur lesquels il serait possible de cibler les campagnes de vaccination, est parmi les objectifs de l'étude socio-économique, ainsi que le calcul du taux de natalité et de mortalité dans une population. Dans ce travail, nous utilisons une base de données socio-économique fournie par la Direction de la Santé et de la Population (DSP) d'Oran ;

Géomatiques

Son objectif, par la présente étude, est la représentation spatiale des données récoltées pour identifier et représenter les résultats d'analyses afin de montrer les foyers épidémiques. La base de données géographique utilisée est fournie par le Centre des Techniques Spatiales (CTS) d'Arzew ;

Médicales

Dans ce travail, nous utilisons une base de données médicale fournie par la Direction de la Santé et de la Population (DSP) d'Oran.

Le croisement de ces informations multidisciplinaires aboutit à l'élaboration de cartes de risques réels et de potentiels qui peuvent permettre de mieux cibler les interventions sur le terrain afin de lutter contre la propagation de la maladie. Ces données alimenteront l'entrepôt de données (Figure 6.7) pour développer un système d'épidémiosurveillance en temps réel et en espace géographique dans le but de répondre aux questions déjà posées.



Figure 6.7 – Représentation schématique des données constituant l'entrepôt de données

6.5.2 Modélisation multidimensionnelle de l'ED spatio-temporel

Dans cette section, nous nous intéressons à la conception d'ED médicales et géospatiales dans le but de bien gérer les données multiples.

Plusieurs sujets d'analyse sont à étudier pour la conception du modèle multidimensionnel d'épidémiosurveillance : les statistiques, l'environnement, la simulation et la démographie, a selon plusieurs axes d'analyse : individus, maladies, catégories de maladie, établissements d'accueil, communes, régions, sexe, tranches d'âge et temps.

La modélisation multidimensionnelle est caractérisée par le fait qu'à chaque fait (sujet d'analyse) sont associées plusieurs valeurs observées (mesures), chose qui n'existe pas dans le modèle relationnel.

Pour la modélisation des données d'épidémiosurveillance, nous avons choisi, pour chaque fait, un schéma en étoile afin de faciliter la navigation entre les tables, et réduire ainsi le nombre de tables et le nombre de jointures entre ces dernières nécessaires pour effectuer des requêtes.

De plus, nous optons pour l'utilisation du moteur ROLAP (Relational OLAP) car nous utilisons directement un modèle relationnel. Le ROLAP est le mieux adapté pour des mises à jour très fréquentes, il extrait les données de la base de données à l'aide de requêtes SQL [Bédard *et al.*, 1997] et les traduit selon une vue multidimensionnelle.

1. *Identification des faits* : Six tables de faits sont identifiées formant au final un modèle en constellation :

- Table de faits MÉDICAL_STATISTICS : liée à six dimensions, et comporte comme mesures des indicateurs statistiques médicaux de morbidité :
 - *Incidence* : le nombre d'incidence est le rapport entre le nombre de nouveaux cas d'une maladie donnée survenus dans une population pendant une période définie ;
 - *Attak_rate* : c'est le rapport entre le nombre de cas confirmés dans une période donnée et le nombre des susceptibles d'être infectés.
- Table de faits ENVIRONMENT : en épidémiologie grippale, on s'intéresse aux mesures suivantes :
 - *High_temperature*, *Low_temperature* : températures maximale et minimale, calculées en Celsius (C°) ;
 - *Humidity relative* : l'humidité relative est la quantité de vapeur d'eau présente dans l'air, exprimée en pourcentage (%).
 - *Wind_speed* : la vitesse du vent est mesurée à 10 mètres de hauteur au-dessus d'un sol dont la rugosité est la plus faible possible, et est exprimée en mètre /seconde ;
 - *Soil_moisture* : l'humidité du sol est la variation des caractéristiques de différents matériaux ou sols ;
 - *Avg_temperature* : la température moyenne est la moyenne entre la température maximale et la température minimale ;
 - *Ratio H/T_max* : c'est le rapport entre l'humidité relative et la température maximale.

- Table de faits HEALTH_STATUS_OF_POPULATION : entourée des dimensions : WEEK, MONTH, COMMUNE et R_INDIVIDU . Elle a comme mesures :
 - *Case_fatality_rate* : le taux de létalité est le rapport entre le nombre de décès et le nombre de malades pendant une période de temps.
- Table de faits DEMOGRAPHY : entourée des dimensions : YEAR et COMMUNE. Ses mesures sont :
 - *NB_Man*, *NB_Woman* : le nombre de population résidente par sexe ;
 - *NB_dead_M*, *NB_dead_W* : nombre de décès dans la population résidente par sexe.
 - *NB_birth_M*, *NB_birth_F* : le nombre de nouveaux nés dans la population résidente par sexe.
 - *NB_population_SC*, *NB_population_MT* : nombre de la population résidente selon la dispersion : Agglomérations Secondaires (SC) et Agglomération Chef- Lieu (MT).
 - *Birthrate* (ou le *taux de natalité* : est le nombre de naissances vivantes de l'année sur la population totale moyenne de la même année.
 - *Mortality* : nombre de décès survenus au cours d'une période donnée généralement une année, dans une population donnée.
- Table de faits SIMULATION : elle est liée aux dimensions : COMMUNE, TEMPS, SEIR-SW. Ses mesures sont :
 - *NB_Susceptible* : fraction de population qui est saine mais pouvant potentiellement devenir infectée par la maladie ;
 - *NB_exposed* : ensemble d'individus infectés et qui ne sont pas encore infectieux (en période de latence) ;
 - *NB_infected* : ensemble d'individus infectieux qui peuvent transmettre la maladie aux individus sains ;
 - *NB_Removed* : ensemble d'individus retirés du processus de la maladie (guérison, décès, etc.)
 - *Attack_rate* (*Taux d'attaque ou d'atteinte*) : le rapport entre le nombre de cas d'infectés dans une période donnée et le nombre des susceptibles d'être infectés.
- Table de faits STATUT : elle est liée aux dimensions : A_INDIVIDU, SEIR_SW et TEMPS. Elle a une seule mesure :
 - *Statut* : permet de représenter le statut d'un individu dans un temps donné pendant la simulation.

2. Définition des axes d'analyse :

Dans ce qui suit, nous décrivons les différentes dimensions du modèle multidimensionnel. Nous distinguons des dimensions qui portent une seule hiérarchie et d'autres plusieurs.

- Axes d'analyse **thématiques** : elle contient les dimensions suivantes : AGE_RANGE : 0-5 ans, 6-17 ans, 18-59 ans, 60+ ans; SEX : homme (H) ou femme (F); HEALTH_CARE_FACILITY : regroupe les différents centres de soin de santé; DISEASE_CATEGORY (désignation catégorie maladie) : plusieurs types de grippe existent : aviaire, porcine, etc.
 Nous nous intéressons ici à « la grippe saisonnière »; DISEASE : trois grandes catégories de grippe saisonnière à citer : A, B, C, la grippe saisonnière rentre dans la catégorie A; R_INDIVIDU (individu réel) : contient comme descripteurs : Age : de 1 mois à 93 ans, décès : deux valeurs : 0 (vivre), 1 (décéder); A_INDIVIDU (individu virtuel) : c'est un individu généré par une simulation et identifié par son identifiant (*A_individu_ID*).
- Axes d'analyse **temporelle** : c'est une dimension conforme. Elle est liée aux six tables de faits. Nous avons défini cinq dimensions temporelles (DATE (jour), WEEK, MONTH, QUARTER, YEAR), avec les niveaux d'hierarchies suivantes : DATE → MONTH → QUARTER → YEAR et DATE → WEEK → YEAR.
 Une autre dimension de temps est générée par le système SEIR-SW : TEMPS : représentant le temps nécessaire pour étudier l'évolution d'épidémie pendant une simulation, l'unité de temps utilisée peut-être : jour, semaine, etc., selon le besoin de l'utilisateur. jour, semaine, etc., selon le besoin de l'utilisateur. Dans ce travail, nous nous intéressons à l'unité de temps : semaine.
- Axes d'analyse **spatiale géométrique** : elle comprend les dimensions suivantes : COMMUNE, décrite par *nom de la commune*, *pollution* : la quantité des déchets produite est mesurée par : Kg/Habitat/jour; *perimeter* : périmètre du polygone commune, *_Name* : longitude, *_Name_2* : altitude, *ID* : référence spatial générée par le SQL Spatial, *_ID* : limite de la zone géographique; RÉGION : dans le découpage administratif de la ville d'Oran, on trouve trois régions, chaque région regroupe un ensemble de communes (Est-Ouest-Sud); *health_care_center* : dans la ville d'Oran, il existe 4 types d'établissements de santé entre : EPSP (Etablissement Public de Santé de Proximité), EHS (Etablissements Hospitaliers Spécialisés), EPH (Etablissements Hospitaliers Restants), EH (Etablissements Hospitaliers), dans les 26 communes.

6.5.2.1 Description du schéma multidimensionnel d'épidémiosurveillance

L'ED ainsi défini est représenté par la Figure 6.8. Nous avons proposé un schéma en constellation qui combine les quatre schémas en étoile des faits suivants : MEDICAL_STATISTICS, ENVIRONMENT, HEALTH_STATUS_OF_POPULATION, DEMOGRAPHY, SIMULATION, STATUT (Les différents schémas sont décrits en détail dans l'annexe A). Le modèle en constellation permet aux tables de dimension d'être partagées entre plusieurs tables de faits. Nous proposons un rafraichissement de l'entrepôt de données et l'ensemble des cubes dérivés par la planification d'une exécution quotidienne du processus ETL.

L'utilisation de l'entreposage de données en temps réel permet essentiellement de comparer les données et les indicateurs clés actuels de la surveillance d'épidémie avec ceux déjà historisés .

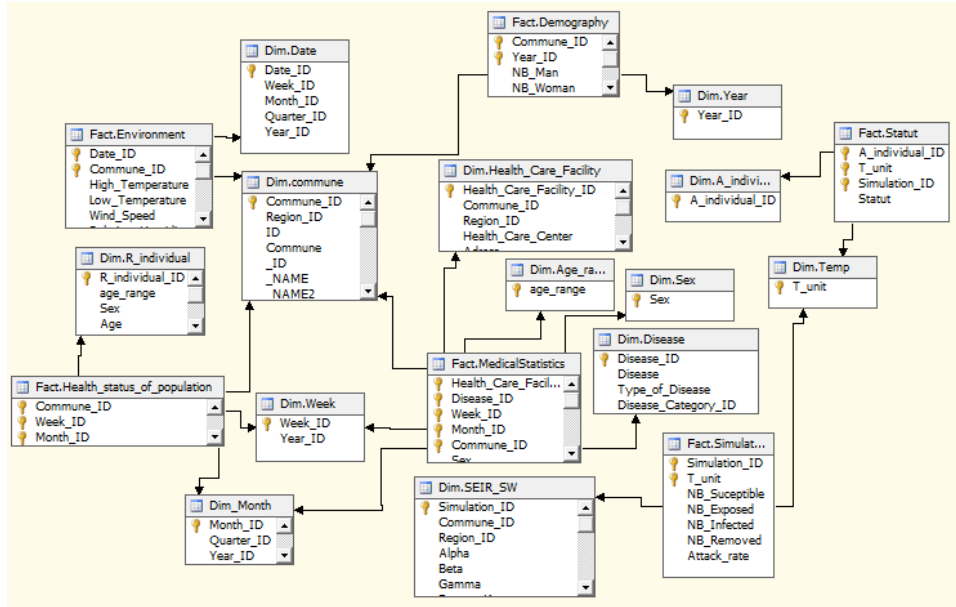


Figure 6.8 – Schéma du modèle multidimensionnel d'épidémiosurveillance

6.6 Mise en œuvre de SYDSEP

Dans cette section, nous décrivons les détails relatifs à la mise en œuvre du système SYDSEP, nous commençons tout d'abord par une description de l'environnement technique utilisé. Ensuite, nous présentons le découpage administratif de la région d'étude et le rapport entre le SIG, l'analyse spatiale et SOLAP. Enfin, nous présentons des exemples sur le suivi et la prévention de l'épidémie grippale.

6.6.1 L'environnement technique

Tout au long de ce travail, nous avons travaillé sur un PC Intel @Core ^{IM} i 5 Mhz équipée de 6 Go de RAM et d'un environnement Windows 7. L'implémentation du modèle de propagation était sur Python, un langage de programmation avec licence compatible GPL. Les calculs mathématique et l'affichage des courbes de l'évolution d'épidémies était assurée par plusieurs bibliothèques telle que : Matplotlib et numpy.

Dans la partie modélisation multidimensionnelle et restitution, nous avons opté pour le Microsoft, le Microsoft, SQL Server 2008 R2 (SP2) et ces outils. Trois plateformes sont mises à notre disposition :

- *SQL Server Integration Services (SSIS)* : permet d'intégrer des données provenant de différentes sources pour les ranger dans l'ED ;
- *SQL Server Analysis Services (SSAS)* : permet d'analyser les données agrégées lors de SSIS, grâce à des fonctions d'analyse multidimensionnelle ;
- *SQL Server Reporting Services (SSRS)* : permet de créer, gérer et publier des rapports résultant des analyses réalisées lors de SSAS.

Plusieurs raisons ont motivé le choix de SQL Server 2008 R2 (SP2). En effet, il permet :

1. d'architecturer, de sécuriser et de gérer des bases de données relationnelles ;
2. d'embarquer dans nos rapports du géospatial grâce à Report Builder 3 prise en charge par SSRS (2008) R2, et

nécessaire afin de lui permettre une capacité importante de traitement des données ainsi qu'une rapidité d'affichage des résultats d'analyse.

Ce couplage a donné naissance au SOLAP. Le Spatial OLAP dispose, comme un SAD (Système d'Aide à la Décision), de toutes les fonctionnalités OLAP et de celles du SIG, avec, en plus, des techniques de Géo-visualisation [Rivest *et al.*, 2005].

6.6.4 Création des cubes d'analyse en ligne

La construction d'un cube de données est une phase nécessaire afin de pouvoir naviguer facilement dans la base des données, et de procéder à la découverte d'informations significatives. Les données issues des cubes SOLAP sont restituées par le décideur via des outils d'analyse ou de « reporting ».

L'objectif principal de l'interface de restitution et d'analyse est double :

- d'une part, elle permet une représentation graphique des structures multidimensionnelles dédiées aux décideurs afin de les aider à spécifier leurs analyses pour visualiser les données et naviguer dedans.
- d'autre part, elle permet une visualisation des cartes de risque d'épidémie de la région d'étude. Dans ce travail, nous utilisons Microsoft Reporting Services (SSRS) pour se connecter à l'ED et récupérer les tables nécessaires pour alimenter les cubes SOLAP et pour créer les dimensions ainsi que les mesures et les champs agrégés des cubes.

Dans notre cas, l'acteur de la santé publique peut effectuer plusieurs analyses selon plusieurs dimensions des données réelles ou simulées. Il peut également imaginer plusieurs scénarios relatifs à la propagation de la maladie et/ou aux facteurs favorisant cette maladie.

6.6.5 Suivi et surveillance de la grippe saisonnière

Le système décisionnel SYDSEP proposé :

- prend en charge les données historiques multidimensionnelles,
- assure la surveillance de la propagation d'épidémie dans un temps efficace afin d'identifier rapidement les situations anormales,
- fournit un environnement analytique qui accélère les enquêtes menées par les épidémiologistes et les autres intervenants.

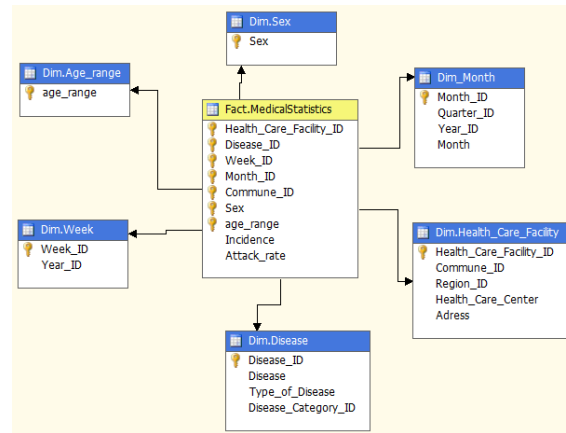
Afin que le décideur puisse effectuer une analyse détaillée, il doit choisir les mesures et les dimensions intéressantes pour cette analyse. Dans la section suivante, nous présentons les différents modes d'affichage fournis par le SYDSEP :

6.6.5.1 L'affichage tabulaire et graphique des analyses multidimensionnelles

Dans ce qui suit, nous présentons quelques exemples d'utilisation d'un cube de données à travers des illustrations. Nous avons choisi un des faits avec ses dimensions (FACT.MEDICAL_STATISTICS) sur lequel nous allons effectuer nos analyses. Le personnel de santé peut choisir d'autres faits selon ses besoins d'analyses.

Le cube de données est construit et déployé afin de répondre aux requêtes spatio-temporelles sur la surveillance épidémiologique. Les Figures 6.10a et 6.10b représentent, respectivement, le schéma du cube de la table FACT.MEDICAL_STATISTICS et la présentation tabulaire du cube. Ce dernier permet de calculer le taux d'attaque de la maladie ainsi que le nombre d'infectés en fonction des dimensions du territoire (RÉGION et COMMUNE),

et du temps (SEMAINE, MOIS, et ANNÉE), du SEXE (M ou F) et de la TRANCHE ÂGE. Des opérations telles que le forage vers le bas (Drill-down) et le forage vers le haut (Roll-up) peuvent être ainsi effectuées.



(a)

Commune ID ▾ Age Range ▾ Sex ▾															
15															
0-5 19-59 60+ 6-18															
F M Total															
Year ID ▾	Month ID ▾	Week ID ▾	Incidence	Attack Rate	Incidence	Attack Rate	Incidence	Attack Rate	Incidence	Attack Rate	Incidence	Attack Rate	Attack Rate		
2009	1/2009		23	1.751335E-03	83	6.320034E-03	78	5.939309E-03	161	1.225934E-02	65	4.949424E-03	36	2.741219E-03	
	10/2009		73	5.558583E-03	138	1.050801E-02	126	9.594268E-03	264	2.010228E-02	67	5.101713E-03	239	1.819865E-03	
	11/2009	49/2009	83	3.274235E-03	113	8.604383E-03	85	6.472325E-03	198	1.507671E-02	29	2.208204E-03	127	9.670414E-03	
	12/2009		49	3.731105E-03	82	6.243889E-03	79	6.015454E-03	161	1.225934E-02	33	2.512785E-03	131	9.974994E-03	
	50/2009		5	4.5687E-04	12	9.137399E-04	15	1.142175E-03	27	2.055915E-03	6	4.5687E-04	23	1.751335E-03	
	51/2009		2	1.5229E-04	1	7.614499E-05	1	7.614499E-05	1	7.614499E-05	1	1.5229E-04	2	1.5229E-04	
	52/2009														
	53/2009														
	Total		200	7.614499E-03	208	1.583816E-02	180	0.0137061	388	2.954429E-02	69	5.254005E-03	283	2.154903E-03	
	8/2009		4	3.0458E-04	5	3.80725E-04	6	4.5687E-04	11	8.375949E-04	4	4.5687E-04	2	1.5229E-04	
	9/2009		25	1.903625E-03	96	7.309919E-03	63	4.797134E-03	159	1.210705E-02	59	4.492554E-03	35	2.665075E-03	
	Total		225	1.713262E-02	530	4.035684E-02	453	3.449368E-02	983	7.485051E-02	266	2.025456E-02	596	4.538241E-03	
	Grand Total		225	1.713262E-02	530	4.035684E-02	453	3.449368E-02	983	7.485051E-02	266	2.025456E-02	596	4.538241E-03	

Value: 983

Total: Incidence

Row Member: Grand Total

Column Member: 15 - 19-59 - Total

(b)

Figure 6.10 – Représentation du cube multidimensionnel des statistiques médicales

A travers une navigation dans les données du cube, nous pouvons illustrer les données selon différents axes, afin de filtrer la quantité de données visibles des dimensions et aussi faire des "forages vers le bas" pour voir les données à un niveau de détails plus fin (Figure 6.10b). Une représentation des hiérarchies de la dimension du territoire (COMMUNE) et la dimension du temps (WEEK) est représentée par la Figure 6.11.

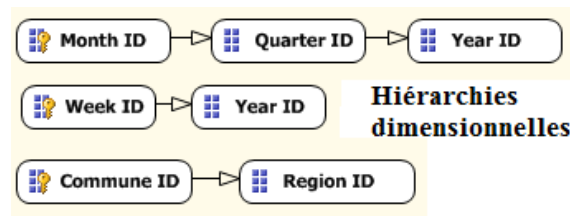


Figure 6.11 – Représentation des dimensions hiérarchisées

Des opérations pour l'obtention des extraits d'information représentés par le cube en fonction des dimensions peuvent être effectuées par les opérateurs *Slice* et *Dice*.

La Figure 6.12 représente les deux opérations sur le cube.

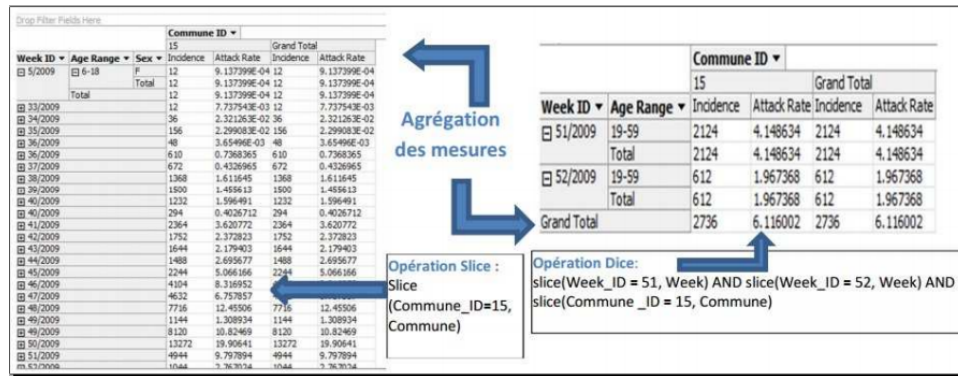


Figure 6.12 – Représentation des opérations *Slice* et *Dice* sur le cube

Les résultats de l'analyse multidimensionnelle peuvent aussi être présentés sous forme de graphiques. Dans la Figure 6.13a nous souhaitons observer l'incidence des deux sexes pour les différentes communes. Tandis que, dans la Figure 6.13b, nous affichons le taux d'attaque de la grippe saisonnière pour les 26 communes de la région et les différentes tranches d'âge de la population. Plusieurs choix d'affichage peuvent être générés par l'outil d'affichage (courbes, diagramme en camembert, etc.).

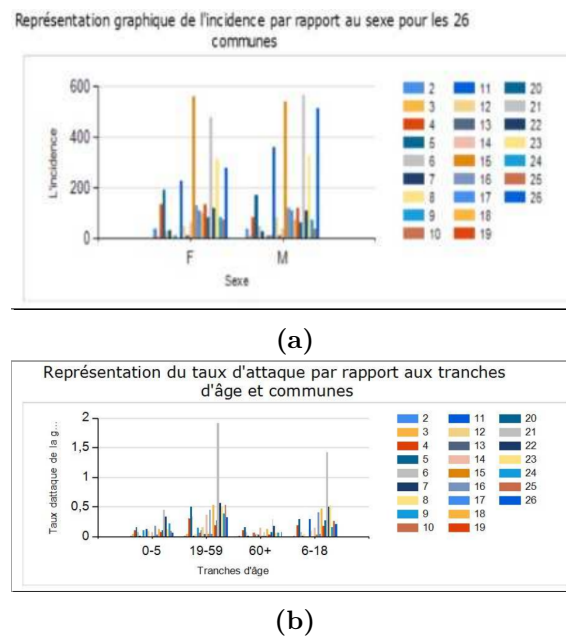


Figure 6.13 – Représentation graphique de requêtes générées : (a) l'incidence; (b) le taux d'attaque

Nous pouvons également améliorer les affichages cités en haut par une représentation cartographique à l'aide d'un outil Spatial OLAP.

6.6.5.2 L'affichage cartographique des analyses multidimensionnelles

Nous nous intéressons, dans ce qui suit, à l'affichage de quelques cartes de la région d'étude, les requêtes proposées selon deux dimensions : temps et communes choisis par le décideur. Dans la Figure 6.14a, nous présentons une distribution du nombre de cas

infectés par commune pour la semaine 52/2009. Nous ne remarquons aucun nouveau cas dans la majorité des communes. Par contre, dans la Figure 6.14b. Nous représentons une distribution du nombre de cas infectés par commune pour les mois d'octobre à décembre 2009. Les communes d'Oran, de Bethioua, d'Ain-el-Bia et d'Arzew représentent des foyers épidémiques. Ceci confirme les résultats obtenus précédemment qui nous montraient une très grande propagation de la grippe dans les communes.

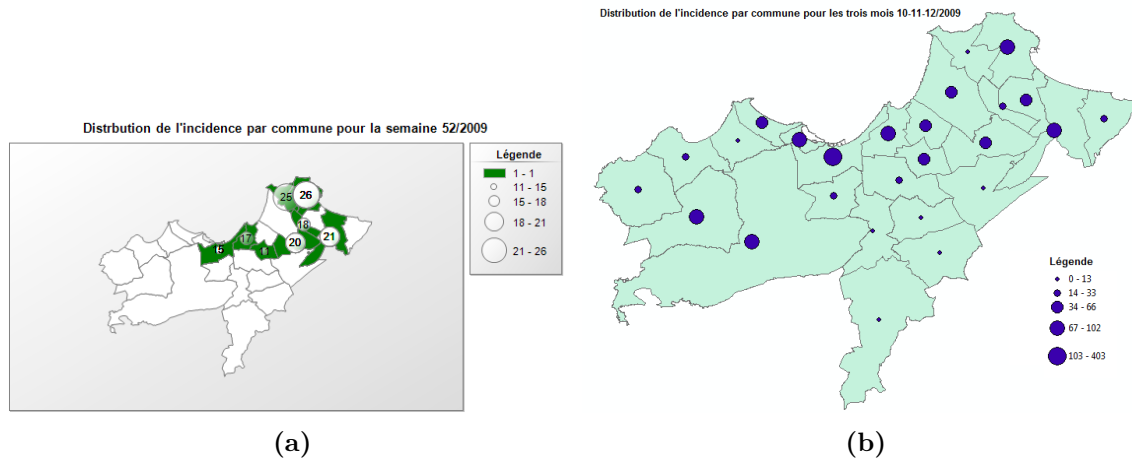


Figure 6.14 – Affichage cartographique du résultat de la requête des incidences : (a) la semaine 52/2009, (b) les mois entre 10/2009 jusqu'à 12/2009

6.6.6 La prédiction

Pour des fins de prévention, le personnel de la santé doit se préparer à d'éventuelles épidémies pouvant survenir. SYDSEP met au service du personnel de la santé un dispositif d'analyse de données ayant un impact sur l'apparition de la maladie. Dans ce qui suit, nous présentons quelques analyses multidimensionnelles de ces données :

6.6.6.1 L'affichage tabulaire et graphique des analyses multidimensionnelles

Pour des fins de prévention contre la grippe saisonnière, le décideur a la possibilité d'analyser les facteurs favorisant la transmission du virus par le choix des mesures et des dimensions sur lesquelles il veut effectuer son analyse, par exemple :

- quelles sont les communes ayant un taux de risque important ?,
- comment la grippe évolue dans le temps pour un ensemble de paramètres de simulation choisi ?,
- est ce que les facteurs climatiques ont un impact important sur la transmission du virus ?, etc.

La Figure 6.15a, illustre deux faits responsables de la prédiction des épidémies, ces deux faits sont alimentés par les sorties du simulateur SEIR-SW.

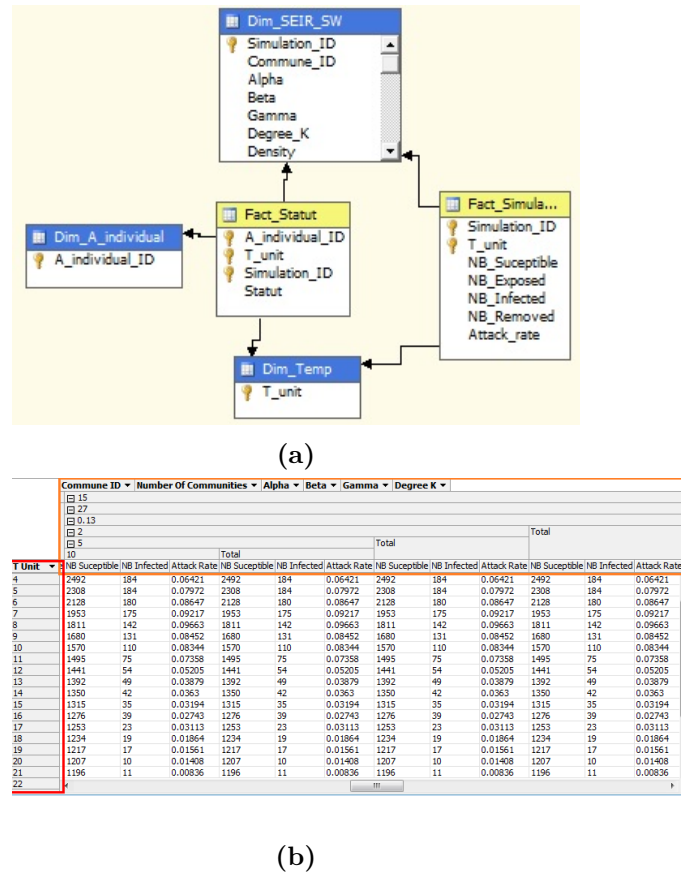
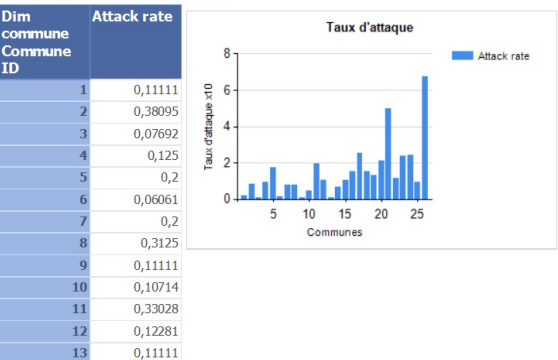


Figure 6.15 – Représentation du cube multidimensionnel des simulations

Dans la Figure 6.15b nous présentons un cube selon plusieurs dimensions : (la commune, le nombre de communauté, le degré k , α , β , γ) et (le temps de simulation), nous souhaitons calculer le nombre de susceptibles, d'infectés, et le taux d'attaque de la maladie par rapport aux indicateurs cités en dessus (dimensions). Nous rappelons, qu'il est aussi possible d'appliquer les différentes opérations OLAP, telles que : Slice, Dice, rotation, etc.

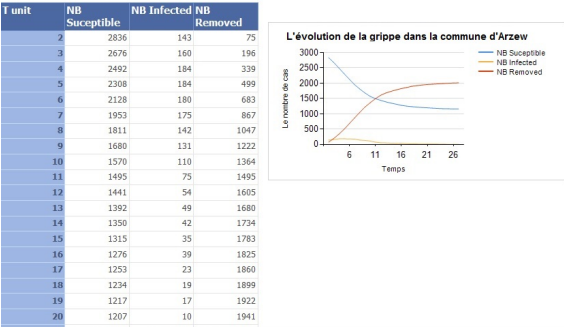
Le décideur peut également publier ces analyses sous forme tabulaire ou graphique. La Figure 6.16a illustre le taux d'attaque pour les 26 commune et Figure 6.16b illustre l'évolution de la grippe dans la commune d'Arzew qui représente un nombre important d'infectés et qui mérite une étude approfondie pour déterminer les causes de l'apparition de la maladie.

Le taux d'attaque de la maladie dans les 26 communes



(a)

L'évolution de la grippe dans la commune d'Arzew



(b)

Figure 6.16 – Représentation graphique de requêtes générées : (a) du taux d'attaque ; (b) l'évolution de la grippe

6.6.6.2 L'affichage cartographie des analyses multidimensionnelles

Les différentes analyses effectuées peuvent également être aussi visualisées sur la carte géographique de la région d'étude. La Figure 6.17a et la Figure 6.17b illustrent, respectivement, les résultats d'analyse du taux d'attaque et l'évolution de la grippe pour $t=5$.

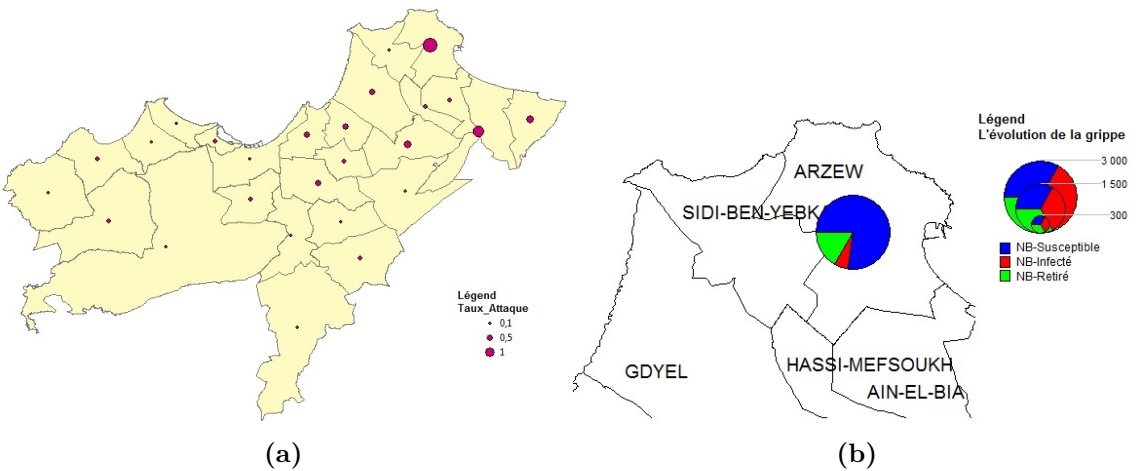


Figure 6.17 – Représentation cartographique de : (a) la distribution de taux d'attaque, (b) l'évolution de la grippe dans la commune d'Arzew

D'autres indicateurs ayant un impact sur la propagation d'épidémie sont liés aux facteurs démographique et environnementale. Toujours dans le contexte de la prévention, lorsque l'humidité relative est faible (air sec) et la température moyenne baisse, le virus de la grippe saisonnière survit plus longtemps, et le taux de transmission augmente. Afin de comprendre ce phénomène, nous avons effectué une analyse sur ces facteurs sur une période donnée. La Figure 6.18, illustre les résultats obtenus.

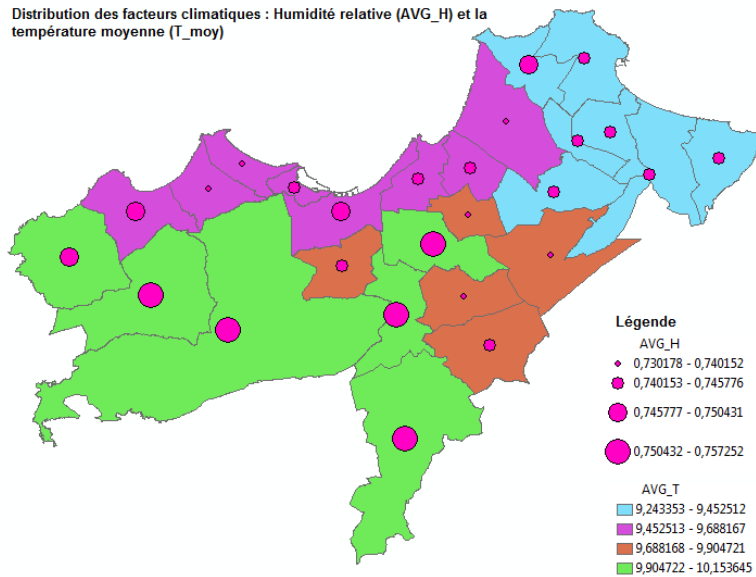


Figure 6.18 – Représentation cartographique de la distribution des facteurs climatologiques

Nous remarquons que la température moyenne (AVG_T) est élevée dans la zone Sud-Ouest de la région d'étude par rapport à la région Nord-Est pendant cette période. Cette variation entre $AVG_T = 9,24\text{ C}^\circ$ et $AVG_T = 10,15\text{ C}^\circ$ est négligeable car la distance entre les deux zones est petite (50 km). Nous remarquons également, que dans les communes où le nombre d'incidences est important, la température est plus au moins basse, ainsi que l'humidité relative (AVG_H). Ce qui explique l'existence d'autres facteurs favorisant la transmission de la maladie : démographiques, sociaux, etc. De plus, selon les résultats obtenus de l'analyse du SW, nous avons conclu que les épidémies ne peuvent pas commencer dans une population à très faible densité. Si elle commence, elle ne peut pas maintenir (c'est à dire, devienne endémique).

6.7 Le système SYDSEP : Discussion

Bien évidemment, le système décisionnel SYDSEP est itératif et séquentiel. Il est fondé sur une démarche participative dans laquelle on met à contribution la compétence et l'expérience de chaque intervenant dans le processus décisionnel. La qualité de ce système est donc influencée par le contenu de l'ED, le choix des mesures et des axes d'analyses ainsi que l'exploitation et l'analyse multidimensionnelle des données multidisciplinaire.

L'objectif de ce système est l'aide à la décision ou bien d'une autre manière l'obtention de la bonne décision. SYDSEP contribue à la conformité à la réalité, d'une part grâce aux données gérées par le système de simulation basées sur des paramètres estimés du monde réel ainsi qu'une analyse du réseau petit monde. D'autre part, grâce aux procédures d'analyse multidimensionnelle qui améliorent la fiabilité de l'évaluation des facteurs favorisant la transmission de l'infection au sein de la population humaine. On peut noter que le SOLAP améliore l'aspect analytique des données grâce à ses fonctionnalités de représentation spatiale et de cartographie.

Nous pensons que le système SYDSEP est plus adapté aux besoins des utilisateurs et aux caractéristiques des problèmes spatiaux. Il intègre plusieurs variantes qui contribuent aux mieux pour analyser le contexte de la santé publique. Notre contribution est d'expérimenter pour la première fois l'intégration d'un générateur de données simulées SEIR-SW

dans un ED spatial pour les problèmes de surveillance et de suivi ainsi que de prévention des épidémies.

6.8 Conclusion

Dans ce chapitre, nous avons présenté d'une façon détaillée le système décisionnel spatiotemporel SYDSEP. Ce dernier est basé essentiellement sur un couplage de trois sous-systèmes, à savoir : système de simulation, système d'entreposage et système d'aide à la décision. Nous avons, également, décrit notre modélisation multidimensionnelle de toutes les activités du processus décisionnel en santé publique, tout en respectant la démarche décisionnelle adoptée lors de la conception multidimensionnelle d'ED.

En termes de génération des cubes spatiotemporels à partir d'ED notre système SYDSEP s'avère être un outil performant pour déterminer les foyers d'épidémie et par conséquent réagir rapidement en cas d'épidémie. Il permet une analyse multidimensionnelle des données multiples ayant un impact sur la propagation d'épidémie grippale. SYDSEP n'est pas adapté à toutes les formes de problème de propagation d'épidémie, il est donc important de préciser tout d'abord le type de l'épidémie à étudier.

Conclusion Générale

LA surveillance épidémiologique est actuellement reconnue comme une fonction essentielle de la santé publique. Son rôle est fondamental dans la planification, la détection précoce des épidémies, ainsi que l'évaluation de l'impact des programmes de prévention.

La complexité croissante de la prise de décision dans le domaine de la santé publique est un problème de plus en plus préoccupant pour les décideurs. Face aux crises sanitaires, les responsables de la santé publique doivent être prêts à agir efficacement et des politiques adéquates doivent être adoptées. Cela mène à une nécessité d'un suivi et d'un contrôle de la propagation des maladies infectieuses afin de mieux les prévenir.

A cet effet, l'objectif principal visé par le présent travail de thèse est la mise en place d'un système d'information décisionnel spatiotemporel SYDSEP pour le suivi et la prévention des épidémies. SYDSEP est basé sur l'élaboration d'un simulateur hybride SEIR-SW et son intégration dans le SID relatif au programme de lutte contre l'épidémie de la grippe saisonnière. En outre, le système proposé adopte une nouvelle démarche décisionnelle pour résoudre la problématique abordée. La démarche proposée met en évidence la distance qui sépare la problématique réelle et la présentation simplifiée, adoptée pour aider le décideur. Dans cette optique, cette thèse vise à lever un certain nombre de verrous scientifiques liés :

- d'une part, à la modélisation du phénomène de propagation d'épidémie par l'utilisation du modèle d'épidémie à compartiments SEIR et le modèle réseau social petit monde SW afin de comprendre le phénomène de transmission des virus grippaux dans la population humaine ;
- d'autre part, à l'intégration des données générées par le SEIR-SW et les données multiples obtenues de différents organismes nationaux dans un SID spatial.

La première partie de cette thèse est consacrée à l'étude des phénomènes de la dynamique de propagation d'épidémie sur les réseaux de contact. Dans le premier chapitre, nous avons présenté les concepts de base de la modélisation mathématique en épidémiologie, plus précisément, les modèles à compartiments. Les différents modèles à compartiments qui existent dans la littérature sont décrits dans le chapitre 2. Chaque modèle utilise ses propres paramètres dans une situation bien particulière.

Dans ce cadre, nous nous sommes particulièrement intéressés à la modélisation d'épidémie par l'utilisation du modèle à compartiments SEIR où il y'a une guérison permanente

d'un individu atteint de la maladie et où les infectés sont distribués d'une façon hétérogène dans l'espace comme dans le temps. Dans cette classe de modèles épidémiques, la population d'hôtes est divisée en un petit nombre de compartiments, chacun contenant des individus identifiés par leurs statuts vis à vis de la maladie. A ce niveau, nous avons considéré le cas d'une population fermée dans laquelle chacun des individus est soit susceptible, soit exposé, soit infecté, soit retiré.

À l'aide d'un réseau complexe où les nœuds représentent les individus et les liens représentent les contacts permettant la transmission de la maladie, il est possible de simuler la propagation de maladies infectieuses sur des réseaux de contacts. En effet, le couplage de SEIR avec SW contribue à une meilleure compréhension de la dynamique du processus de propagation sur les réseaux du monde réel.

Dans ce sens, nous avons proposé de nouveaux algorithmes afin de générer un réseau SW et appliquer le modèle à compartiments SEIR [Younsi *et al.*, 2015a] [Younsi *et al.*, 2014a].

Habituellement, le nombre de voisins dans un réseau SW est fixe, dans l'algorithme SEIR-SW, nous avons considéré que cette valeur suit une loi exponentielle en se basant sur des travaux de recherche actées [Beretta et Kuang, 2001], [Bonachela et Levin, 2014]. Par ailleurs, nous avons estimé les paramètres inconnus de l'épidémie grippale et de la structure sociale du modèle SEIR-SW par la proposition d'un processus d'estimation de paramètres. Pour se faire, nous avons appliqué l'algorithme de Levenberg-Marquardt qui permet d'obtenir efficacement le minimum. Le modèle SEIR-SW proposé et développé dans le chapitre 3 a été validé par des méthodes statistiques, à savoir : la corrélation linéaire et le test de Kolmogorov-Smirnov.

Pour comprendre la structure sociale de la population cible, nous avons effectué une analyse du réseau social SW. En effet, l'étude des structures sociales d'une population par les réseaux complexes permet de connaître les risques qu'une épidémie survienne. Ceci ouvre la porte à la possibilité de tester diverses stratégies de prévention ou d'intervention avant même qu'une quelconque épidémie ne soit déclarée et ainsi de permettre aux populations et aux intervenants de la santé de se préparer en conséquence. Une autre force du modèle de propagation SEIR-SW, est de permettre non seulement de représenter et comprendre la dynamique des maladies infectieuses mais il est aussi considéré comme un moteur fournissant des informations essentielles sur la dynamique de la maladie.

Dans la deuxième partie de cette thèse, nous avons décrit d'une façon détaillée le système SYDSEP. Nous avons présenté les concepts fondamentaux d'un ED, ainsi que son moteur spatial SOLAP dans le chapitre 4. Nous avons, également, détaillé les différentes architectures pour la construction d'un ED (logique et physique). Le chapitre 5 a été consacré à l'exposé d'un état de l'art sur les systèmes de surveillance d'épidémie. Nous avons présenté un panorama des principaux systèmes existant à l'échelle nationale et mondiale ainsi que les systèmes de surveillance à référence spatiale. L'utilisation du SID pour la surveillance d'épidémie n'est pas nouvelle mais peu de travaux ont abordé la problématique de surveillance et de prévention d'épidémie. Nous avons, également, complété ce

panorama par d'autres travaux de recherche. La description de SYDSEP, la construction de l'ED ainsi que son exploitation ont fait l'objet du chapitre 6. Nous avons conçu un schéma en constellation pour le système SYDSEP et nous avons présenté une description de ce schéma en décrivant ses différents composants (faits et dimensions).

Afin de montrer l'apport de SYDSEP dans le suivi, la surveillance et la prévention de propagation d'épidémie dans la région d'étude, une analyse multidimensionnelle SOLAP est primordiale. Elle permet d'effectuer une analyse spatio-temporelle du phénomène étudié par l'utilisation des requêtes plus spécialisées.

Nous avons montré à travers une revue bibliographique présentée dans le chapitre 5, l'intérêt du SOLAP en santé publique. Cette étude a démontré que les variables spatiales sont d'une grande importance dans la résolution des problèmes de santé publique en cas de survenu d'une épidémie, et que le développement d'un système d'aide à la décision basé sur une intégration du simulateur SEIR-SW dans le SID fournit aux décideurs un outil qui permet la génération de solutions faisables basées sur les différents points de vue sociaux, économiques et environnementaux.

En termes de génération des cubes à partir d'ED, notre système SYDSEP s'avère être un outil performant pour déterminer les autres facteurs favorisant la transmission d'épidémie hormis les caractéristiques de la maladie et le nombre de contacts d'un individu obtenus par SEIR-SW. Nous avons constaté que la génération des cubes à partir d'ED implémenté avec l'intégration de SEIR-SW donne de meilleurs résultats.

Perspectives

Les travaux accomplis durant cette thèse ouvrent plusieurs perspectives de travaux futurs. Les principales perspectives de recherche concernent la réutilisabilité de notre travail.

Tout d'abord, il est possible d'utiliser les automates cellulaires pour modéliser la transition des états de santé d'un individu. Les automates cellulaires fournissent une représentation simplifiée des phénomènes réels, dans lequel nous pouvons étudier la problématique des interactions entre un comportement local et un comportement global. Chaque cellule de l'automate est indépendante et réagit de façon déterministe aux conditions de son entourage. Cependant, cette formalisation ne permet pas aisément de prendre en compte des entités décisionnelles qui influencent le comportement des automates cellulaires. C'est pourquoi les systèmes multi-agents (SMA) trouvent un réel intérêt dans le domaine de la modélisation dynamique appliquée à des problématiques étudiant un système complexe.

A ce titre, nous recommanderons un couplage des automates cellulaires avec les systèmes multi-agents (SMA). Le but de ce couplage réside dans la mobilité des agents (représentant des cellules) qui peuvent ainsi transmettre la maladie ou agir sur leur environnement. Les travaux futurs peuvent aussi s'orienter vers l'utilisation du modèle de métapopulation où les nœuds représentent les communes et les liens valués représentent les flux de déplacement d'individus, tout en exploitant la notion de graphe valué au lieu de graphe non orienté utilisé dans le présent travail. Le modèle de métapopulation permet

d'identifier d'autres paramètres influant sur le mécanisme de diffusion d'épidémie liés aux déplacements des individus.

Enfin, dans ce travail nous nous sommes focalisés uniquement sur des contacts "directs", car l'épidémie étudiée (grippe saisonnière) se propage par le mode direct, il sera préférable d'étudier aussi le contact de manière « indirect » à travers des vecteurs de transmission intermédiaires (ex. mouchoir, gobelet, monnaie, etc.). Ces objets sont des vecteurs passifs dans la transmission des maladies infectieuses.

Troisième partie

Annexes

L'épidémiologie

A.1 Quelques définitions

- **Agent pathogène** : cause qui provoque ou peut provoquer directement ou indirectement une maladie.
- **Exposé** : un individu est dit exposé lorsqu'il est atteint d'une maladie infectieuse et qu'il n'est pas contagieux.
- **Germes** : virus, bactérie, parasite, champignon, prions.
- **Grippe saisonnière** : est une infection virale aiguë provoquée par un virus grippal. Les anticorps mémoire restent dans le corps après la guérison et servent à réaliser une autre réaction immunitaire plus rapide que la première en cas d'un nouvel contact avec le même antigène . Comme il a été dit, une certaine quantité d'AC reste dans le corps après guérison. Ils constituent une mémoire qui permettra une réponse (détection et multiplication) bien plus rapide lors d'une prochaine infection du même antigène ¹.
- **Infecté** : un individu qui a été atteint par la maladie et qui est contagieux.
- **Infectiosité** : propriété d'êtres infectieux ou encore capacité (quantifiée) d'un individu à infecter les autres individus qui dépend de la souche responsable de l'infection.
- **Retiré** : un individu qui a été atteint par la maladie et qui n'est plus infecté est dit retiré (immunisé ou décédé).
- **Réservoirs** : espèce(s), milieu(x) ou mécanisme(s) permettant de façon prépondérante le survie d'un agent pathogène. Par exemple : sujet malade ou sain, animaux, environnement.
- **Susceptible** : un individu qui n'a pas encore été infecté et qu'il est exposé à la maladie est dit susceptible.
- **Supports** : eau, aliments, sol, objets souillés.
- **Vecteurs** : arthropode hématophage qui assure la transmission biologique ou mécanique active d'un agent infectieux d'un vertébré à un autre vertébré. Par exemple : arthropodes, rats, chiens, insectes, etc.

1. [http : //e - sante.futura - sciences.com/ forum/ virus - grippe - guerison - anticorps.html](http://e-sante.futura-sciences.com/forum/virus-grippe-guerison-anticorps.html)

A.2 Formes épidémiologiques des maladies

Dans la littérature d'épidémiologie, ils existent trois formes épidémiologiques des maladies :

1. **Endémie** : est la présence habituelle d'une maladie dans une région.
2. **Epidémie** : développement ou propagation rapide d'une maladie infectieuse, le plus souvent par contagion, touchant simultanément un grand nombre de personnes. Ce nombre est supérieur au nombre de cas habituellement attendu.
3. **Pandémie** : on parle d'une pandémie (flambée ou bouffée épidémique) dans le cas où une vaste part de la population mondiale est touchée par un virus contre lequel il n'existe pas encore de réaction immunitaire appropriée.

Les mesures fréquentes utilisées en épidémiologie, sont : proportions, risques, taux.

A.3 Indicateurs épidémiologiques

Les indicateurs épidémiologiques permettent d'estimer l'impact d'une maladie sur population une donnée. L'identification des facteurs associés à une augmentation ou une réduction de risque de développer une maladie et l'estimation de la force de l'association sont deux objectifs principaux en épidémiologie [Moto, 1995].

Il existe deux indicateurs épidémiologiques majeurs :

1. *Les indicateurs de mortalité* : décrivent la fréquence des décès
2. *Les indicateurs de morbidité* : décrivent la fréquence des maladies (prévalence et incidence)

1. Les indicateurs de morbidité :

- **Prévalence** : est un indicateur statistique de morbidité, il se définit comme : « la proportion du nombre de cas observée à un instant donné sur la population dont sont issus ces cas ». C'est donc la proportion de sujets malades à un instant donné dans une population.

Prévalence = nombre de cas observés à un instant t / population à risque à cet instant t .

- **Incidence** : un taux qui prend en compte la vitesse de survenu de la maladie dans une population. C'est le rapport entre le nombre de nouveaux cas d'une maladie donnée survenus dans une population pendant une période définie. On utilise aussi le terme *Incidence cumulative* pour représenter la proportion de l'incidence.

Incidence = nombre de nouveaux cas observés à un instant t / population à risque à cet instant t .

2. Les indicateurs de mortalité :

- **La mortalité** : Nombre de décès survenus au cours d'une période donnée généralement une année, dans une population donnée. Le taux peut inclure toutes les causes de mortalité ou il peut être spécifique à une cause particulière.
- **Le taux de létalité** : La proportion de personnes décédées ayant une maladie particulière dans une période précise. Habituellement, la période pendant laquelle le patient est malade en raison de la maladie, cela s'applique bien à une maladie infectieuse. Comme nous pouvons citer d'autres indicateurs.

3. Autres indicateurs :

- **Taux d'attaque** : Le taux d'attaque, ou d'atteinte, indique l'incidence cumulative de l'infection sur une période. Ce terme est utilisé surtout en cas d'épidémie. Il se pourrait que la durée de la période ne soit pas indiquée.
Taux d'attaque = le nombre de cas confirmés dans une période donnée / nombre de susceptibles.
- **Le facteur de risque** : ensemble de facteurs individuels et environnementaux qui peuvent avoir une influence positive ou négative sur la santé de la population humaine.

Les facteurs climatologiques

Les variations des grands fléaux infectieux, notamment parasitaires, sont liées fortement à l'environnement. La lutte à grande échelle contre les vecteurs de ces maladies ne peut passer que par la connaissance de ces facteurs.

Aussi, les SIG sont mis à contribution pour traiter des données de provenance forte différente, afin de mieux connaître les conditions de transmission et de diffusion de ces épidémies, et donc de mieux les prévenir. Dans ce contexte, nous nous intéressons à l'étude des facteurs environnementaux favorisant la survie du virus grippale ainsi que sa propagation au sein d'une population humaine.

B.1 Description des données climatiques

Dans le cas de la grippe saisonnière, nous considérons certaines données climatiques liées à la propagation de la grippe. Ces données sont collectées du service de météorologie de la ville d'Oran :

1. **L'humidité relative** : un facteur climatique qui dépend de la quantité absolue de vapeur d'eau et de la température, il dépend aussi à l'altitude, la quantité d'eau contenue dans l'atmosphère diminue très rapidement avec l'altitude et vice versa. Contrairement à un air qui semble très humide peut contenir beaucoup moins de vapeur, s'il est à une température faible, un air qui pourra sembler très sec peut contenir de grandes quantités de vapeur, s'il est à une température élevée.
2. **L'humidité du sol** : l'humidité du sol influence les échanges de surface entre le sol, la végétation et l'atmosphère. C'est un facteur essentiel pour l'augmentation de la production agricole car il étudie la quantité d'eau stockée dans le sol qui peut produire plusieurs effets, tels que : une amélioration des rendements, moins de risques de pertes en rendements dus à la sécheresse et une recharge de la nappe phréatique. L'humidité est une grandeur sans dimension. C'est un rapport et elle est indiquée en pourcentage.
3. **La température** : la température est une grandeur physique mesurée à l'aide d'un thermomètre et étudiée en thermométrie. Elle peut être mesurée par plusieurs unités (Celsius, Fahrenheit, kelvin, etc.) la plus répandue est le **degré Celsius**. En météorologie, nous mesurons la température maximum et minimum pendant la journée et la nuit.
L'humidité relative varie en fonction de la température. Elle baisse lorsque la température s'élève et augmente lorsque la température baisse. En effet, l'influence des

variations de température sur l'humidité relative peut être très importante.

4. **La vitesse du vent** : est un flux d'air, il naît sous l'effet des différences de températures et de pression, sa direction est celle de sa provenance (nord, sud, est, ouest, etc.). Il est catégorisé selon diverses échelles, les plus connues : échelle de Beaufort et en nœuds. Dans notre base données, nous avons utilisé les données d'échelle de Beaufort illustrées par la table suivante :

L'échelle de Beaufort					
Nombre de Beaufort	Terme générique	Vitesse du vent à 10 m de hauteur			Que se passe-t-il au large ?
(degré)	(descriptif)	en m/sec	en km/h	en nœuds	(remarques)
0	Calme	moins de 1	moins de 1	moins de 1	La fumée s'élève verticalement La mer est comme un miroir
1	Très légère brise	1 à 1,59	1 à 5	1 à 3	Il se forme des rides, mais il n'y a pas d'écume
2	Légère brise	1,60 à 2,99	6 à 11	4 à 6	Vaguelettes courtes Leurs crêtes ne déferlent pas
3	Petite brise	3,00 à 5,19	12 à 19	7 à 10	Très petites vagues Écume d'aspect vitreux
4	Jolie brise	5,20 à 7,69	20 à 28	11 à 15	Petites vagues devenant plus longues Moutons nombreux
5	Bonne brise	7,70 à 10,49	29 à 38	16 à 21	Vagues modérées, allongées Moutons nombreux
6	Vent frais	10,5 à 13,59	39 à 49	22 à 26	Des lames se forment Crêtes d'écume blanche plus étendues
7	Grand frais	13,60 à 16,99	50 à 61	27 à 33	La mer grossit, l'écume est soufflée en trainées Lames déferlantes
8	Coup de vent	17,00 à 20,49	62 à 74	34 à 40	Lames de hauteur moyenne De leurs crêtes se détachent des tourbillons
9	Fort coup de vent	20,50 à 24,09	75 à 87	41 à 47	Grosses lames Leur crête s'écroule et déferle en rouleaux
10	Tempête	24,10 à 28,39	88 à 102	48 à 55	Très grosses lames à longues crêtes en panache Déferlement en rouleaux intense et brutal
11	Violente tempête	28,40 à 32,49	103 à 117	56 à 63	Lames exceptionnellement hautes Mer recouverte de bancs d'écume blanche
12	Ouragan	plus de 32,50	plus de 118	plus de 64	Air plein d'écume et d'embruns Mer entièrement blanche ; visibilité très réduite

Figure B.1 – La table d'échelle de Beaufort

L'analyse multidimensionnelle

C.1 Description du schéma MEDICAL_STATISTICS et ses dimensions

Les décideurs auront besoin d'avoir le nombre d'incidences ainsi que le taux d'attaque pour chaque commune pendant une période donnée. En outre, ces deux mesures peuvent être calculées par rapport aux tranches d'âge et type de sexe. La Figure C.1 représente un schéma composé du fait MEDICAL_STATISTICS avec ses dimensions.

Nous définissons : Un schéma de cube est un n-uplet $C_s = (cn, M, D)$, où

$cn = \text{MEDICAL_STATISTICS}$

$M = \{\text{incidence, attack_rate}\}$

$D = \{\text{Dim_sex, Dim_age_range, Dim_Disease, Dim_Month, Dim_health_care_Facility, Dim_commune, Dim_Week}\}$

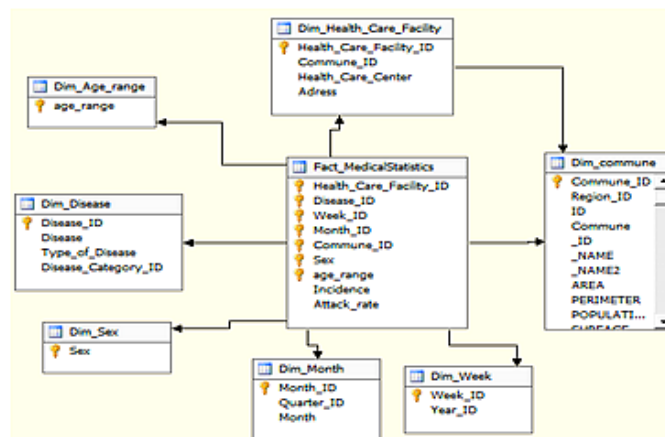


Figure C.1 – Modèle Dimensionnel en étoile du fait « MEDICAL_STATISTICS »

C.2 Description du schéma ENVIRONNEMENT et ses dimensions

La deuxième activité que nous avons jugée importante est l'activité ENVIRONNEMENT. Celle-ci calcule les différentes mesures liées aux facteurs climatiques. Si le virus grippal

survit dans un environnement favorable à sa survie et à sa transmission donc il est capable d'infecter l'hôte et se développer rapidement. En effet, il est important d'étudier l'impact de ces facteurs dans la propagation d'épidémie grippale. Dans la Figure C.2, nous représentons le schéma du fait ENVIRONMENT avec ses dimensions.

Nous définissons : Un schéma de cube est un n-uplet $C_s = (cn, M, D)$, où

$cn = \text{ENVIRONMENT}$

$M = \{High_temperature, Low_temperature, Wind_speed, Soil_moisture, Avg_temperature, RatioH/T_max, Humidity_relative\}$

$D = \{Dim_date, Dim_commune\}$

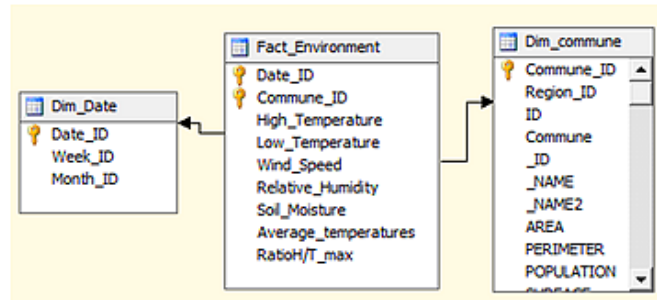


Figure C.2 – Modèle Dimensionnel en étoile de l'activité « ENVIRONMENT »

C.3 Description du schéma HEALTH_STATUS_OF_POPULATION et ses dimensions

: L'objectif de cette activité est de connaître le taux de létalité pendant une période de temps. Cette mesure aide le décideur dans son calcul de statistiques, c'est l'un des facteurs épidémiologiques. La Figure C.3 représente le schéma du fait HEALTH_STATUS_OF_POPULATION avec ses dimensions.

Un schéma de cube est un n-uplet $C_s = (cn, M, D)$, où

$cn = \text{HEALTH_STATUS_OF_POPULATION}$

$M = \{Case_fatality_rate\}$

$D = \{Dim_Week, Dim_Month, Dim_commune, Dim_R_individual\}$

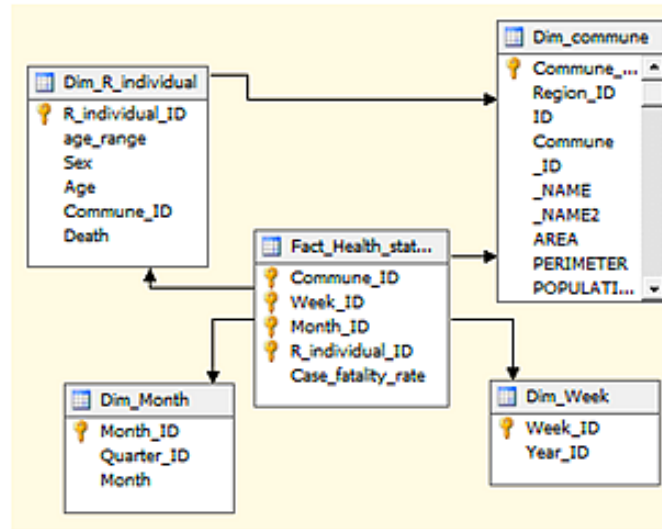


Figure C.3 – Modèle Dimensionnel en étoile de l'activité « HEALTH_STATUS_OF_POPULATION »

C.4 Description du schéma DEMOGRAPHY et ses dimensions

Dans ce fait, plusieurs facteurs liés à l'aspect démographie et à la population sont pris en considération. Dans notre étude, l'épidémie grippale a duré quelques semaines, par conséquent, les variations démographiques de la population sont négligeables. C'est pour cette raison que nous n'avons pas inclus la démographie dans le modèle SEIR-SW. Dans le cas, où nous nous intéressons à ce qui se passe à long terme, nous devons prendre en compte la démographie de la population. La Figure C.4 représente le schéma du fait DEMOGRAPHY avec ses dimensions.

Un schéma de cube est un n-uplet $C_s = (cn, M, D)$, où

$cn = \text{DEMOGRAPHY}$

$M = \{NB_Man, NB_Woman, NB_dead_M, NB_dead_W, NB_birth_M, NB_birth_F, NB_population_SC, NB_population_MT, Birth_rate, Mortality\}$

$D = \{Dim_Year, Dim_commune\}$

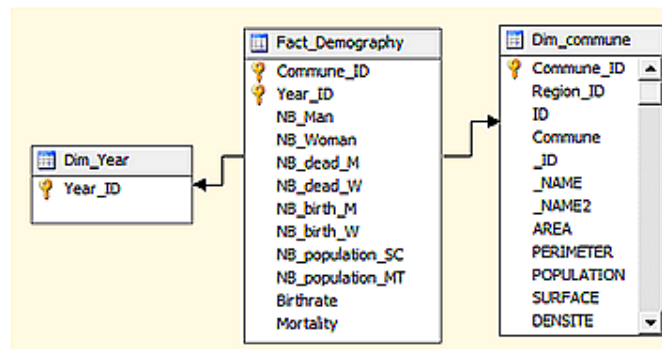


Figure C.4 – Modèle Dimensionnel en étoile de l'activité « DEMOGRAPHY »

C.5 Description du schéma SIMULATION et ses dimensions

Cette activité est très importante, les décideurs auront besoin de savoir, en cas de simulation, l'évolution et le nombre de susceptibles, d'exposés, d'infectés et de retirés dans le temps ainsi que le taux d'attaque pour chaque commune dans une période donnée. La Figure C.5 illustre le schéma du fait SIMULATION avec ses dimensions.

Un schéma de cube est un n-uplet $C_s = (cn, M, D)$, où

$cn = \text{SIMULATION}$

$M = \{NB_Susceptible, NB_exposed, NB_infected, NB_Removed, Attack_rate\}$

$D = \{Dim_Temp, Dim_SEIR_SW\}$

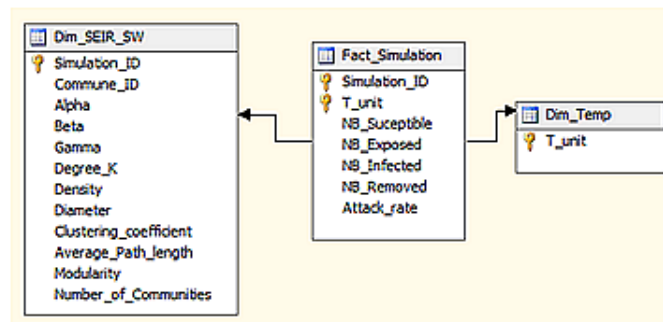


Figure C.5 – Modèle Dimensionnel en étoile de l'activité « SIMULATION »

C.6 Description du schéma STATUT et ses dimensions

Dans ce processus d'activité, il est important de connaître l'état de santé de l'individu par rapport à une simulation dans une période donnée. Nous représentons dans la Figure C.6 le schéma du fait SIMULATION avec ses dimensions.

Un schéma de cube est un n-uplet $C_s = (cn, M, D)$, où

$cn = \text{STATUT}$

$M = \{\text{Statut}\}$

$D = \{\text{Dim_Temp}, \text{Dim_SEIR_SW}, A_individual\}$

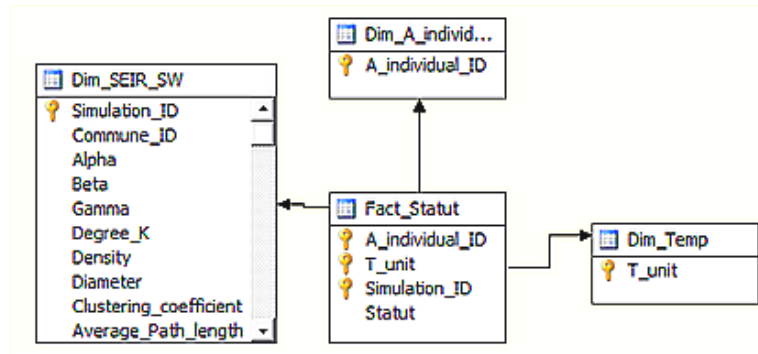


Figure C.6 – Modèle Dimensionnel en étoile de l'activité « STATUT »

Bibliographie

- [Abbey, 1952] ABBEY, H. (1952). An examination of the reed-frost theory of epidemics. *Human biology*, 24(3):201.
- [Agrawal *et al.*, 1997] AGRAWAL, R., GUPTA, A. et SARAWAGI, S. (1997). Modeling multidimensional databases. In *Data Engineering, 1997. Proceedings. 13th International Conference on*, pages 232–243. IEEE.
- [Anderson et May, 1979] ANDERSON, R. M. et MAY, R. M. (1979). Population biology of infectious diseases : Part i. *Nature*, (280):361–7.
- [Anderson *et al.*, 1992] ANDERSON, R. M., MAY, R. M. et ANDERSON, B. (1992). *Infectious diseases of humans : dynamics and control*, volume 28. Wiley Online Library.
- [Apolloni *et al.*, 2014] APOLLONI, A., POLETTI, C., RAMASCO, J. J., JENSEN, P., COLIZZA, V. *et al.* (2014). Metapopulation epidemic models with heterogeneous mixing and travel behaviour. *Theoretical Biology and Medical Modelling*, 11(1):3.
- [Axelrod, 1997] AXELROD, R. (1997). Advancing the art of simulation in the social sciences. In *Simulating social phenomena*, pages 21–40.
- [Bailey *et al.*, 1975] BAILEY, N. T. *et al.* (1975). *The mathematical theory of infectious diseases and its applications*.
- [Bakari *et al.*, 2012] BAKARI, W., ALI, M. et BEN-ABDALLAH, H. (2012). Automatic approach for generating etl operators. *arXiv preprint arXiv :1212.6051*.
- [Balcan *et al.*, 2010] BALCAN, D., GONÇALVES, B., HU, H., RAMASCO, J. J., COLIZZA, V. et VESPIGNANI, A. (2010). Modeling the spatial spread of infectious diseases : The global epidemic and mobility computational model. *Journal of computational science*, 1(3):132–145.
- [Barabási et Albert, 1999] BARABÁSI, A.-L. et ALBERT, R. (1999). Emergence of scaling in random networks. *science*, 286(5439):509–512.
- [Barreca et Shimshack, 2012] BARRECA, A. I. et SHIMSHACK, J. P. (2012). Absolute humidity, temperature, and influenza mortality : 30 years of county-level evidence from the united states. *American journal of epidemiology*, 176(suppl 7):S114–S122.
- [Barrett *et al.*, 2001] BARRETT, C., BECKMAN, R., BERKBIGLER, K., BISSET, K., BUSH, B., CAMPBELL, K., EUBANK, S., HENSON, K., HURFORD, J., KUBICEK, D. *et al.* (2001). Transims : Transportation analysis simulation system. *Los Alamos National Laboratory Unclassified Report*.
- [Barthélemy *et al.*, 2005] BARTHÉLEMY, M., BARRAT, A., PASTOR-SATORRAS, R. et VESPIGNANI, A. (2005). Dynamical patterns of epidemic outbreaks in complex heterogeneous networks. *Journal of theoretical biology*, 235(2):275–288.

- [Basileu *et al.*, 2010] BASILEU, C., BOUNEKKAR, A., KABACHI, N. et LAMURE, M. (2010). Vers un modèle de diffusion spatiale d'une pandémie. In *Colloque international Veille Stratégique Scientifique et Technologique, VSST 2010*.
- [Becker, 1979] BECKER, N. (1979). The uses of epidemic models. *Biometrics*, pages 295–305.
- [Bédard *et al.*, 1997] BÉDARD, Y., BERNIER, E., LARRIVÉE, S., NADEAU, M., PROULX, M. et RIVEST, S. (1997). Spatial olap. In *Forum annuel sur la RD, Géomatique VI : Un monde accessible*, pages 13–14.
- [Bédard *et al.*, 2005] BÉDARD, Y., PROULX, M.-J. et RIVEST, S. (2005). Enrichissement du olap pour l'analyse géographique : exemples de réalisation et différentes possibilités technologiques. In *EDA*, pages 1–20.
- [Ben Alaya *et al.*, 2008] BEN ALAYA, N., BELLALI, H., KHEDER, I., BETTAIB, J., CHLIF, S., DELLAGI, K., LOUZIR, H. et SALAH, A. B. (2008). Apports des systèmes d'information géographique dans la surveillance épidémiologique de la leishmaniose cutanée zoonotique. *Revue d'Épidémiologie et de Santé Publique*, 56(5):305.
- [Ben Messaoud, 2006] BEN MESSAOUD, R. (2006). *Couplage de l'analyse en ligne et de la fouille de données pour l'exploration, l'agrégation et l'explication des données complexes*. Thèse de doctorat, Université Lyon 2.
- [BenAmor, 2008] BENAMOR, S. (2008). *Percolation, prétopologie et multialéatoires, contributions à la modélisation des systèmes complexes : exemple du contrôle aérien*. Thèse de doctorat, École doctorale de l'École pratique des hautes études (Paris).
- [Beretta et Kuang, 1998] BERETTA, E. et KUANG, Y. (1998). Modeling and analysis of a marine bacteriophage infection. *Mathematical Biosciences*, 149(1):57–76.
- [Beretta et Kuang, 2001] BERETTA, E. et KUANG, Y. (2001). Modeling and analysis of a marine bacteriophage infection with latency period. *Nonlinear Analysis : Real World Applications*, 2(1):35–74.
- [Berndt *et al.*, 2007] BERNDT, D. J., FISHER, J. W., CRAIGHEAD, J. G., HEVNER, A. R., LUTHER, S. et STUDNICKI, J. (2007). The role of data warehousing in bioterrorism surveillance. *Decision Support Systems*, 43(4):1383–1403.
- [Bernoulli, 1760] BERNOULLI, D. (1760). Essai d'une nouvelle analyse de la mortalité causée par la petite vérole et des avantages de l'inoculation pour la prévenir. *Histoire de l'Acad. Roy. Sci.(Paris) avec Mém. des Math. et Phys. and Mém.*, pages 1–45.
- [Bertrandias et Bertrandias, 1997] BERTRANDIAS, J.-P. et BERTRANDIAS, F. (1997). *Mathématiques pour les sciences de la vie, de la nature et de la santé*.
- [Bonachela et Levin, 2014] BONACHELA, J. A. et LEVIN, S. A. (2014). Evolutionary comparison between viral lysis rate and latent period. *Journal of theoretical biology*, 345:32–42.
- [Börner *et al.*, 2007] BÖRNER, K., SANYAL, S. et VESPIGNANI, A. (2007). Network science. *Annual review of information science and technology*, 41(1):537–607.
- [Bouba *et al.*, 2013] BOUBA, F., BAH, A., NDIONE, J.-A. et NDIAYE, S. (2013). Conception d'un modèle multidimensionnel sur les maladies à vecteurs : Cas de la fièvre de la vallée du rift à barkedji (sénégal). In *COnférence sur la Modélisation Mathématique et Informatique des Systèmes Complexes, CoMMISCO 2013*.
- [Bouillot, 2011] BOUILLOT, F. (2011). Acquisition, modélisation et mise en œuvre d'un entrepôt de données pour l'analyse d'informations issues de twitter.

- [Boulil, 2012] BOULIL, K. (2012). *Une approche automatisée basée sur des contraintes d'intégrité définies en UML et OCL pour la vérification de la cohérence logique dans les systèmes SOLAP : Applications dans le domaine agri-environnemental*. Thèse de doctorat, Université Blaise Pascal-Clermont-Ferrand II.
- [Boussaid, 2015] BOUSSAID, O. (2015). Introduction aux système d'information décisionnels. *Support de cours*.
- [Brandes et Wagner, 2004] BRANDES, U. et WAGNER, D. (2004). Analysis and visualization of social networks. In *Graph drawing software*, pages 321–340.
- [Brauer, 2008] BRAUER, F. (2008). Epidemic models with heterogeneous mixing and treatment. *Bulletin of mathematical biology*, 70(7):1869–1885.
- [Capasso, 1993] CAPASSO, V. (1993). *Mathematical structures of epidemic systems*, volume 88.
- [Carrat et al., 2006] CARRAT, F., LUONG, J., LAO, H., SALLÉ, A.-V., LAJAUNIE, C. et WACKERNAGEL, H. (2006). A 'small-world-like' model for comparing interventions aimed at preventing and controlling influenza pandemics. *BMC medicine*, 4(1):26.
- [Carroll et al., 2014] CARROLL, L. N., AU, A. P., DETWILER, L. T., FU, T.-c., PAINTER, I. S. et ABERNETHY, N. F. (2014). Visualization and analytics tools for infectious disease epidemiology : A systematic review. *Journal of biomedical informatics*, 51:287–298.
- [Casagrandi et al., 2006] CASAGRANDE, R., BOLZONI, L., LEVIN, S. A. et ANDREASEN, V. (2006). The sirc model and influenza a. *Mathematical biosciences*, 200(2):152–169.
- [Castillo-Chavez, 1989] CASTILLO-CHAVEZ, C. (1989). Review of recent models of hiv/aids transmission. In *Applied Mathematical Ecology*, pages 253–262.
- [Cauchemez et al., 2008] CAUCHEMEZ, S., VALLERON, A.-J., BOELLE, P.-Y., FLAHAULT, A. et FERGUSON, N. M. (2008). Estimating the impact of school closure on influenza transmission from sentinel data. *Nature*, 452(7188):750–754.
- [Choquet et al., 2008] CHOQUET, R., DANIEL, C., BOUSSAID, O. et JAULENT, M. (2008). Etude méthodologique comparative de solutions d'entreposage de données de santé à des fins décisionnelles. In *9th International Conference on System Science in Health Care (ICSSHC 08)*.
- [Colizza et al., 2007] COLIZZA, V., PASTOR-SATORRAS, R. et VESPIGNANI, A. (2007). Reaction–diffusion processes and metapopulation models in heterogeneous networks. *Nature Physics*, 3(4):276–282.
- [Colizza et Vespignani, 2008] COLIZZA, V. et VESPIGNANI, A. (2008). Epidemic modeling in metapopulation systems with heterogeneous coupling pattern : Theory and simulations. *Journal of theoretical biology*, 251(3):450–467.
- [Cori et al., 2012] CORI, A., VALLERON, A., CARRAT, F., TOMBA, G. S., THOMAS, G. et BOËLLE, P. (2012). Estimating influenza latency and infectious period durations using viral excretion data. *Epidemics*, 4(3):132–138.
- [Cuggia et al., 2009] CUGGIA, M., BAYAT, S., GARCELON, N., SANDERS, L., ROUGET, F., COURSI, A. et PLADYS, P. (2009). A full-text information retrieval system for an epidemiological registry. *Studies in health technology and informatics*, 160(Pt 1):491–495.
- [Culpepper et Kelley, 2002] CULPEPPER, R. et KELLEY, P. (2002). Dod-global emerging infections surveillance and response system. *Navy Medicine*, 93(5):10–14.
- [Daley et al., 2001] DALEY, D. J., GANI, J. et GANI, J. M. (2001). *Epidemic modelling : an introduction*, volume 15.

- [Degenne et Forsé, 2004] DEGENNE, A. et FORSÉ, M. (2004). Les réseaux sociaux.
- [Diestel, 2005] DIESTEL, R. (2005). Graph theory. 2005. *Grad. Texts in Math.*
- [Dijkstra, 1959] DIJKSTRA, E. W. (1959). A note on two problems in connexion with graphs. *Numerische mathematik*, 1(1):269–271.
- [d’Ortenzio *et al.*, 2008] D’ORTENZIO, E., SISSOKO, D., LANDREAU, D., BENOIT-CATTIN, T., RENAULT, P. et PIERRE, V. (2008). Épidémie de rougeole à mayotte, océan indien, 2005–2006. *Médecine et maladies infectieuses*, 38(11):601–607.
- [Dreyer et Gabriel, 2010] DREYER, N. et GABRIEL, J.-P. (2010). Daniel bernouilli et la variole1.
- [Dufour et Hendriks, 2007] DUFOUR, B. et HENDRIKS, P. (2007). Surveillance épidémiologique en santé animale. 2 éd. *AEEMA et Quae, Paris*.
- [Dufour *et al.*, 2006] DUFOUR, B., HENDRIKS, P. et TOMA, B. (2006). Élaboration et mise en place de systèmes de surveillance épidémiologique des maladies à haut risque dans les pays développés. *Rev. sci. tech. Off. int. Epiz*, 25(1):187–198.
- [Ebel *et al.*, 2003] EBEL, H., DAVIDSEN, J. et BORNHOLDT, S. (2003). Dynamics of social networks. *Complexity*, 8(2):24–27.
- [El Akkaoui et Zimányi, 2009] EL AKKAOU, Z. et ZIMÁNYI, E. (2009). Defining etl workflows using bpmn and bpel. In *Proceedings of the ACM twelfth international workshop on Data warehousing and OLAP*, pages 41–48.
- [Erdős et Rényi, 1959] ERDÖS, P. et RÉNYI, A. (1959). On random graphs i. *Publicationes Mathematicae*, 6:290–297.
- [Eubank *et al.*, 2004] EUBANK, S., GUCLU, H., KUMAR, V. A., MARATHE, M. V., SRINIVASAN, A., TOROCZKAI, Z. et WANG, N. (2004). Modelling disease outbreaks in realistic urban social networks. *Nature*, 429(6988):180–184.
- [Euler, 1741] EULER, L. (1741). Solutio problematis ad geometriam situs pertinentis. *Commentarii academiae scientiarum Petropolitanae*, 8:128–140.
- [Favier *et al.*, 2005] FAVIER, C., SCHMIT, D., MÜLLER-GRAF, C. D., CAZELLES, B., DEGALLIER, N., MONDET, B. et DUBOIS, M. A. (2005). Influence of spatial heterogeneity on an emerging infectious disease : the case of dengue epidemics. *Proceedings of the Royal Society of London B : Biological Sciences*, 272(1568):1171–1177.
- [Ferguson *et al.*, 2005] FERGUSON, N. M., CUMMINGS, D. A., CAUCHEMEZ, S., FRASER, C., RILEY, S., MEEYAI, A., IAMSIRITHAWORN, S. et BURKE, D. S. (2005). Strategies for containing an emerging influenza pandemic in southeast asia. *Nature*, 437(7056):209–214.
- [Few, 2004] FEW, S. (2004). Dashboard confusion. *Perceptual Edge*.
- [Few, 2007] FEW, S. (2007). Dashboard design for real-time situation awareness. *Inova Solutions*.
- [Filipe et Maule, 2004] FILIPE, J. et MAULE, M. (2004). Effects of dispersal mechanisms on spatio-temporal development of epidemics. *Journal of theoretical biology*, 226(2):125–141.
- [Flahault *et al.*, 1994] FLAHAULT, A., DEGUEN, S. et VALLERON, A.-J. (1994). A mathematical model for the european spread of influenza. *European Journal of Epidemiology*, 10(4):471–474.
- [Flahault *et al.*, 1988] FLAHAULT, A., LETRAIT, S., BLIN, P., HAZOUT, S., MENARES, J. et VALLERON, A.-J. (1988). Modelling the 1985 influenza epidemic in france. *Statistics in medicine*, 7(11):1147–1155.

- [Flasche *et al.*, 2011] FLASCHE, S., HENS, N., BOËLLE, P.-Y., MOSSONG, J., van BALLEGOOIJEN, W. M., NUNES, B., RIZZO, C., POPOVICI, F., SANTA-OLALLA, P., HRUBÁ, F. *et al.* (2011). Different transmission patterns in the early stages of the influenza a (h1n1) v pandemic : a comparative analysis of 12 european countries. *Epidemics*, 3(2):125–133.
- [Fuentes et Kuperman, 1999] FUENTES, M. et KUPERMAN, M. (1999). Cellular automata and epidemiological models with spatial dependence. *Physica A : Statistical Mechanics and its Applications*, 267(3):471–486.
- [Garba *et al.*, 2013] GARBA, S. M., SAFI, M. A. et GUMEL, A. (2013). Cross-immunity-induced backward bifurcation for a model of transmission dynamics of two strains of influenza. *Nonlinear Analysis : Real World Applications*, 14(3):1384–1403.
- [Gattepaille et Jay, 2007] GATTEPAILLE, L. et JAY, F. (2007). Modèle de propagation d'épidémie avec contraintes spatiales. *Projet de Spécialité, ENSIMAG*.
- [Gaudart, 2007] GAUDART, J. (2007). *Analyse spatio-temporelle et modélisation des épidémies : application au paludisme à P. falciparum*. Thèse de doctorat, Aix Marseille 2.
- [Gilbert et Pfeiffer, 2012] GILBERT, M. et PFEIFFER, D. U. (2012). Risk factor modelling of the spatio-temporal patterns of highly pathogenic avian influenza (hpaiv) h5n1 : a review. *Spatial and spatio-temporal epidemiology*, 3(3):173–183.
- [Gill *et al.*, 1981] GILL, P., MURRAY, W. et WRIGHT, M. (1981). The levenberg-marquardt method. *Practical optimization*, pages 136–137.
- [González-Parra *et al.*, 2015] GONZÁLEZ-PARRA, G., VILLANUEVA, R.-J., RUIZ-BARAGANO, J. et MORAÑO, J.-A. (2015). Modelling influenza a (h1n1) 2009 epidemics using a random network in a distributed computing environment. *Acta tropica*, 143:29–35.
- [Goodchild, 2000] GOODCHILD, M. F. (2000). New horizons for the social sciences : geographic information systems. *Social Sciences for a Digital World Building Infrastructure and Databases for the Future : Building Infrastructure and Databases for the Future*, page 173.
- [Grais *et al.*, 2003] GRAIS, R. F., ELLIS, J. H. et GLASS, G. E. (2003). Assessing the impact of airline travel on the geographic spread of pandemic influenza. *European journal of epidemiology*, 18(11):1065–1072.
- [Gray *et al.*, 1997] GRAY, J., CHAUDHURI, S., BOSWORTH, A., LAYMAN, A., REICHART, D., VENKATRAO, M., PELLOW, F. et PIRAHESH, H. (1997). Data cube : A relational aggregation operator generalizing group-by, cross-tab, and sub-totals. *Data Mining and Knowledge Discovery*, 1(1):29–53.
- [Grenfell et Harwood, 1997] GRENFELL, B. et HARWOOD, J. (1997). (meta) population dynamics of infectious diseases. *Trends in ecology & evolution*, 12(10):395–399.
- [Guégan et Choisy, 2008] GUÉGAN, J.-F. et CHOISY, M. (2008). *Introduction à l'épidémiologie intégrative*. De Boeck Supérieur.
- [Guigal, 1995] GUIGAL, P. (1995). *Modélisation de la propagation infectieuse dans un réseau organisé d'individus. Apport de la prétopologie et de la géométrie fractale*. Thèse de doctorat, Institut national des sciences appliquées de Lyon.
- [Hamer, 1906] HAMER, W. H. (1906). *The Milroy lectures on epidemic disease in England : the evidence of variability and of persistency of type*.
- [Hanafi-Bojd *et al.*, 2012] HANAFI-BOJD, A., VATANDOOST, H., OSHAGHI, M., CHARRAHY, Z., HAGHDOOST, A., ZAMANI, G., ABEDI, F., SEDAGHAT, M., SOLTANI, M.,

- SHAHI, M. *et al.* (2012). Spatial analysis and mapping of malaria risk in an endemic area, south of iran : a gis based decision making for planning of control. *Acta tropica*, 122(1):132–137.
- [Harris, 1974] HARRIS, T. E. (1974). Contact interactions on a lattice. *The Annals of Probability*, pages 969–988.
- [Hethcote, 1988] HETHCOTE, H. W. (1988). Optimal ages of vaccination for measles. *Mathematical Biosciences*, 89(1):29–52.
- [Hethcote, 2000] HETHCOTE, H. W. (2000). The mathematics of infectious diseases. *SIAM review*, 42(4):599–653.
- [Hooten *et al.*, 2010] HOOTEN, M. B., ANDERSON, J. et WALLER, L. A. (2010). Assessing north american influenza dynamics with a statistical sirs model. *Spatial and spatio-temporal epidemiology*, 1(2):177–185.
- [Huang *et al.*, 2010] HUANG, G., TAKEUCHI, Y., MA, W. et WEI, D. (2010). Global stability for delay sir and seir epidemic models with nonlinear incidence rate. *Bulletin of mathematical biology*, 72(5):1192–1207.
- [Inmon, 1996] INMON, W. H. (1996). The data warehouse and data mining. *Communications of the ACM*, 39(11):49–50.
- [INSP, 2010] INSP (2010). Bilan de la saison (2009-2010) du réseau sentinelle algérien de surveillance de la grippe saisonnière. *Institut National de Santé Publique*.
- [Keeling *et al.*, 2004] KEELING, M., BJØRNSTAD, O. et GRENFELL, B. (2004). Metapopulation dynamics of infectious diseases. *Ecology, genetics, and evolution of metapopulations*. Elsevier, Amsterdam, pages 415–445.
- [Kerkri *et al.*, 1998] KERKRI, E., QUANTIN, C., YETONGNON, K., ALLAERT, F. et DUS-SERRE, L. (1998). Application of the medical data warehousing architecture epidware to epidemiological follow-up : data extraction and transformation. *Studies in health technology and informatics*, 68:414–418.
- [Kermack et McKendrick, 1927] KERMAK, W. O. et MCKENDRICK, A. G. (1927). A contribution to the mathematical theory of epidemics. *In Proceedings of the Royal Society of London A : Mathematical, Physical and Engineering Sciences*, volume 115, pages 700–721. The Royal Society.
- [Khormi et Kumar, 2011] KHORMI, H. M. et KUMAR, L. (2011). Modeling dengue fever risk based on socioeconomic parameters, nationality and age groups : Gis and remote sensing based case study. *Science of the Total Environment*, 409(22):4713–4719.
- [Kimball, 1996] KIMBALL, R. (1996). The data warehouse toolkit : practical techniques for building dimensional data warehouse. NY : *John Willey & Sons*, 248:4.
- [Kimball *et al.*, 2005] KIMBALL, R., REEVES, L. et ROSS, M. (2005). *Le data warehouse : Guide de conduite de projet*. Eyrolles.
- [Kimball *et al.*, 2002] KIMBALL, R., ROSS, M. *et al.* (2002). The data warehouse toolkit : the complete guide to dimensional modelling. *Nachdr.* New York [ua] : Wiley.
- [Kleinberg, 2000] KLEINBERG, J. (2000). The small-world phenomenon : An algorithmic perspective. *In Proceedings of the thirty-second annual ACM symposium on Theory of computing*, pages 163–170.
- [Kristensen, 2004] KRISTENSEN, M. R. (2004). Parameter estimation in nonlinear dynamical systems. Mémoire de D.E.A., Department of Chemical Engineering, Technical University of Denmark.

- [Kroenke et Auer, 2011] KROENKE, D. M. et AUER, D. J. (2011). Database processing : fundamentals, design, and implementation. *Pearson Education*.
- [Lahleb, 2015] LAHLEB, C. (2015). Conception et élaboration d'un système décisionnel dédié à la surveillance épidémique. Mémoire de D.E.A., Université d'Oran.
- [Landais *et al.*, 2002] LANDAIS, P., SIMONET, A., GUILLON, D., JACQUELINET, C., SAÏD, M. B., MUGNIER, C. et SIMONET, M. (2002). Sims@ rein : Un système d'information multi-sources pour l'insuffisance rénale terminale. *Comptes Rendus Biologies*, 325(4): 515–528.
- [Langlois *et al.*, 2013] LANGLOIS, P., BLANPAIN, B. et DAUDÉ, E. (2013). Magéo, une plateforme de simulation multi-agents pour tous. *In SimTools 2013*.
- [Lebhar, 2005] LEBHAR, E. (2005). *Algorithmes de routage et modeles aléatoires pour les graphes petits mondes*. Thèse de doctorat, Ecole normale supérieure de lyon-ENS LYON.
- [Legrand *et al.*, 2004] LEGRAND, J., VIBOUD, C., BOELLE, P.-Y., VALLERON, A.-J. et FLAHAULT, A. (2004). Modelling responses to a smallpox epidemic taking into account uncertainty. *Epidemiology and infection*, 132(01):19–25.
- [Levenberg, 1944] LEVENBERG, K. (1944). A method for the solution of certain non-linear problems in least squares.
- [Levins, 1969] LEVINS, R. (1969). Some demographic and genetic consequences of environmental heterogeneity for biological control. *Bulletin of the Entomological Society of America*, 15(3):237–240.
- [Li et Mackaness, 2014] LI, S. et MACKANESS, W. A. (2014). A multi-agent-based, semantic-driven system for decision support in epidemic management. *Health informatics journal*.
- [Liccardo et Fierro, 2013] LICCARDO, A. et FIERRO, A. (2013). A lattice model for influenza spreading.
- [Longini, 1988] LONGINI, I. M. (1988). A mathematical model for predicting the geographic spread of new infectious agents. *Mathematical Biosciences*, 90(1):367–383.
- [Longini *et al.*, 2005] LONGINI, I. M., NIZAM, A., XU, S., UNGCHUSAK, K., HANSHAWORAKUL, W., CUMMINGS, D. A. et HALLORAN, M. E. (2005). Containing pandemic influenza at the source. *Science*, 309(5737):1083–1087.
- [Loth *et al.*, 2010] LOTH, L., GILBERT, M., OSMANI, M. G., KALAM, A. M. et XIAO, X. (2010). Risk factors and clusters of highly pathogenic avian influenza h5n1 outbreaks in bangladesh. *Preventive veterinary medicine*, 96(1):104–113.
- [MacDonald *et al.*, 2004] MACDONALD, P., GREEN, A., COTRELL, T., WEBBER, D. et BURR, P. (2004). Developments in near real-time surveillance : the portable remote illness and symptom monitor (prism). *XXXV International Congress on Military Medicine*.
- [Marcel, 1998] MARCEL, P. (1998). *Manipulations de données multidimensionnelles et langages de règles*. Thèse de doctorat, Institut National des Sciences Appliquées de Lyon.
- [Marder, 2007] MARDER, M. (2007). Dynamics of epidemics on random networks. *Physical Review E*, 75(6):066103.
- [Marquardt, 1963] MARQUARDT, D. (1963). An algorithm for least-squares estimation of nonlinear parameters. *Journal of the Society for Industrial & Applied Mathematics*, 11(2):431–441.

- [May, 1997] MAY, R. M. (1997). The scientific wealth of nations. *Science*, 275(5301):793.
- [Mejhed *et al.*, 2009] MEJHED, H., BOUSSAA, S. et el houda MEJHED, N. (2009). Data warehouse and the deployment of data mining process to make decision for leishmaniasis and development of mathematical models for phlebotomine sandflies seasonality in marrakech city. *WSEAS Transactions on. Biology and Biomedicine*, 6(1).
- [Melnik et Korobeinikov, 2013] MELNIK, A. V. et KOROBEINIKOV, A. (2013). Lyapunov functions and global stability for sir and seir models with age-dependent susceptibility. *Math. Biosci. Eng*, 10(2):369–378.
- [Meloni *et al.*, 2011] MELONI, S., PERRA, N., ARENAS, A., GÓMEZ, S., MORENO, Y. et VESPIGNANI, A. (2011). Modeling human mobility responses to the large-scale spreading of infectious diseases. *Scientific reports*, 1.
- [Meyers *et al.*, 2006] MEYERS, L. A., NEWMAN, M. et POURBOHLOUL, B. (2006). Predicting epidemics on directed contact networks. *Journal of theoretical biology*, 240(3):400–418.
- [Meynard *et al.*, 2008] MEYNARD, J.-B., CHAUDET, H., TEXIER, G., QUEYRIAUX, B., DEPARIS, X. et BOUTIN, J.-P. (2008). Surveillance épidémiologique en temps réel dans les armées ; concepts, réalités et perspectives en france. *Revue d'Épidémiologie et de Santé Publique*, 56(1):11–20.
- [Meynard *et al.*, 2004] MEYNARD, J.-B., TEXIER, G., SBAI IDRISSE, K., OLLIVIER, L., MICHEL, R., GAUDRY, M., ROGIER, C., MIGLIANI, R., SPIEGEL, A. et BOUTIN, J.-P. (2004). Surveillance épidémiologique en temps réel pour les armées. *Médecine et armées*, 32(4):360–365.
- [Midouni *et al.*, 2009] MIDOUNI, S. A. D., DARMONT, J. et BENTAYEB, F. (2009). Approche de modélisation multidimensionnelle des données complexes : Application aux données médicales. In *5èmes Journées francophones sur les Entrepôts de Données et l'Analyse en ligne (EDA 09)*, pages 155–166. Cépaduès.
- [Milgram, 1963] MILGRAM, S. (1963). Behavioral study of obedience. *The Journal of Abnormal and Social Psychology*, 67(4):371.
- [Miller, 2011] MILLER, J. C. (2011). A note on a paper by erik volz : Sir dynamics in random networks. *Journal of mathematical biology*, 62(3):349–358.
- [Molloy et Reed, 1995] MOLLOY, M. et REED, B. A. (1995). A critical point for random graphs with a given degree sequence. *Random structures and algorithms*, 6(2/3):161–180.
- [Moncion, 2008] MONCION, T. (2008). *Modélisation de la complexité et de la dynamique des simulations multi-agents : application pour l'analyse des phénomènes émergents*. Thèse de doctorat, Evry-Val d'Essonne.
- [Moto, 1995] MOTO, L. (1995). *Santé et multidisciplinarité : choix et décisions*. Hermès.
- [Nagel *et al.*, 2013] NAGEL, A. C., TSOU, M.-H., SPITZBERG, B. H., AN, L., GAWRON, J. M., GUPTA, D. K., YANG, J.-A., HAN, S., PEDDECORD, K. M., LINDSAY, S. *et al.* (2013). The complex relationship of realspace events and messages in cyberspace : case study of influenza and pertussis using tweets. *Journal of medical Internet research*, 15(10).
- [Naqvi *et al.*, 2015] NAQVI, S. A. A., KAZMI, S. J. H., SHAIKH, S. et AKRAM, M. (2015). Evaluation of prevalence patterns of dengue fever in lahore district through geo-spatial techniques. *Journal of Basic and Applied Sciences*, 11:20–30.

- [Neuman et Mizruchi, 2010] NEUMAN, E. J. et MIZRUCHI, M. S. (2010). Structure and bias in the network autocorrelation model. *Social Networks*, 32(4):290–300.
- [Newman, 2002] NEWMAN, M. E. (2002). Spread of epidemic disease on networks. *Physical review E*, 66(1):016128.
- [Newman et Girvan, 2004] NEWMAN, M. E. et GIRVAN, M. (2004). Finding and evaluating community structure in networks. *Physical review E*, 69(2):026113.
- [Newman et Watts, 1999] NEWMAN, M. E. et WATTS, D. J. (1999). Renormalization group analysis of the small-world network model. *Physics Letters A*, 263(4):341–346.
- [Njagarah et Nyabadza, 2014] NJAGARAH, J. et NYABADZA, F. (2014). A metapopulation model for cholera transmission dynamics between communities linked by migration. *Applied Mathematics and Computation*, 241:317–331.
- [Nkoghe Mba *et al.*, 2005] NKOGHE MBA, D., FORMENTY, P., LEROY, E., NNEGUE, S., OBAME EDPU, S., BA, I., ALLARANGAR, Y., CABORE, J., BACHY, C., ANDRAGHETTI, R. *et al.* (2005). Plusieurs épidémies de fièvre hémorragique due au virus ebola au gabon, d’octobre 2001 à avril 2002. *Bulletin de la Société de pathologie exotique*, 98(3).
- [Noël *et al.*, 2009] NOËL, P.-A., DAVOUDI, B., BRUNHAM, R. C., DUBÉ, L. J. et POURBOHLOUL, B. (2009). Time evolution of epidemic disease on finite and infinite networks. *Physical Review E*, 79(2):026101.
- [Ochab et Góra, 2011] OCHAB, J. et GÓRA, P. (2011). Shift of percolation thresholds for epidemic spread between static and dynamic small-world networks. *The European Physical Journal B-Condensed Matter and Complex Systems*, 81(3):373–379.
- [Omumbo *et al.*, 2005] OMUMBO, J., HAY, S., SNOW, R., TATEM, A. et ROGERS, D. (2005). Modelling malaria risk in east africa at high-spatial resolution. *Tropical Medicine & International Health*, 10(6):557–566.
- [Opatowski *et al.*, 2013] OPATOWSKI, L., de CELLÈS, M. D., SOUISSI, S., KARDAS-SŁOMA, L., TEMIME, L. et GUILLEMOT, D. (2013). Contribution des modèles mathématiques à la compréhension de la dynamique de diffusion des bactéries multi-résistantes à l’hôpital. *Journal des Anti-infectieux*, 15(4):193–203.
- [Ortiz-Pelaez *et al.*, 2006] ORTIZ-PELAEZ, A., PFEIFFER, D., SOARES-MAGALHAES, R. et GUITIAN, F. (2006). Use of social network analysis to characterize the pattern of animal movements in the initial phases of the 2001 foot and mouth disease (fmd) epidemic in the uk. *Preventive veterinary medicine*, 76(1):40–55.
- [Pastor-Satorras et Vespignani, 2001] PASTOR-SATORRAS, R. et VESPIGNANI, A. (2001). Epidemic spreading in scale-free networks. *Physical review letters*, 86(14):3200.
- [Pavé, 1994] PAVÉ, A. (1994). *Modélisation en biologie et en écologie*. Aléas.
- [Peguiron et Thiery, 2005] PEGUIRON, F. et THIERY, O. (2005). A propos d’un entrepôt de données universitaire : modélisation des acteurs et méta données. *In 1ère Journée Francophone sur les Entrepôts de Données et l’Analyse en Ligne-EDA’2005*.
- [Proulx *et al.*, 2002] PROULX, M.-J., BÉDARD, Y., NADEAU, M., GOSSELIN, P. et LEBEL, G. (2002). Géomatique et santé environnementale : innovations résultant du projet icem/se. *Géomatique 2002*.
- [Pusceddu et Chakhar, 2005] PUSCEDDU, C. et CHAKHAR, S. (2005). Un processus pour la prise de décision spatiale. *ROADEF’05, Tours, France*, pages 14–16.
- [Rakotomanana, 2009] RAKOTOMANANA, F. (2009). *Apport d’un système d’information géographique et de la télédétection dans la prévention du risque de survenue d’épidémie de paludisme sur les Hautes Terres Centrales de Madagascar*. Thèse de doctorat, Paris Est.

- [Ramírez-Ramírez *et al.*, 2013] RAMÍREZ-RAMÍREZ, L. L., GEL, Y. R., THOMPSON, M., de VILLA, E. et MCPHERSON, M. (2013). A new surveillance and spatio-temporal visualization tool simid : Simulation of infectious diseases using random networks and gis. *Computer methods and programs in biomedicine*, 110(3):455–470.
- [Rechenmann, 2010] RECHENMANN, F. (2010). Systèmes dynamiques et équations différentielles. *Interstices*.
- [Reluga *et al.*, 2006] RELUGA, T. C., MEDLOCK, J. et GALVANI, A. P. (2006). A model of spatial epidemic spread when individuals move within overlapping home ranges. *Bulletin of mathematical biology*, 68(2):401–416.
- [Rivest *et al.*, 2005] RIVEST, S., BÉDARD, Y., PROULX, M.-J., NADEAU, M., HUBERT, F. et PASTOR, J. (2005). Solap technology : Merging business intelligence with geospatial technology for interactive spatio-temporal exploration and analysis of data. *ISPRS journal of photogrammetry and remote sensing*, 60(1):17–33.
- [Rizzo *et al.*, 2008] RIZZO, C., LUNELLI, A., PUGLIESE, A., BELLA, A., MANFREDI, P., TOMBA, G. S., IANNELLI, M. et DEGLI ATTI, M. C. (2008). Scenarios of diffusion and control of an influenza pandemic in italy. *Epidemiology and infection*, 136(12):1650–1657.
- [Rvachev et Longini, 1985] RVACHEV, L. A. et LONGINI, I. M. (1985). A mathematical model for the global spread of influenza. *Mathematical biosciences*, 75(1):3–22.
- [Sallet, 2010] SALLET, G. (2010). Inria & ird epicasa09 avril 2010. Rapport technique, •.
- [Sander *et al.*, 2002] SANDER, L., WARREN, C., SOKOLOV, I., SIMON, C. et KOOPMAN, J. (2002). Percolation on disordered networks as a model for epidemics. *Math. Biosci*, 180:293–305.
- [Santos *et al.*, 2013] SANTOS, R., MALHEIROS, S., CAVALHEIRO, S. et DE OLIVEIRA, J. P. (2013). A data mining system for providing analytical information on brain tumors to public health decision makers. *Computer methods and programs in biomedicine*, 109(3):269–282.
- [Saramäki et Kaski, 2005] SARAMÄKI, J. et KASKI, K. (2005). Modelling development of epidemics with dynamic small-world networks. *Journal of Theoretical Biology*, 234(3):413–421.
- [Sauvage et Pontier, 2005] SAUVAGE, F. et PONTIER, D. (2005). Interêts des modèles déterministes et stochastiques en épidémiologie des maladies infectieuses : exemple du hantavirus puumala. *Epidémiologie et Santé Animale*, 47:63–82.
- [Sharomi et Gumel, 2011] SHAROMI, O. et GUMEL, A. (2011). Dynamical analysis of a sex-structured chlamydia trachomatis transmission model with time delay. *Nonlinear Analysis : Real World Applications*, 12(2):837–866.
- [Simoës, 2007] SIMOËS, J. M. A. (2007). *An agent-based approach to spatial epidemics through GIS*. Thèse de doctorat, University of London.
- [Stattner, 2012] STATTNER, E. (2012). *Contributions à l'étude des réseaux sociaux : propagation, fouille, collecte de données*. Thèse de doctorat, Université des Antilles-Guyane.
- [Tibiche, 2011] TIBICHE, A. (2011). La surveillance épidémiologique. http://www.ummto.dz/IMG/pdf/La_Surveillance_Epidemiologique_cle4e19c4.pdf.
- [Toma *et al.*, 1991] TOMA, B., BÉNET, J.-J., DUFOUR, B., ELOIT, M., MOUTOU, F., SANNA, M. *et al.* (1991). *Glossary of animal epidemiology*. Editions du Point Vétérinaire, 25 rue de Bourgelat.

- [Tran *et al.*, 2004] TRAN, A., DEPARIS, X., DUSSART, P., MORVAN, J., RABARISON, P., REMY, F., POLIDORI, L. et GARDON, J. (2004). Dengue spatial and temporal patterns, french guiana, 2001. *Emerg Infect Dis*, 10(4):615–621.
- [Turbelin et Boëlle, 2010] TURBELIN, C. et BOËLLE, P.-Y. (2010). Improving general practice based epidemiologic surveillance using desktop clients : the french sentinel network experience. *Studies in health technology and informatics*, 160(Pt 1):442.
- [Vazquez, 2006] VAZQUEZ, A. (2006). Spreading dynamics on small-world networks with connectivity fluctuations and correlations. *Physical Review E*, 74(5):056101.
- [Volz, 2008] VOLZ, E. (2008). Sir dynamics in random networks with heterogeneous connectivity. *Journal of mathematical biology*, 56(3):293–310.
- [Wang *et al.*, 2014] WANG, J., WANG, J., LIU, M. et LI, Y. (2014). Global stability analysis of an sir epidemic model with demographics and time delay on networks. *Physica A : Statistical Mechanics and its Applications*, 410:268–275.
- [Watts et Strogatz, 1998] WATTS, D. J. et STROGATZ, S. H. (1998). Collective dynamics of ‘small-world’ networks. *nature*, 393(6684):440–442.
- [Wen et Yang, 2008] WEN, L. et YANG, X. (2008). Global stability of a delayed sirs model with temporary immunity. *Chaos, Solitons & Fractals*, 38(1):221–226.
- [West et Thompson, 1997] WEST, R. W. et THOMPSON, J. R. (1997). Models for the simple epidemic. *Mathematical biosciences*, 141(1):29–39.
- [Wilson *et al.*, 2004] WILSON, J., CAREW, M. et STRAUSS, B. (2004). Canadian forces evaluation of the epinato health surveillance system in bosnia-herzegovina. *Military medicine*, 171(10):955.
- [Wu *et al.*, 2006] WU, J. T., RILEY, S., FRASER, C. et LEUNG, G. M. (2006). Reducing the impact of the next influenza pandemic using household-based public health interventions. *PLoS Med*, 3(9):e361.
- [Yan *et al.*, 2006] YAN, P., ZENG, D. et CHEN, H. (2006). A review of public health syndromic surveillance systems. In *Intelligence and Security Informatics*, pages 249–260.
- [Younsi *et al.*, 2014a] YOUNSI, F. Z., BOUNEKKAR, A. et HAMDADOU, D. (2014a). SEIR-SW : Un modèle de suivi de propagation d’épidémies. Hammamet, Tunisie.
- [Younsi *et al.*, 2015a] YOUNSI, F. Z., BOUNEKKAR, A., HAMDADOU, D. et BOUSSAID, O. (2015a). SEIR-SW, simulation model of influenza spread based on the small world network. *Tsinghua Science and Technology*, 20(5).
- [Younsi *et al.*, 2014b] YOUNSI, F. Z., HAMDADOU, D. et BOUSSAID, O. (2014b). Towards a spatiotemporal decision support system for epidemiological surveillance. Oran, Algérie.
- [Younsi *et al.*, 2015b] YOUNSI, F. Z., HAMDADOU, D. et BOUSSAID, O. (2015b). A surveillance and spatiotemporal visualization model for infectious diseases using social network. *International Journal of Decision Support System Technology*, 7(4).
- [Zekri et Clerc, 2002] ZEKRI, N. et CLERC, J. P. (2002). Étude statistique et dynamique de la propagation d’épidémies dans un réseau de petit monde. *Comptes Rendus Physique*, 3(6):741–747.
- [Zemri *et al.*, 2015] ZEMRI, F. A., HAMDADOU, D. et ZEITOUNI, K. (2015). Vers un système d’aide à la décision multicritères et spatiotemporel pour la surveillance épidémiologique. *GAST–Gestion et Analyse de données Spatiales et Temporelles*.

- [Zhou *et al.*, 2010] ZHOU, X., CHEN, S., LIU, B., ZHANG, R., WANG, Y., LI, P., GUO, Y., ZHANG, H., GAO, Z. et YAN, X. (2010). Development of traditional chinese medicine clinical data warehouse for medical knowledge discovery and decision support. *Artificial Intelligence in Medicine*, 48(2):139–152.

Résumé

Les maladies infectieuses représentent aujourd'hui un problème majeur de santé publique. Devant l'augmentation des résistances bactériennes, l'émergence de nouveaux pathogènes et la propagation rapide de l'épidémie, le suivi et la surveillance de la transmission de la maladie devient particulièrement importants. Face à une telle menace, la société doit se préparer à l'avance pour réagir rapidement et efficacement si une telle épidémie est déclarée. Cela nécessite une mise en place des dispositifs de suivi et de prévention.

Dans ce contexte, nous nous intéressons, dans le présent travail, à l'élaboration d'un Système d'Information Décisionnel Spatio-temporel pour le suivi et la surveillance du phénomène de propagation de l'épidémie de la grippe saisonnière au sein de la population de la ville d'Oran (Algérie). L'objectif de ce système est double : il consiste, d'une part, à comprendre comment l'épidémie se propage par l'utilisation du réseau social Small World (SW) et du modèle à compartiments d'épidémie SEIR (Susceptible-Exposed-Infected-Removed), et d'autre part, à stocker dans un entrepôt les données multiples tout en les analysant par un outil d'analyse en ligne de donnée Spatiale dit SOLAP (Spatial On-Line Analytical Processing).

Mots clés : Modèle à compartiments SEIR (Susceptible-Exposed-Infected-Removed), Réseau social Small World (SW), Modèle hybride SEIR-SW, Entrepôt de données, Spatial OLAP, Simulation, Système d'information à référence spatiale, Analyse du réseau social.

Abstract

Today, infectious diseases represent a major public health problem. With the increase of bacterial resistance, the emergence of new pathogens and the rapid spread of epidemic, monitoring and surveillance of disease transmission becomes important. In the face of such a threat, the society must prepare in advance to respond quickly and effectively if an outbreak is declared. This requires setting up monitoring mechanisms and prevention.

In this context, we are particularly interested by development a Spatiotemporal decision support system for monitoring and preventing the phenomenon of seasonal influenza epidemic spread in the population of Oran (city at Algeria). The objective of this system is twofold : on one hand, to understand how epidemic is spreading through the social network by using SEIR (Susceptible-Exposed-Infected-Removed) compartmental model within Small World network, and on the other hand, to store multiple data in data warehouse and analyzing it by a specific online analysis tool Spatial OLAP (Spatial on-line Analytical Processing).

Keywords : Compartmental SEIR (Susceptible-Exposed-Infected-Removed) model, Small World social network (SW), SEIR-SW hybrid model, Data warehouse, Spatial OLAP, Simulation, Spatial Reference System, Social network analysis.