



**HAL**  
open science

# Multi-scale-socio-environmental modeling of epidemiological process : a way for organizing humain environments and rythms to control and prevent the spread of contagious diseases

Mohammad Hessem Hessami

► **To cite this version:**

Mohammad Hessem Hessami. Multi-scale-socio-environmental modeling of epidemiological process : a way for organizing humain environments and rythms to control and prevent the spread of contagious diseases. Human health and pathology. Université Grenoble Alpes, 2016. English. NNT : 2016GREAS036 . tel-01693272

**HAL Id: tel-01693272**

**<https://theses.hal.science/tel-01693272>**

Submitted on 26 Jan 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## THÈSE

Pour obtenir le grade de

### **DOCTEUR DE LA COMMUNAUTÉ UNIVERSITÉ GRENOBLE ALPES**

Spécialité : MBS - Modèles, méthodes et algorithmes en biologie,  
santé et environnement

Arrêté ministériel : 25 mai 2016

Présentée par

**Mohammad Hessam HESSAMI**

Thèse dirigée par **Nicolas GLADE**

préparée au sein du **Laboratoire Techniques de L'Ingénierie  
Médicale et de la Complexité - Informatique, Mathématiques et  
Applications.**

dans l'**École Doctorale Ingénierie pour la santé la Cognition et  
l'Environnement**

**Modélisation multi-échelle et hybride des  
maladies contagieuses - Vers le  
développement de nouveaux outils de  
simulation pour contrôler les épidémies**

**Multi-scale-socio-environmental modeling of  
epidemiological process. A way for  
organizing human environments and  
rhythms to control and prevent the spread of  
contagious diseases.**

Thèse soutenue publiquement le **23 juin 2016**,  
devant le jury composé de :

**Monsieur OLIVIER BASTIEN**

INGENIEUR CHERCHEUR, CEA GRENOBLE, Examineur **Monsieur  
BRIAN WILLIAMS**

CHERCHEUR, SACEMA - AFRIQUE DU SUD, Rapporteur

**Monsieur PASCAL BALLET**

MAITRE DE CONFERENCES, UNIVERSITE DE BRETAGNE  
OCCIDENTALE, Rapporteur

**Monsieur EMMANUEL DROUET**

PROFESSEUR, UNIVERSITE GRENOBLE ALPES, Président





I would like to dedicate this thesis to my loving parents.



## **Declaration**

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is a presentation of my own research work. Wherever contributions of others are involved, every effort is made to indicate this clearly, with due reference to the literature, and acknowledgement of collaborative research and discussions.

The work was done under the supervision of Research Professor Nicolas Glade, at the TIMC-IMAG laboratory , Grenoble.

Hessam Hessami

23 Juin 2016



## **Acknowledgements**

Firstly, I would like to express my sincere gratitude to my advisor Nicolas Glade for the continuous support of my Ph.D study and related research, for the interesting discussions, his patience, motivation, and immense knowledge in science and research. His guidance helped me in all the time of research and writing of this thesis.

Besides my advisor, I would like to thank the rest of my thesis committee: Prof. Brian Williams, Dr. Pascal Ballet, Dr. Emmanuel Drouet and Dr. Olivier Bastien for their insightful comments and encouragement, but also for the hard question which incited me to widen my research from various perspectives.

Last but not the least, I would like to thank my family for supporting me spiritually throughout writing this thesis and my life in general.



## Abstract

Theoretical studies in epidemiology mainly use differential equations, often under unrealistic assumptions ((*e.g.* spatially homogeneous populations), to study the development and spreading of contagious diseases. Such models are not, however, well adapted to understand epidemiological processes at different scales, nor are they efficient for correctly predicting epidemics. Yet, such models should be closely related to the social and spatial structure of populations. In the present thesis, we propose a series of new models in which different levels of spatiality (*e.g.* local structure of population, in particular group dynamics, spatial distribution of individuals in the environment, role of resistant people *etc*) are taken into account, to explain and predict how communicable diseases develop and spread at different scales, even at the scale of large populations. Furthermore, the manner in which our models are parametrised allow them to be connected together so as to describe the epidemiological process on a large scale (population of a big town, country *etc.*) and with accuracy in limited areas (office buildings, schools) at the same time.

We first succeeded in including the notion of groups in SIR (Susceptible, Infected, Recovered) differential equation systems by a rewriting of the SIR dynamics in the form of an enzymatic reaction in which group-complexes of different compositions in S, I and R

individuals form and where R people behave as non-competitive inhibitors. Then, global group dynamics simulated by stochastic algorithms in a homogeneous space, as well as emerging ones, obtained in multi-agent systems, are coupled to such *SIR* epidemic models. As our group-based models provide fine-grain information (i.e. microscopical resolution of time, space and population) we propose an analysis of criticality of epidemiological processes. We think that diseases in a given social and spatial environment present characteristic signatures and that such measurements could allow the identification of the factors that modify their dynamics.

We aim here to extract the essence of real epidemiological systems by using various methods based on different computer-oriented approaches. As our models can take into account individual behaviours and group dynamics, they are able to use big-data information yielded from smart-phone technologies and social networks. As a long term objective derived from the present work, one can expect good predictions in the development of epidemics, as well as tools to reduce epidemics by guiding new environmental architectures and by changing human health-related behaviours.

# Contents

<b>List of Figures</b>	<b>xv</b>
<b>List of Tables</b>	<b>xix</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Epidemiology and historical perspective . . . . .	1
1.1.1 Early epidemiology . . . . .	1
1.1.2 Epidemiology in 20 <sup>th</sup> century . . . . .	3
1.2 Epidemiological models, their restrictions and what we aim to do . . . . .	6
<b>2 Models and Computer Simulations in Epidemiology</b>	<b>15</b>
2.1 Background: Classical Modelling in Epidemiology . . . . .	16
2.1.1 Overview of Epidemiology & Mathematics . . . . .	16
2.1.1.1 Deterministic models . . . . .	16
2.1.1.2 Statistical models . . . . .	19
2.1.1.3 Stochastic models . . . . .	19
2.1.2 Compartmental models in epidemiology . . . . .	20

2.2	New Mathematical Models: A Modified S-I-R Model & A Group Dynamics	
	Model . . . . .	24
2.2.1	SIR epidemic models & enzyme kinetics . . . . .	24
2.2.1.1	A modified version of the classical SIR model: SIR kinetics without resistance (and with Delay) . . . . .	26
2.2.1.2	A modified version of the classical SIR model : SIR kinetics with resistance . . . . .	31
2.2.2	Group Dynamics in a Population . . . . .	34
2.2.2.1	Modelling the group-size in homogeneous SSA-based models . . . . .	40
2.2.2.2	Modelling the group-size in MAS-based Models . . . . .	46
2.2.2.3	Group-Lifetime . . . . .	50
2.3	Computer Implementation of The SIR Models . . . . .	57
2.3.1	Different Models for Representing Local and Global Spatiality . . . . .	59
2.3.2	Modified SIR in ODE (Homogeneous Space) and in PDE (Locally Homogeneous) . . . . .	65
2.3.3	The SSA Algorithm - Application to SIR and Group Dynamics . . . . .	68
2.3.4	Implementation of SIR Dynamics in SSA-Group Based Models . . . . .	71
2.3.5	Implementation of SIR Dynamics in Stochastic Cellular Automata (SSA-CA) Based Models . . . . .	79
2.3.6	Multi-Agent System Modelling of Group Structure and SIR Dynamics (MAS-G) . . . . .	82

---

2.3.6.0.1	Group dynamics in the MAS-G model. . . . .	82
2.3.6.0.2	SIR dynamics in the MAS-G model - dynamics.	86
2.3.6.0.3	SIR dynamics in the MAS-G model - algorithm.	90
2.4	Measurements . . . . .	93
2.4.1	Global Measurements . . . . .	93
2.4.2	Epidemic outbreaks and criticality analysis . . . . .	95
2.4.3	Micro and Macro Events – Fine to Coarse-Grain Observations . . .	100
2.4.4	Infection Charge as Signatures of Diseases . . . . .	111
<b>3</b>	<b>Results of our Epidemiological Simulations</b>	<b>115</b>
3.1	SIR Epidemic process in ODE-based Model: Continuous models . . . . .	115
3.1.1	Modified SIR dynamics without resistance . . . . .	115
3.1.2	Modified SIR dynamics with resistance . . . . .	119
3.2	SIR Epidemic process in SSA & MAS-based Models-Discrete models . . .	125
3.2.1	SSA-G Model . . . . .	126
3.2.2	SSA-CA Model . . . . .	129
3.2.3	MAS-G Models . . . . .	133
<b>4</b>	<b>Discussion and Perspective</b>	<b>143</b>
4.1	General discussion . . . . .	143
4.2	Complementary Approaches . . . . .	147
4.2.1	Participatory Simulations and video games . . . . .	148
4.2.2	Smart Device Applications, Social Network and Big-Data . . . . .	149

---

4.3 Possible Future Works . . . . .	153
<b>Appendix References</b>	<b>161</b>
<b>Appendix A <i>SI</i> and <i>SIR</i> Dynamics</b>	<b>177</b>
A.1 Modified <i>SI</i> kinetics without resistance (and with latency) . . . . .	178
A.2 modified <i>SIR</i> kinetics with resistance . . . . .	180
A.3 About Mass Conservation Equations of $[S(t)]_{tot}$ . . . . .	184
<b>Appendix B Effective Surface</b>	<b>187</b>

# List of Figures

1.1	Herd Immunity . . . . .	14
2.1	Bernoulli's Compartmental Model . . . . .	18
2.2	A classical epidemic model . . . . .	20
2.3	A classical endemic model . . . . .	21
2.4	Graphical diagram of a SIRS epidemic model . . . . .	22
2.5	Impacts of Free-population Density and Movement Speed on Mean-free-time and Group-size . . . . .	44
2.6	Group Size Distribution . . . . .	45
2.7	Group Formation of Individuals . . . . .	47
2.8	Box Plot of Group-sizes During Simulated 8-hour Work Day . . . . .	48
2.9	Histogram of Group-size in MAS Model . . . . .	49
2.10	The Goodness of Fit for MAS Data . . . . .	50
2.11	The Gamma Probability Density Function for Different Group-size . . . . .	52
2.12	Group Dynamics in a Simulated Big City (Slow Movement Speed) . . . . .	55
2.13	Group Dynamics in a Simulated Big City (Fast Movement Speed) . . . . .	56

2.14	Locally or Globally Spatialised Numerical Models . . . . .	65
2.15	SIR infection process in a group-based SSA model . . . . .	73
2.16	Problems relating to interactions between bimolecular and monomolecular reactions-events in the SSA-G model in continuous time . . . . .	75
2.17	Stochastic Simulation Algorithm based on Cellular Automaton (SSA-CA Model) . . . . .	82
2.18	Individual and Group classes of MAS-G Model . . . . .	84
2.19	Multi Agent Simulation in Group-based System . . . . .	86
2.20	Group-Forming and SIRS-epidemic Simulation Diagram . . . . .	92
2.21	Critical Dynamics . . . . .	98
2.22	Micro and Macro Events at the Individual Level . . . . .	102
2.23	SIRS Micro and Macro Events . . . . .	103
2.24	Stochastic SIR model: micro and macro events . . . . .	104
2.25	Micro and Macro Events at Population Level . . . . .	111
2.26	Probability Distribution of Infection Charge . . . . .	113
3.1	Dissociation constant and Population Density Impacts on SIR Epidemic Dynamics . . . . .	118
3.2	Effects of $R$ -individuals on SIR Epidemic Dynamics . . . . .	122
3.3	Population Density and Stochastic Simulation Algorithm of Group-based Model (SSA-G) . . . . .	128
3.4	Impact of resistance and moving rates on SIR Dynamics . . . . .	132
3.5	SIR Kinetics Propagation Fronts . . . . .	133

---

3.6	Group Formation and Population Density in the MAS-G Model . . . . .	137
3.7	The Group Formation and Number of Outbreaks in MAS-G Model . . . . .	138
3.8	SIR Epidemic model with & without Group Formation . . . . .	140
3.9	Daily Incidences at Group and Population Levels . . . . .	141
4.1	Discussion . . . . .	145
4.2	FluPhone Project . . . . .	149
4.3	Social Network Data . . . . .	150
4.4	FaceBook Data . . . . .	152
4.5	Multi-Modal Data Approaches . . . . .	155
4.6	Hybrid Models to Organise the Human Environment . . . . .	158
4.7	Disease Surveillance System . . . . .	158
4.8	Personal Diseases Surveillance Assistant Applications such as Virtual Com- panion . . . . .	160
B.1	Occupied Spaces . . . . .	188
B.2	Grenoble Map . . . . .	189



# List of Tables

2.1	Parameters used in our different numerical simulations . . . . .	63
2.2	SSA-G numerical model implementation . . . . .	78
2.3	Gillespie algorithm implementation for a given group $k$ . . . . .	91



# Chapter 1

## Introduction

### 1.1 Epidemiology and historical perspective

Epidemiology is the study of the distribution and determinants of health-related states or events, and the application of this study to the control of diseases and other health problems. Various methods can be used to carry out epidemiological investigations, such as surveillance and descriptive studies, which lead to an understanding of the critical factors. This corresponds to the current definition given by World Health Organisation (WHO). By taking a glance back in time we will see how epidemiology has developed from "early" to "contemporary" concepts and arrives at this point [80].

#### 1.1.1 Early epidemiology

The history of epidemiology started with Hippocrates (C.470-C.400 BC) who performed simple medical observations to help explain the spreading of diseases that decimated popu-

lations. The early epidemiology period lasted until the first third of the nineteenth century, during which the germ theory of disease (e.g. the first description for typhus and syphilis) was addressed by Gerolamo Fracastoro (1478-1553), the clinical observation method was developed by Thomas Sydenham (1624-1689) and the first epidemiological study of a non-communicable condition (*i.e.* disease that is not transmitted; including the investigation of a series of sudden deaths in Rome) was elaborated by Giovanni Maria Lancisi (1654-1720) [87].

In the beginning of the 18<sup>th</sup> century, after the probability theory arose, the first changes in epidemiological studies, appeared. Daniel Bernoulli (1700-1782), Mathematician-Physicist, was the originator of the first epidemiological model for an infectious disease. He was probably the first to systematically use differential equations for deducing a number of formulae [97] (refer to section 2.1.1).

Almost 100 years later, after Bernoulli's model, epidemiological modelling was further developed by using mathematical and statistical tools. Sir William Farr (1807-1883) was one of the first and few who successfully attempted modelling veterinary problems by using the equation of second and third ratios. His studies consisted of the attempts to predict the outcome of a rinderpest epidemic in England (1865). The considerable work of John Snow (1813-1858) [87] in calculating the valid rate of cholera outbreaks<sup>1</sup>, was one of the first observational statistics studies in epidemiology.

---

<sup>1</sup>Broad Street cholera outbreak in Soho district of London in 1854

### 1.1.2 Epidemiology in 20<sup>th</sup> century

In the early 20<sup>th</sup> century, several mathematical models were developed, e.g. W. Hamer's discrete model to analyse and formulate the measles epidemics (1906) [45], or the ordinary differential equation (ODE) -based model of Ronald Ross (1857-1932) [87] formulated to study and control malaria (1911), which was also the first model that determined the threshold density of man-to-mosquitoes below which malaria would disappear from the human population [87, 88]. Later, this model was developed by Kermack and McKendrick (1927), which today is known as the Susceptible-Infective-Recovered (SIR) model (some SIR-based models are explained in section 2.1.2 ). This model consists of a system of three coupled nonlinear ordinary differential equations [61]. Historically, the first epidemiological studies were based on ODE modelling of the variations of continuous amounts of persons of different health states, usually classified as susceptible people  $S$  that can be infected in the presence of contagious germs, infected people  $I$  (which can also be separated into only infected persons and infectious ones), and potentially recovered people  $R$  that are immuned for a certain time until they loose their immunity and become susceptible once again. Such challenges include the development of numerical models to understand epidemics from the smaller to the larger scales.

In general, the twentieth century was a period of a rapid growth for the application of mathematical and computational tools in the field of epidemiology on public health policies. We might exclude the period in which science was affected by world wars and even the several years after the second world war when there was no effective dialog between applied

epidemiology and mathematical modelling [90]. In the late 1970s and early 1980s, epidemiologists that focused more on developing the field of statistics and dynamic mathematical models have been reinforced by computer simulation [52].

Since the last decades of 20<sup>th</sup> century, human life patterns have changed : an increase in international travels, a restructuration of work and an increase in the urban population to the detriment of the rural population, changes in economic patterns and in the social norms. Moreover, the world population has increased exponentially from 2.5 billion in 1945 to over 7 billion in 2015. All this provides new challenges in the analysis and control of the spatio-temporal spread of contagious diseases including emerging ones (AIDS, SARS, ebola, chikungunya, *etc.*) or re-emerging diseases (resistant tuberculosis, plague, malaria,*etc.*) [66, 92].

The rapid growth in mathematical and informatics applications in epidemiology has led to a new methodological revolution. Newly developed compartmental models (such as SEIS; SEI; SIRS; SEIRS; SEIR; MSEIR; MSEIRS ...) are used for this purpose and most of those recently developed have involved aspects such as passive immunity, stages of infection, gradual loss of vaccine and disease-acquired immunity, vertical transmission, age structure, social and sexual mixing groups, spatial spread, vaccination, quarantine, *etc.* [3, 6, 71, 74, 31].

Computer sciences coupled to epidemiological methods should now be well adapted to simulate epidemics in our strongly industrialised and computerised countries given the access to more and more important individual data. Such data, from *omics* – including the *exposome*

that informs about the vulnerability of individuals to contagious diseases – to pre-topological data that inform about the manner people live (or work) and on their preferential contacts and pathways, are now frequently (if not easily) available since they can be extracted from *The BigData*. Such data provides information about people's life styles, their contacts, their frequented places and their trips. This information is now easy to access by data-mining approaches from smartphones, social networks, online newsletters and online consumer preferences (including travel, food *etc.*). This can lead to an understanding of individual behavioural aspects in a given population facing an epidemic process, and should be taken into account in multi scale numerical models. However, only few recent epidemiological studies focus on these approaches to bring new alternatives to mathematical based modelling, health monitoring, control and prevention (see Broniatowski et al [14], Li et al [63]).

We believe "tomorrow's" challenges will depend on *hybrid* (models that couple different modelling formalisms) and multi-scale (models that give access to an understanding of the process at different scales) models including multi-modal data. This is the purpose of the present work, which will be described in more detail in the next chapters, as well as perspectives and future works (section 4.3).

## 1.2 Epidemiological models, their restrictions and what we aim to do

Studying historical perspectives in epidemiology shows different motivations for solving epidemic problems. Most of them have economic and financial reasons. For example, the main objective of D. Bernoulli was to calculate the gain in life expectancy at birth if smallpox were to be eliminated as a cause of death, because at the time annuities were being sold, his work on the prolongation of life expectancy at any age had immediate financial impact. Another example is Sir William Farr's attempt to reduce the economical impacts of rinderpest epidemics on human life in England, or William Hamer's and Ronald Ross's works in the prevention of malaria, each of which was somehow initiated by economic motivations [22, 24, 94].

These examples show how early epidemiology was trying to answer epidemic problems by providing concrete solutions according to actual human needs in real time. However it is important to understand that the coarse grain models they used in those centuries were well adapted to their economical and epidemiological problems. This is not necessarily the case nowadays. Their contemporary counterparts are more mathematically developed but less focused on solving explicitly epidemic problematics vis-a-vis individuals' need. Even the most advanced current mathematical models are more based on theoretical aspects than on the aim to provide direct resolutions of epidemics at the individual level. Even at the population level, they have some important restrictions. For example the deterministic ODE models,

which assume that a population (host-agent) is spatially uniform and homogeneous (as an unrealistic assumption), or locally homogeneous and globally heterogeneous in the case of partial differential equations (PDE), often provide inappropriate results. As an illustration of this point, when modelling the transmission of malaria, lack of consideration for the heterogeneity of contact (in human-mosquito populations) leads to an overestimation of the infection prevalence [68, 90, 52]. The stochastic models derived from their deterministic formulations of stochastic differential equations (SDE) are more complicated and difficult to formulate in comparison with differential equation based (deterministic) counterparts (see Allen [1]).

Generally speaking, the occurrence of disease is a result of interactions between components of the agent-host-environment (including the biological characteristics), and the spread of disease may be influenced by social-cultural and geographical mechanisms. While studying an epidemic process, despite the aforementioned methods (*i.e.* deterministic, stochastic), it is important to take into account both the elements that cause the epidemic disease occurrence and the mechanisms that influence the spreading of disease.

Therefore, considering all possible interactions and mechanisms makes the epidemiological studies more rigorous, but an epidemic process involving all interfering factors (e.g. biological, environmental, geographic, demographic and socio-cultural) quickly becomes complex. Such a system wouldn't be formulated easily by some differential equations and in the case that it does, the differential equations will be very difficult or impossible to solve

explicitly. This appears especially when the complex behaviour is admitted in a system. Computer simulations are an alternative method for relaxation of the unrealistic assumptions and also in dealing with rigorous differential equation based models. Through dealing with such complex systems, the simulation methods based on spatio-temporal stochastic simulation algorithms (SSA) or intrinsic spatio-temporal multi-agent systems (MAS), provide more information : there are more explicative models in comparison with differential equation based models. As two explicit examples, one can refer to [21] in which a stochastic discrete-event simulation was used for modelling the spread of airborne infections (by considering the mixing patterns of social contact network) or see [74] in which the theoretical results for an SEIRS deterministic model has been analysed by computer simulation.

These point out the importance of providing new methods that characterise the epidemic process more conceivably at different levels (individual/population) and also models that lead to an understanding of the effects of social and local spatial structures on disease spread dynamics. In light of this, this thesis, regarding to such a challenge, proposes a hybrid modelling approach implemented in series of numerical models. In this way we are able to study the effect of socio-environmental factors and spatio-temporal population dynamics, including their local structure, on the S-I-R (Susceptible, Infected, Recovered) epidemic process using 3 different simulation models :

- An ODE-based model including the notion of group formation
- A hybrid discrete group based models (using a simulated continuous time SSA and discrete time SSA) that compute epidemics within groups

- A MAS-based model describing group formation with discrete time SSA computed epidemics within groups

One of the interests of this thesis is to study how group-formation as a basic social structure can affect the epidemic dynamics in the context of communicable disease transmission. The spatial and population density factors as well as the number of "resistant" individuals (e.g. vaccinated population as natural barriers against diseases spread) in populations will be studied using the above models, notably depending on the group structure of populations.

What does "groups" mean in epidemiological context? Before going into more detail about the definition of a "group" and its meaningful sense in epidemiology, we should describe what kind of diseases introduce the social groups as an important key to unlock the propagation path. Some communicable diseases such as hepatitis A, gastrointestinal illnesses, skin infections, cold and flu, which are caused by germs, can spread easily via hands; directly (by shaking hands) or indirectly (by objects containing germs), droplets in the air, sharing clothes and towels or infected foods and water. Considering this fact, in each society there is always a risk of encountering at least one of the above situations. In our social life we meet other people, referred to as social groups, such as our family, friends, colleagues or even strangers in many different conditions and places in which we are potentially at risk. This means the groups can have a significant role in the spread of communicable diseases.

Group and individual dynamics should be the heart of epidemiological models as they describe the fundamental social structure of populations in which we aim to study the spread

of diseases. However, these dynamics are not easy to represent in differential equation based models (in which it is difficult to implement discrete features such as groups), nor in MAS models (in which the formation of groups by emerging processes is also hard to obtain). In the first case this requires the introduction of multi-scale population dynamics. Dynamics of an epidemic process can vary differently depending on *group-size* and *group-lifetime*. For example, whether a group of 6 people stay together for 30 seconds in an elevator or 10 people form a group in a big meeting room for 2 hours, affects the epidemic dynamics. In the case of individuals-based models (which are particularly complicated to control) we can ensure that the agents behave realistically enough to observe emerging social behaviour (i.e. the particular group dynamics).

Spatially speaking, the groups form dense space situations (e.g. 10 people in a lift of  $4\text{ m}^2$  represent a density of  $2.5 \times 10^6$  people per  $\text{km}^2$ ; so they can be compared with the population densities as they are usually considered (e.g. the average density of Paris is between 10,000 to 20,000 people per  $\text{km}^2$ ), i.e. considerably lower than the above lift example. The average distances for sufficient contacts can also vary extremely, for example from a few millimetres in very intimate relations (e.g. sexual contacts) or centimetres (the lift example) to a few meters (e.g. people with flu can potentially spread the germs to others over a distance of up to 2 meters). The group concept is truly unclear, for example could we consider the people in a  $30\text{m}^2$  dining room as a group, or two individuals working together in a office of  $6\text{m}^2$ ? On the other hand, groups can last for short or long times, e.g. from 2 seconds to a few minutes or several hours. These kind of questions will be tackled in chapter 3 as *group-size*

and *group-lifetime* by studying their probability distributions as well as their impact when modelling epidemic processes.

Population density matters to epidemic outbreaks. Diseases in dense populated societies are spreading faster than in populations with low densities. For example, an epidemic of severe acute respiratory syndrome (SARS) started in Asia on November 2002 and lasted until July 2003, with a majority of cases in Hong-Kong<sup>2</sup> the 4<sup>th</sup> populated country in the world, or the recent ebola outbreaks (West Africa, 2014) occurred more in urban than rural areas, because of population density. We studied in sections 3.1, 3.2.1 and 3.2.3 how population density can influence S-I-R epidemic dynamics. The moving rates of individuals and group-tendency as latent variables are also integrated in those modelling methods depending on related circumstances.

The  $R$  (recovered) class, in S-I-R compartmental models, as the "resistant" (immune or vaccinated) individuals, can also influence epidemic dynamics. The presence of resistant individuals in a given population leads to a sort of barrier against the spread of disease as fronts against propagation. The  $R$ -individuals, in communicable disease modelling, are usually (e.g. in classical SIR models) considered as the transition class for  $I$ -individuals after recovery or before the reinfection process (see [43, 44, 96, 105]) or as flow of individuals from  $S$  population class in the context of vaccination strategies in population (see [51, 82, 81, 95, 113]). Since the role of  $R$ -individuals is summarised as a simple transition class in classical SIR

---

<sup>2</sup>9.6% fatality rate according to the World Health Organisation (WHO)

modelling approaches, there is no chance of taking into account the *community* or *herd* immunity feature of "resistant" individuals. They form an indirect protection (a front of  $R$ ) for  $S$  population class from infectious disease. The herd immunity feature is illustrated in figure 1.1, in the population with a majority of  $R$ -individuals, the chains of infection are likely to be disrupted, which leads to a stop or slows the spread of disease Merrill [75].

The role of immune individuals in populations seems more important than it is generally thought and this role is not yet integrated into S-I-R epidemic modelling, particularly for those whose population is considered homogenous, and so we cannot study the local structure's impact on disease spread. The spatial models called correlation (Dietz & Hadelar 1988; Altmann 1995; Keeling et al. 1999) apply this feature by considering the interactions between individuals (or sites) as something occurring on a network (i.e. the case of communicable disease). These interactions can be studied as a contact structure that forms a network of links (pretopology) between individuals, which provides a more general framework and neighbourhood structure than those considered in traditional spatial models Keeling [60]. For example, the pairs of type  $I-R$  ( $[IR]$ ) and triple of type  $S-I-R$  ( $[SIR]$ ) represent respectively a couple of  $I-R$  and group of  $S,I,R$ -individuals in a given node in the network in which we can explain absence of  $I$ -individuals because of being in a group with  $R$  individuals. In this thesis we go further, and by applying the chemical kinetics theory, we modify the **classical S-I-R model** by integrating notions from enzymatic reactions (inhibition and competition). In this way, in our **modified S-I-R model**, we are able to take into account the local spatial structure of the epidemic process, even in homogenous populations, for ODE and SDE-based

models (see in section 2.2.1). It should be noted that modified S-I-R kinetics won't be applied in individuals based models (such as stochastic cellular automata (CA) or MAS models). In the first case (CA) this is because the notion of a group is intrinsically encoded in the neighbourhood's relationships between individuals. In the second case (MAS) this is because epidemic process within the groups are considered homogeneous.

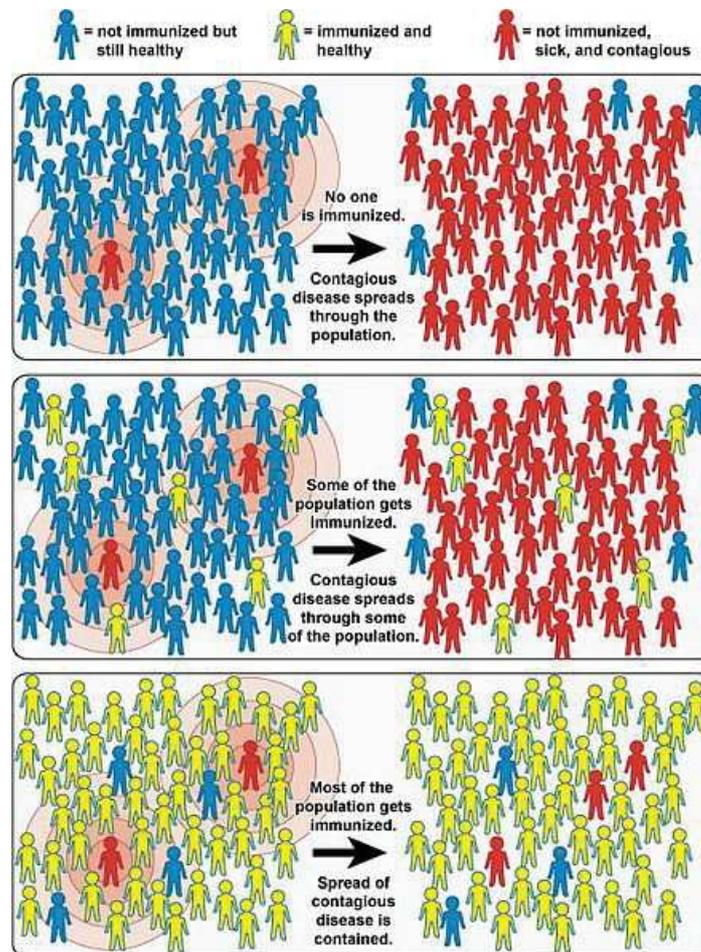


Figure 1.1 **Herd Immunity** : The top box shows an outbreak in a community in which a few people are infected (shown in red) and the rest are healthy but unimmunised (shown in blue); the illness spreads freely through the population. The middle box shows a population where a small number have been immunised (shown in yellow); those not immunised become infected while those immunised do not. In the bottom box, a large proportion of the population have been immunised; this prevents the illness from spreading significantly, including to unimmunised (or susceptible) people. (Source : National Institutes of Health NIH)

## Chapter 2

# Models and Computer Simulations in Epidemiology

This chapter is dedicated to the description of the mathematical models used in epidemiology and their numerical implementation. The first part (2.1) presents an overview of this, focusing particularly on *SIR* compartmental models and the related *ODE* mathematical models. In the second part we describe (2.2) our own form of a modified *SIR* mathematical model. In the same part, we also explain how we model group dynamics in a population independently from any epidemiological context. Once expressed, we show in part 2.3 how we implement numerically (and by using different numerical schemes like cellular automata, multi-agent systems, *etc.*) these mathematical models to perform simulations in a computer. Finally, in part 2.4 we explain precisely how to get the numerous different measurements from our models, *i.e.* both measurements related to population dynamics (group sizes, lifetimes, spatio-temporal distribution of individuals *etc.*) and indicators of the epidemic dynamics.

## 2.1 Background: Classical Modelling in Epidemiology

In this part we describe epidemiology viewed by mathematics and computer sciences. A brief description of different models as well as some classical compartmental SIR models, will be presented.

### 2.1.1 Overview of Epidemiology & Mathematics

Mathematical based models can provide a clear view of a given disease by simplifying the epidemic process elements (*e.g.* infected people, transmission, lack of immunity etc.) like variables, parameters and relevant factors. In literature, one can find a considerable list of advantages in using mathematical models to improve epidemiological understanding (*e.g.* see notably [46]). One can also refer to [3] who provided a brief summary of the mathematical models' utility in the study of HIV transmission and the epidemiology of AIDS.

Depending on the nature of variables used, we can consider three global modelling approaches based on deterministic, statistical and stochastic models.

#### 2.1.1.1 Deterministic models

The deterministic models are the simplest models for studying epidemic dynamics. They are based on homogenous populations (*e.i.* they assume there is no difference between individuals in the population). Theoretical models in epidemiology consist of ordinary differential equations (ODE) to study the diseases spreading. The mathematical analysis of these *ODE* leads to analytical solutions, evaluation of equilibrium points, determination of stability and

sensitivity analysis.

Daniel Bernoulli is frequently quoted as originator of the first epidemiological models for an infectious disease. His work was one of the first epidemiological studies to assess quantitative conjectures by using such a deterministic model [12, 16, 22, 24]. The main objective of Bernoulli was to calculate the gain in life expectancy at birth if smallpox were to be eliminated as a cause of death. His work on the prolongation of life expectancy at any age had immediate financial impacts because at this time it allowed to evaluate correctly the annuities (the economic considerations in relation to population dynamics often motivated the original epidemiological studies).

Bernoulli's model is one of the first compartmental models (Figure 2.1). The aim of this model was to describe the conversion of susceptible  $S$  people into immune individuals  $R$  in the case of an immunising infection. Susceptible individuals are defined as those who have not yet been infected, and immune individuals are those who have been immunised for the rest of their life (or at least for a long time) after one infection. Notice that Bernoulli didn't aim to evaluate the number of infected people  $I$  and he only wanted to calculate the number of people who would be healthy the next year and that would be able to pay annuities. The death rate (denoted by  $\mu(a)$ ) refers to death due to all causes except the infection. The force of infection  $\lambda(a)$  is the rate according to which susceptible individuals  $S$  are infected and then immunised ( $R$ ). Only a fraction  $s(a)$  survives to become immune and the rest  $c(a) = 1 - s(a)$  die due to the infection. Let  $u(a)$  and  $w(a)$  be the probability for newborn individuals to be alive and susceptible, and the probability to be immune and alive at age  $a$  respectively,

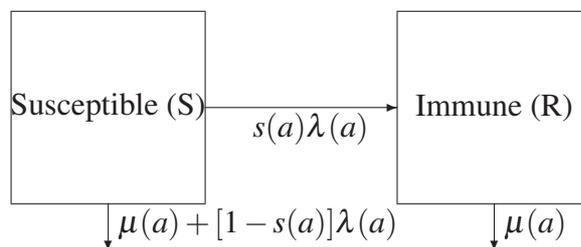


Figure 2.1 **Bernoulli's Compartmental Model** : The number of susceptible people either decreases through infection, by transition to the immune class or by decreasing through fatality due to the infection (or other diseases); the number of immune people in this class increases through infection of susceptible individuals who have been immunised or decreases by fatality.

then the differential equations corresponding to these compartments are described by the following system, where  $u(0) = 1$  and  $w(0) = 0$  are the initial conditions :

$$\begin{cases} \frac{du}{da} = -[\lambda(a) + \mu(a)]u \\ \frac{dw}{da} = s(a)\lambda(a)u(a) - \mu(a)w \end{cases}$$

The analytical solutions of the above equations are  $u(a) = e^{-[\Lambda(a)+M(a)]}$  and  $w(a) = e^{-M(a)} \int_0^a [1 - c(\tau)]\lambda(\tau)e^{-\Lambda(\tau)}d\tau$ , where  $\Lambda(a) = \int_0^a \lambda(\tau)d\tau$  and  $M(a) = \int_0^a \mu(\tau)d\tau$ . The probability of surviving to age  $a$  is defined by  $l(a) = u(a) + w(a)$ . Then, the survival function in the population in the absence of smallpox is determined by  $l_0(a) = e^{-M(a)}$ , and the survival function in the presence of smallpox is  $l(a) = l_0(a)[e^{-\Lambda(a)} + \int_0^a [1 - c(\tau)]\lambda(\tau)e^{-\Lambda(\tau)}d\tau]$ .

As mentioned before, Bernoulli's main purpose was to calculate the gain in life expectancy if smallpox were eliminated as a cause of death, then the life expectancies at birth with and with-

out smallpox are calculated by integrating the survival curves over all ages.  $L = \int_0^\infty l(a)da$  denotes the life expectancy at birth in the presence of smallpox and  $L_0 = \int_0^\infty l_0(a)da$  denotes the life expectancy at the birth in the absence of smallpox (see [22] for more details).

### 2.1.1.2 Statistical models

In recent decades, the mathematical studies in epidemiology have depended more and more on the developing field of statistics. Statistical models are data-based models that completely depend on observed data. However, statistical models, as well as deterministic models, ignore some information and mechanisms of the system such as the spatial-temporal dynamics of populations, but they are widely used to estimate threshold  $R_0$  (the basic reproduction number) and to assess the effectiveness of control strategies, see [3, 4, 17, 23, 25, 31, 66, 106, 109]. Since our approach aims to be explicative, it is not concerned by such data-based methods.

### 2.1.1.3 Stochastic models

Stochastic models compared to deterministic ones assume the stochastic variation of variables in epidemic systems. As all diseases are subject to stochasticity in terms of a natural chance of transmission (the probability of transmission is not the same for everyone), then the stochastic models can be more appropriate (and realistic) for modelling infectious diseases [57, 92, 103]. This stochastic modelling framework is particularly well adapted for modelling disease extinctions or the spread of diseases in small communities [18, 78]. The mobility of the population, populations with different mixing patterns, demographic effects ... are

all example of factors that can be considered in stochastic models, *e.g.* in the modelling of *AIDS* and *HIV* epidemics [93, 106–108].<sup>1</sup>

### 2.1.2 Compartmental models in epidemiology

Models called *compartmental* are formed of two parts: classes and reactions. The population compartmental classes are the possible (diagnostic) states of the disease (*e.g.* infected class  $I$ ) and reactions specify the proportion of individuals moving from one state to another (*e.g.* the proportion  $\gamma I$  when  $I \xrightarrow{\gamma I} R$ ). For example a *SIRS* compartmental model consists of three different classes: Susceptible( $S$ ), Infected ( $I$ ) and Recovered ( $R$ ) population, with a transition between  $S$  to  $I$ ,  $I$  to  $R$  and then a return from  $R$  to  $S$ . As shown below in the general transfer diagram (Figure 2.2 and Figure 2.3), in a classical epidemic SIR model, the labels  $S, I$  and  $R$  represent respectively susceptible, infectious and recovered people.

The epidemic models are used to describe the rapid outbreaks that occur in less than one year. In these models the vital dynamics (birth and deaths, represented by the  $\mu$  parameter) of the population are generally not taken into account except in endemic models. The horizontal

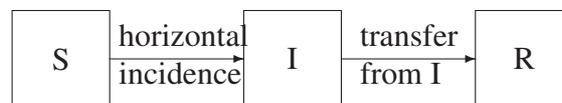


Figure 2.2 A classical epidemic model

incidence is the infection rate of susceptible individuals through their contacts with infectious individuals [47]. This means that if  $\beta$  denotes the average number of adequate contacts of a

<sup>1</sup>Some mixed stochastic-statistical models called state space, are as well known as modelling approaches in epidemiology

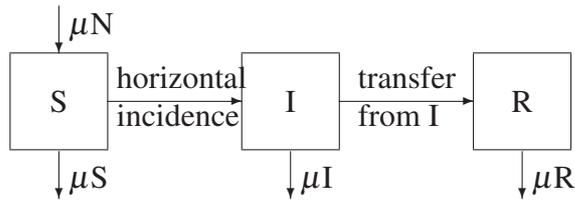


Figure 2.3 A classical endemic model

person per unit time,  $\beta I/N$  is the average number of contacts with infectious individuals per unit time of one susceptible individual (where  $N$  is the total population size). Then  $(\beta I/N)S$  gives the number of new cases per unit time when the entire population is considered to be susceptible. This form of the horizontal incidence is called *standard incidence*. Note that the notation of  $\beta$  doesn't always denote the average number of adequate contacts in all the epidemiological literature. Sometimes  $\beta$  is called *the contagion rate* and determined by  $\beta=cq/N$  where  $q$  denotes the per interaction probability of infection transmission from one  $I$  person to one  $S$  person by their contacts and  $c$  is the number of different interactions per person in the population [106]. Therefore the horizontal incidence might be defined differently in some references. Transfer from the  $I$  compartment and the movement into the  $R$  compartment is governed by the term  $\gamma I$ . Supposing that the duration of an infection is exponentially distributed with the mean waiting time  $1/\gamma$  (it is also called *the average infectious period*), then the rate at which  $I$ -individuals recover is  $\gamma$  [69].

Figure 2.4 shows the graphical diagram of a conventional *SIRS* epidemic model in a population of size  $N$ , where  $\beta$  is the transmission parameter and  $\beta I/N$  expresses the average number of contacts of one susceptible individual in the presence of infectious people in the population per unit time [47]. The recovery and the loss of immunity rates are defined respectively by  $\gamma I$  and  $\lambda R$ .

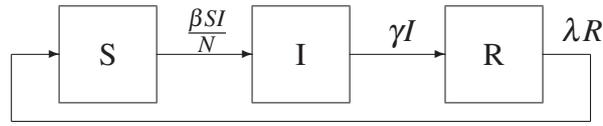


Figure 2.4 Graphical diagram of a SIRS epidemic model

The above *SIRS* model can be formulated by a set of ordinary differential equations as follows:

$$\frac{dS}{dt} = -\beta IS/N + \lambda R, \quad \frac{dI}{dt} = \beta IS/N - \gamma I, \quad \frac{dR}{dt} = \gamma I - \lambda R \quad (2.1)$$

The analytical solution of these equations allows the computation of the state vector  $X(t) = [S(t), I(t), R(t)]$  at time  $t$ . The change in the number of individuals in a given class is calculated by its time derivative (*e.g.*  $dS/dt$  is the variation in the number of individuals  $S$  at time  $t$ ). A negative balance means a loss of  $S$  individuals, while a positive balance means adding new individuals in the  $S$  population class. The balance  $dI/dt$  is the infection rate because it represents the number of new infected people suffering from the disease.

The well-known compartmental *SIR* models (Susceptible-Infected-Recovered) of [61] is always a source of inspiration in epidemiological studies. From that historical model, several models have been developed and improved [23, 48, 54, 109, 106, 53]. Recently developed models are considering different aspects of epidemic processes such as heterogeneity of the populations, spatial effects, different clinical stages and social and cultural aspects [74, 104, 71, 3, 31, 6].

The *basic reproduction number* ( $R_0$ ), the *contact number* ( $\delta$ ) and the *replacement number* ( $R$ ) are important thresholds used in the recent epidemiological modelling literature. Although they are all equal at the beginning of infection spreading, this changes to  $R_0 \geq \delta \geq R$  after the infection has invaded a population [47]. The basic reproduction number is defined as the average number of secondary infections generated by one typical infectious individual taken in the susceptible population [3].  $R_0$  is often used to determine whether a disease can invade a population. A  $R_0$  less than 1 means that the disease disappears from the population, when greater than or equal to 1, the disease will spread in the population. In a classical epidemic model (as shown schematically in Figure 2.2), the basic reproduction number is the average number of adequate contacts multiplied by the average infectious period, then  $R_0 = \beta \times dI/\gamma$ . The contact number  $\delta$  is the average number of adequate contacts of a typical infectious subject during the infection period. As mentioned before, when an infection starts to spread, this quantity is equal to the basic reproduction number. Note that after the spreading has started, this equality ( $R_0 = \delta$ ) will remain stable for most models and the replacement number  $R$ , which is also called *the reproduction number*, is defined as the average number of secondary infections provided by a typical infectious subject during the infection period.

## 2.2 New Mathematical Models: A Modified S-I-R Model & A Group Dynamics Model

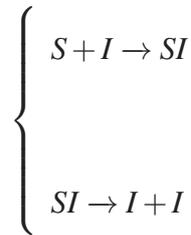
Here is described a part of our work, *i.e.* the conception of new mathematical models on which are based our numerical epidemiological models. In particular we modified the classical *SIR* epidemic dynamics so as to take correctly into account the influence of immune subjects *R* and to introduce the notion of groups. Moreover, we describe a manner of simulating a population structure, *i.e.* with a model of group formation and their dynamics.

### 2.2.1 SIR epidemic models & enzyme kinetics

The classical *SIR* models can be represented by a simple succession of enzymatic reactions expressed in the chemical kinetics theory. In this case, the transitions of individuals between the compartmental classes, are considered as chemical molecular reactions (precisely enzymatic reactions). Then the classical *SIR* models can be represented by a set of unimolecular (*e.g.*  $I \rightarrow R$ ) or bimolecular (*e.g.*  $S + I \rightarrow I + I$ ) reactions that can be expressed by a set of differential equations which are known as *enzymatic systems in the chemical kinetics* [2, 77]. Enzymes, by reacting with a substance during a catalytic reaction, provide intermediate products which are known as *enzyme-substance-complexes* (refer to theory of S. Arrhenius). In the *SIR* epidemic context the substance and enzyme are represented by individuals in the susceptible *S* and infected *I* classes respectively: infected people cause the transition from *S* subjects to *I* subjects; from the other point of view *S* people suffer from the effect of infection and behave as passive substrates. However there is a fundamental difference between a *SIR*

system and an enzymatic system: while enzymes and substrates are of a different nature, this is not the case between  $S$ ,  $I$  and  $R$  that are all people.

Then, the infection reaction in a  $SIR$  model ( $S + I \rightarrow I + I$ ) is written like its analogue in an enzymatic systems as follows:



In our model the intermediate product  $SI$  will be called *group-complex* of individuals  $S$  and  $I$  once they have established a contact, *i.e.* once they are able to react. Infection in communicable human diseases spreads from place to place by jumping from person to person as infected people transmit their diseases to healthy people, which usually happens with close contacts between two people. Possible contacts are shaking hands, kisses, touching the infected surfaces with germs, mosquito bites, *etc.* Here individuals  $S$  and  $I$  are considered as free individuals before being in contact with each other, and then after being in contact with a given  $p_1$  probability, they become an  $SI$  *group-complex*. At this time the contact is not sufficient to provoke the transmission of infection. Then another probability term  $p_2$  is defined to determine the likelihood of effective contacts for infection transmission. This can be expressed as :



The intermediate group-complex terms are not only limited to  $SI$  group-complexes but can also result from the interactions between  $S$ ,  $I$  and  $R$  individuals. Therefore other *group-complex* products such as  $SR, RI$  or even  $SIR$  should be considered in SIR kinetic models. These can affect the epidemic process by giving a real role to  $R$  individuals in a population.

When in classical  $SIR$  models  $R$  individuals are considered just as a passage from the  $I$  compartmental class to the  $S$  class, the  $R$ -individuals have slightly similar importance to the  $S$ -individuals in providing the intermediate  $SI$  *group-complex*, the difference being the introduction of an average (time) delay of  $1/\lambda$ . In our modified  $SIR$  model,  $R$ -individuals are now considered as non-competitive inhibitors. Then their role is changed to produce  $SR$ ,  $RI$  and  $SIR$  *group-complexes*, which means that the production of  $SI$  *group-complexes* decreases when  $R$  inhibitors are present: a lower number of free  $I$  individuals in a given population will be able to make contacts with free  $S$  individuals.

Consequently, depending on the importance allowed to  $R$  individuals in an epidemic process (depending on whether one considers a classical  $SIR$  model or our modified one), we can define two different types of SIR kinetic reactions as follows:

### **2.2.1.1 A modified version of the classical SIR model: SIR kinetics without resistance (and with Delay)**

Here we present how we modified an  $SIR$  model by applying the formalism used in enzyme kinetics.

An  $SIR$  model without resistance can be represented by a simple succession of reversible

or irreversible enzymatic processes in which  $S$  individuals are infected by contact with  $I$ -individuals (in the form of bimolecular reactions), and the  $I$  individuals are converted into individuals being temporarily "resistant" or "recovered"  $R$  people. Finally,  $R$  people may gradually lose their "immunity" and become susceptible  $S$  again. In this model,  $R$  individuals can naturally not be infected, but do not participate in possible interactions with either  $S$ ,  $I$  or  $SI$  group-complexes.

In this context,  $R$  individuals do not play the role of "resistants", *i.e.* inhibitors of the infection as it will be discussed in section 2.2.1.2. This class consists of isolated individuals without any possible (bimolecular) reaction with other individuals belonging to  $S$  or  $I$  class. Then they cannot play the role of the resistance fronts (barriers) against epidemic spread. This doesn't corresponds to real situations in which all types of individuals ( $S$ ,  $I$  or  $R$ ) in a population interact and mix together (indeed, the population mixing patterns, including the speed of mixing, can have significant impacts on epidemic dynamics see [100] ). Then the presence of  $R$  individuals is strictly limited here to introducing a delay in the transition from  $I$  to  $S$ .

A fast way to express the SIR epidemic process with delay is as follow:

- Infection dynamics :  $S + I \xrightarrow{\beta_{app}} I + I$
- Recovery dynamics :  $I \xrightarrow{\gamma} R$
- Loss of immunity dynamics:  $R \xrightarrow{\lambda} S$

where  $\beta_{app} = \beta/[N]$  is an apparent constant, in a second-order reaction and is expressed

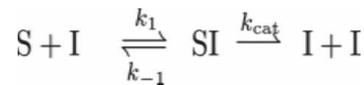
in  $km^2.individual^{-1}.day^{-1}$ . This results from an adjustment by the total density  $[N] = [S] + [I] + [R]$  (all in  $individual.km^2$ ), with  $\beta$  being a catalytic rate constant in a first-order reaction (in  $day^{-1}$ ). Moreover  $\gamma$  and  $\lambda$  are the constant rates of the monomolecular reactions of recovery and loss of immunity both expressed in  $km^{-2}.day^{-1}$ . Then the variations of the population density (in  $individual/km^2$ , where  $A$  is the surface in  $km^2$  of the environment where the epidemic process develops) in  $S$ ,  $I$  and  $R$  can be formulated by a set of ordinary differential equations as follows:

$$\left\{ \begin{array}{l} \frac{d[S]}{dt} = -\beta_{app}[S][I] + \lambda[R] \times A \\ \frac{d[I]}{dt} = \beta_{app}[S][I] - \gamma[I] \times A \\ \frac{d[R]}{dt} = \gamma[I] \times A - \lambda[R] \times A \end{array} \right.$$

It has to be noted that  $\beta_{app}$  is considered as a constant rate. Here  $\beta$  is modulated by  $[N]$  where  $[N] = [S] + [I] + [R]$  meaning that here the role of  $R$  is the same as the one of  $S$  or  $I$ , *i.e.*  $R$  is intended to play a role in the infection process itself. As we will see, in the modified *SIR* model with resistance (section 2.2.1.2), the role of  $R$  is different; in reality  $R$  doesn't play a role in the infection process itself. Moreover the rates of the epidemic reactions (infection, recovery and loss of immunity) differ strongly: in most cases, the infection is largely faster than the rate of recovery and "infinitely" faster than the loss of immunity. Then one can limit the epidemic process to the infection process, in which only  $S$  and  $I$  play a role. Compared to the turnover of group-complex formation, the rate of infection is "infinitely" lower. As

mentioned before,  $\beta$  is intended to be modulated by  $[N]$ , then once considering the last evoked points, its adjustment depends only on the protagonists that endure the infection, *i.e.* meaning  $[N] = [S]$  (see Annexe A.3). In section 2.2.1.2 we will see how this can influence the infection dynamics in an epidemic process, because while a lot of  $R$  or  $I$  individuals appear in a population,  $[N] = [S]$  becomes small. Then the adjustment of  $\beta$  we make here is totally different from the case where it is modulated by a value of  $[N]$  equal to the total sum of individuals in each class ( $[N] = [S] + [I] + [R]$ ).

The SIR epidemic process is described under enzyme kinetics dynamics as an  $SI$  epidemic system with a delay (influence of  $R$ ) as following schematic :



where  $k_1$ ,  $k_{-1}$  are the constant rates and  $k_{cat}$  is catalytic rate.

The catalyst role of an enzyme corresponds here to  $I$  individuals that cause the  $S$  individuals (seen as substrates) to change their status and become infected. The individuals in these classes are considered to be *free individuals* or *group-complexes* composed of individuals of the same class. Such individuals have to be in close contact (*i.e.* like to be combined in an enzymic reaction) by providing the *group-complex*. Once these intermediate group-complexes (here  $SI$ ) are formed, the infection process may be accomplished. The *group-complex*  $SI$  is of a different nature compared to simple group-complexes (or free individuals)  $S$  or  $I$ . Then only the  $I$  individuals present in  $SI$  group-complexes count in the infection process. The formation of  $SI$  intermediate complexes is due to reactions associated

with a second-order rate constant  $k_1$  (in  $km^2.individual^{-1}.day^{-1}$ ) and their dissociation into free  $S$  and  $I$  results from reactions associated to a first-order constant rate  $k_{-1}$  (in  $day^{-1}$ ).

The dissociation constant described by  $K = k_{-1}/k_1$  (in  $individual.km^2$ ) is a constant rate that determines the average turnover of formation and dissociation of  $SI$  group-complexes. The catalytic rate  $k_{cat}$  (in  $day^{-1}$ ) drives the reaction to the final reaction product (*i.e.* occurrence of a new infection outbreak), *i.e.* new free  $I$  individuals in the population.

Then  $\beta_{app}$  can be described by a combination of these above constant rates involved in kinetic dynamics. This leads to the following equation (see appendix A.1 for more details) :

$$\beta_{app} = \frac{k_1 k_{cat}}{k_1 [N] + k_{-1} + k_{cat}} = \frac{k_1 k_{cat}}{k_1 [S] + k_{-1} + k_{cat}} \quad (2.3)$$

As known in Michaelis-Menten kinetics, the rate of enzymatic reaction, here in an epidemic context, the infection rate (infection flux) can be derived by :

$$\frac{d[I]}{dt} = -\frac{d[SI]}{dt} = k_{cat} [IS] = \frac{k_{cat} [I][S]}{[S] + \frac{(k_{-1} + k_{cat})}{k_1}} = \frac{k_{cat} [I][S]}{[S] + K_M} \quad (2.4)$$

The Michaelis constant  $K_M$  (in  $individual.km^2$ ) corresponds to the dissociation constant (of  $SI$  group-complexes) adjusted by the catalytic rate  $k_{cat}$ . Its value determines a concentration threshold (*i.e.* here  $S$  population density) at which the reaction rate is increased to 50% of  $V_{max}$ , where  $V_{max} = k_{cat} [I]$  represents the maximal infection velocity that measures the number of  $S$  individuals that become infected per unit time. Consequently the infection flux is :

$$\frac{d[I]}{dt} = V_{max} \cdot f^{sat}(S, K_M) \quad (2.5)$$

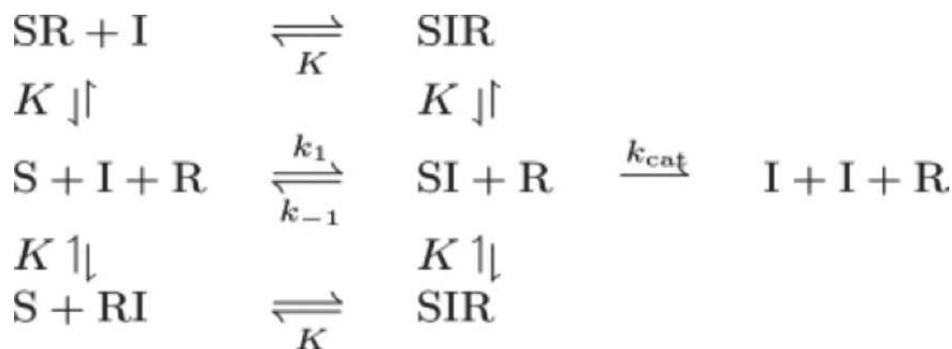
where  $f^{sat}(S, K_M) = \frac{[S]}{[S] + K_M}$ .

The saturation function  $f^{sat}(S, K_M)$  expresses the fact that occurrence of reactions that result from  $SI$  complexes, cannot be faster than the  $V_{max}$  when  $[S]$  is large enough (*i.e.*  $S$  in excess:  $[S] \gg K_M$ ) or it will approach to 50% of  $V_{max}$  when decreasing  $[S]$ , precisely when  $[S] = K_M$ .

As explained before, in this  $SIR$  model without resistance, there is no interaction between  $S$  and/or  $I$  with  $R$  individuals class. The  $R$  class seems to be invisible as their "resistant" role is not integrated in the kinetic-based model. In the epidemic models based on  $SIR$  kinetics with resistance (as we will see in the next section),  $R$  individuals are considered to be inhibitors that prevent, somehow, the infection process. We will see through the simulation results how the "resistance" factor can affect epidemic dynamics.

### 2.2.1.2 A modified version of the classical $SIR$ model : $SIR$ kinetics with resistance

By integrating the  $R$  individuals as non-competitive inhibitors in a kinetic system, the general  $SIR$  reaction scheme is described as follows :



Here the  $R$  individuals have the same role that non-competitive inhibitors have in enzymatic systems. A definition of non-competitive inhibition is given by "ChemWiki" : *in non-competitive inhibition, the inhibitor binds to the enzyme at a location other than the active site in such a way that the inhibitor and substrate can simultaneously be attached to the enzyme. The substrate and the inhibitor have no effect on the binding of the other and can bind and unbind the enzyme in either order.* In other words, the inhibitor may associate individually with the enzyme, the substrate, or the enzyme-substrate-complex. In the epidemic context,  $R$  individuals can be in contact with other  $S$  or  $I$  individuals and even with  $SI$  group-complexes (e.i.  $R + I \rightleftharpoons RI$ ,  $S + R \rightleftharpoons SR$ ,  $SI + R \rightleftharpoons SIR$  and  $SR + I \rightleftharpoons SIR$ ). In this SIR kinetic model, among all possible complexes, only the  $SI$  might directly result in an infection transmission. Note that in our model we assume that  $R$  is a perfect inhibitor meaning that there is no possible infection from  $SIR$  group-complexes. The other group-complexes that include  $R$  individuals are important because of their role in affecting the formation of  $SI$  group-complexes. In other words  $SI$  group-complexes form less frequently.

The processes of formation and conversely dissociation of each *group-complex* are related to given association/dissociation constant rates (e.g.  $S + I \xrightarrow{k_1} SI$  or  $SI \xrightarrow{k_{-1}} S + I$ ). Since the nature of individuals ( $S$ ,  $I$  and  $R$ ) is the same for all individuals (e.g. they are all humans), i.e. the status  $S$ ,  $I$  or  $R$  is unknown, then they can form any *group-complex* (e.g. presumably people can not know who is infected or who is not). Therefore the rate constants are assumed to be all the same and equal to  $K = k_{-1}/k_1$ . They represent the *average turnover* of association/dissociation of all possible types of *group-complexes* (see annexe A.2).

It should be noted that the same value of  $K$  may be obtained by different values of  $k_1$  and

$k_{-1}$ . For example, the same dissociation constant  $K$  might result from both a slow formation and dissociation of *group-complexes* (meaning that less groups are formed ( $k_1$  is small, e.g.  $10 \text{ day}^{-1}$ ) and last for long time because of a slow reverse dissociation process ( $k_{-1}$  is small too e.g.  $10 \text{ km}^2 \cdot \text{individual}^{-1} \cdot \text{day}^{-1}$ ), or on the contrary it can result from a case in which *group-complexes* are formed quickly (many groups form with a important value of  $k_1$ , e.g.  $10000 \text{ km}^2 \cdot \text{individual}^{-1} \cdot \text{day}^{-1}$ , these groups lasting for a very short time, and an important value of  $K_{-1}$  too, e.g.  $10000 \text{ day}^{-1}$ ). We will see in the section concerning the group dynamics ( 2.2.2.1) and the models based on group dynamics (sections 3.1, 3.2.1 and 3.2.3) how the dissociation constant  $K$  (and its components) can impact the epidemic dynamics.

By reformulating the enzymatic kinetic equations (see details in appendix A.2) in the context of epidemic processes including  $R$  individuals as inhibitors, we obtain :

$$\beta_{app} = \frac{k_1 k_{cat}}{k_1 \left(1 + \frac{[R]}{3K}\right) [S] + (k_{-1} + k_{cat}) \left(1 + \frac{[R]}{K} \left(1 + \frac{2[S]}{3K}\right)\right)} \quad (2.6)$$

then the equation 2.4 of the infection flux can be rewritten as follows :

$$\frac{d[I]}{dt} = \frac{k_{cat} [I]}{\left(1 + \frac{[R]}{3K_A}\right)} \times \frac{[S]}{[S] + \frac{(k_{-1} + k_{cat})}{k_1 \left(1 + \frac{[R]}{3K}\right)} \left(1 + \frac{[R]}{K} \left(1 + \frac{2[S]}{3K}\right)\right)} \quad (2.7)$$

$K_M$  (the Michaelis constant rate) and  $V_{max}$  (the maximal infection velocity) are affected by the presence of  $R$  individuals as inhibitors (resistants) in the SIR kinetic system :

$$V_{max\ app} = \frac{k_{cat}[I]}{\left(1 + \frac{[R]}{3K_A}\right)} = \frac{V_{max}}{\left(1 + \frac{[R]}{3K_A}\right)} \quad (2.8)$$

$$K_{M\ app} = \left(\frac{k_{-1} + k_{cat}}{k_1}\right) \frac{\left(1 + \frac{[R]}{K_A} \left(1 + \frac{[S]}{3K_A} + \frac{[S]}{3K_B}\right)\right)}{\left(1 + \frac{[R]}{3K_A}\right)} = K_M \frac{\left(1 + \frac{[R]}{K_A} \left(1 + \frac{[S]}{3K_A} + \frac{[S]}{3K_B}\right)\right)}{\left(1 + \frac{[R]}{3K_A}\right)} \quad (2.9)$$

These effects are clearly explicable: the presence of  $R$  individuals in a homogeneous space (there are different effects in a heterogeneous space, see chapter 3) leads to the formation of *group-complexes* such as  $SR$ ,  $RI$  and  $SIR$  to the detriment of forming the  $SI$  group-complexes, those leading to the infection process. This causes the quick saturation in the formation of  $SI$  group-complexes (the reaction rate reaches 50% of the  $V_{max}$ ) in the presence of  $R$  individuals. In conclusion it is more difficult to provide new infections in the presence of resistances: it is understandable since one thinks that the infection process results from collisions between  $I$  and  $S$  individuals. Due to the presence of  $R$  individuals, these collisions are less likely to happen, therefore the maximal infection velocity ( $V_{max\ app}$ ) is affected.

### 2.2.2 Group Dynamics in a Population

There are different manners of defining *groups* in group theoretical studies. [70] mentioned some definitions of groups and cited, notably, Forsyth's definition as the most helpful one:

*Hundreds of fish swimming together are called a school. A pack of foraging baboons*

*is a troupe. A half dozen crows on a telephone line is a murder. A gam is a group of whales. But what is a collection of human beings called ? A group is collections of people may seem unique, but each possesses that one critical element that defines a group: connections linking the individual members, members are linked together in a web of interpersonal relationships. Thus, a group is defined as two or more individuals who are connected to one another by social relationships [Donelson R.Forsyth (2006)].*

There are two sets of categories in which to classify groups: groups can be known as *primary groups*, like families and friendship circles with intimate interactions, or they can be categorised as *secondary groups* that are formally organised (*i.e.* in a certain context, like organisations, institutions and companies, cultural events, etc.), their members not necessarily being in direct contact with each other.

Groups may be classified as *planned* or *emergent* groups. Planned groups are specifically formed for a specific purpose (business meetings, concerts, etc.). Emergent groups appear relatively spontaneously when people find themselves together in the same place, or when the same collection of people gradually comes to know each other through conversation and interactions over a certain period of time (Cartwright and Zander 1968 in [70]). According to Arrow et Al (2000) (see [70]) *emergent* groups are circumstantial when unplanned and often temporary groups develop when external forces bring people together (*e.g.* people in the metro or in the bus queue) and called self-organising emergent groups when people gradually cooperate and engage with the others around some task or interest. The differences in the definition of group occur because the authors usually have different criteria for group

existence according to their centre of interest (Benson 2001 in [70]).

Despite these classifications, groups are not always categorised as an absolute type. This means that in different circumstances, group types can switch from one category to another, and a group might consist in sub-groups or, on the contrary, split into sub-groups, each having different types. As an example, educational institutions (*e.g.* high-schools) are classified as *secondary* groups that can contain other *secondary* (and sometimes *primary* like groups of friends) sub-groups. A number of students in the school meal queue is a circumstantial emergent group. Any spontaneous cooperation among groups of students, their families, teachers or administration employees can be classified as self-organising secondary groups. In an urban context, in big cities for example, in every office, bar, shop, street corner, different groups of people are constantly forming and splitting up. They are sometimes ephemeral (*e.g.* the time to shake hands with a colleague in the corridor : 2 seconds), sometimes longer (2 hours or a multi-days for a conference). Then to avoid being limited to promiscuity questions such as *where does the group stop?*, we should define groups based on their purpose, *e.g.* different groups are formed for several reasons: a conference, meeting a friend, a purchase in a shopping store etc. In their social lives, individuals have different goals and interests to form different groups with different numbers of people, for example a small group of friends at a bar or a large number of people in a concert.

Understanding correctly the group phenomenon needs more socio-behavioural studies at the individual level and more investigation into the factors that play an important role in social aggregations. However, to model group dynamics we need to simplify these real examples instead of investing too much in individual behavioural characteristics (somehow

this is possible in individual-based models) and then to focus more on less realistic but more controllable simplified models of groups inspired by mathematics and physics.

We propose to limit the description of groups to three characteristics: (i) the frequency at which groups appear, (ii) their size and (iii) their lifetime. It is reasonable to consider that *group-appearance*, *group-sizes* and *group-lifetimes* obey the probability laws (see sections 2.2.2.1 and 2.2.2.3), notably in the case of a sufficiently large population (the question is *how big?*).

Another manner of giving a structure to simulated populations is to use pre-topological data to define how big some groups are and how long they last (*e.g.* number of student in a given class during their mathematics lecture) and how they live (which person knows another, where do people usually spend their time...), or to introduce a new variable that describes whether a given individual is more likely to form a group with others or not, for example by considering a grouping tendency ( $0 \leq \alpha < 1$ ) in multi-agent modelling (see sections 2.2.2.2 and 2.2.2.3). We can go further and use participatory simulations in which all or part of the agents are the avatars played by real humans. In such serious games a player can decide to form a group with other people.

These assumptions are still debatable, but they allow us to model group dynamics and study epidemic processes in more realistic conditions. Of course these assumptions are less implementable in ODE, PDE and homogeneous-space stochastic models than in individual-based models or group-based models.

Group formation is based on each individual's grouping tendency. It is not necessarily easy to simulate this when we are not dealing with individual-based numerical models. Then we must define a group-size distribution. The study of group formation dynamics in animals shows scaling aspects of group-size distribution (*e.g.* exponential distribution). For example the modelling of the herd size of mammals (*e.g.* African buffaloes) or the group-size of tuna fish or sardines shows a truncated power law distribution [15, 79]. According to Okubo ([15]), the group-sizes are exponentially distributed: *Any group-size distribution should be exponentially decreasing by applying a maximum entropy principle to the distribution under the constraint of fixed average size, which implicitly includes the strong assumption that there exists a well-defined mean and therefore overlooks slowly decaying distributions such as power laws with  $b \leq 2$ . It is well known to physicists that such a procedure leads to exponential (Gibbs-Boltzmann) distributions.*

These group-size (connectivity-size) distribution properties have also been found in complex systems with large connectivity features. In [7], the distribution of local connectivity of many large random networks shows a power law scaling. In an epidemiological context, [5] described by stochastic simulations of how scaling laws may arise from the distribution of various population sizes (the number of sexual partners considered as group-sizes). in [8], the statistical analysis of realistic social networks based on empirical or semi-empirical data (for six different populations) shows a best fit of an exponential degree distribution for all populations among Poisson, pure power law and truncated power law distributions.

Like social networks in [8], we propose to use an exponential distribution to represent the group-size variable. In brief, we can identify a group as a collection of individuals who participate in a particular goal and having different connection degrees during a specific time; the group lifetime.

The group lifetime is not particularly well developed as an important key element in group theories, but from the epidemiological point of view, this may play a crucial role. Indeed, as long as an individual stays in a given group, there is a higher likelihood that they are in contact with other members and this potentially increases the risk of being infected by an infectious agent. For example, let's consider one infected individual (infected by pathogens like Flu, gastroenteritis, etc.) among a number of employees in a company. These individuals, during an eight hour work a day, might split off into different sub-groups (*e.g.* several people in a meeting room, in a cafeteria or in the same office...) and spend a certain time with each others, a time related to the context of that group. The individuals who stay a longer time with the given infected person are more likely to be infected than others who have less direct contact with this person.

Then a group can last a couple of days, hours or just a few minutes, and even just momentarily (*e.g.* shaking hands with a colleague: 2 seconds) can be significant in increasing the probability of being infected by an infectious agent in a given group. In fact, it is more likely that a group with an important size (*e.g.*  $n = 1000$  persons such as in a concert) last for an average of 2 hours while the smaller groups ( $n < 5$ ) can last just 2 seconds (the most probable lifetime) as well as more than 4 hours (*e.g.* a long business meeting, less probable event), or even a few days (*e.g.* two people that go camping, an even less probable situation).

Group dynamics can be affected by other characteristics such as the population density, the individual movement speeds and the proportion of free individuals (those that don't belong to any group) compared to grouped individuals. However, these three group characteristics (group-appearance frequency, group-size and group-lifetime) as probabilistic variables are not necessary in the cases where organisational structures and rhythms are pre-defined (like in high-school classes where a given daily working-hours or formal meeting schedules are defined in advance). In the following subsections we give details on these topics.

### 2.2.2.1 Modelling the group-size in homogeneous SSA-based models

Group dynamics presented here are based on a set of assumptions on individual behaviours. These assumptions are still debatable but they allow us to model group dynamics and study epidemic processes in more realistic conditions. We will show how the appearance and the size of groups can be modelled.

In a simulation of group dynamics three characteristics are important: *group-appearance* ( $B_{g_i}$ ), *group-size* ( $S_{g_i}$ ) and *group-lifetime* ( $LT_{g_i}$ ). Group appearance is the time between the appearance of two successive groups ( $g_i$  and the previous group  $g_{i-1}$ ) and is called  $B_{g_i}$ . This random variable plays an important role in non individual-based models since there is no emergence in such models like in MAS. As non individual-based models correspond to event-driven simulation methods (meaning that the simulations have to be determined by a succession of events, here with groups appearance), we must impose the manner that these events will occur. We assume here that when one group forms at a certain time, no other groups can form at the same time, implying that a certain time separates the formation of two

successive groups. We also assume that groups are not formed by merging with other existing groups or by adding single individuals (nor can any group be split into new sub-groups). To obtain the formulation of group formation dynamics we suppose a simplified view of how populations behave, *i.e.* we suppose that groups form due to random collisions between individuals, their movement being based on a simple random walk process with a given speed like the molecules in a fluid that constantly collide between each other. Of course we are conscious that such a model is far from reality, since people don't move randomly.

Group appearance  $B_{g_i}$  and group size  $S_{g_i}$  are intimately related. Indeed, gathering a large number of people in a given place, takes time. Then this one can expect that the time between the formation of two groups is long enough when this group size is expected to be large. Inspired from the concepts of the kinetic theory of gasses, the successive values of  $B_{g_i}$  are randomly chosen regarding to the following *the probability density function*:

$$P(t) = \frac{1}{\bar{\tau}} e^{-t/\bar{\tau}}$$

where  $\bar{\tau}$  is *mean free collision time*, which is obtained by (a) the *free* population density at time  $t$  per effective surface  $F(t)/s_{eff}$ , where  $s_{eff}$  is the effective surface frequented by individuals normalised by the total area ( $s_{eff}$  is proportional (e.g 1%) to the total area); (b) the effective diameter (the average interaction radius)  $d$ , that determines the actual occupied space by one individual or the average distance between two individuals considered to have established a contact. Then in the SIR kinetic process, the average distance for an efficient

contact during the time  $t$ , is defined by  $\beta_{app} \times t \times 1 \text{ individual}$ . For example, in a given communicable disease with  $\beta_{app} = 5.10^{-5} (m^2 \cdot indiv^{-1} \cdot day^{-1})$ , if the contact between two individuals comprises between 10 and 60 minutes, then the effective average distance for an efficient contact (effective transmission of disease) comprises between 35 cm and 210 cm respectively. This seems acceptable because in a real situation, 35 cm might represent the distance between two people in an elevator or in a concert; a distance of 2.1 m could represent the standard distance between two individuals at work in an office or at a meeting.  $\bar{\tau}$  depends also on (c) the average relative velocity  $\bar{v}_{rel}$  calculated from the speed of individuals  $\bar{v}$  (here assumed constant).

$$\bar{\tau} = \frac{s_{eff}}{F(t) \times 2d \times \bar{v}_{rel}}$$

consequently  $S_{g_i}$  is randomly chosen according to the following density probability function :

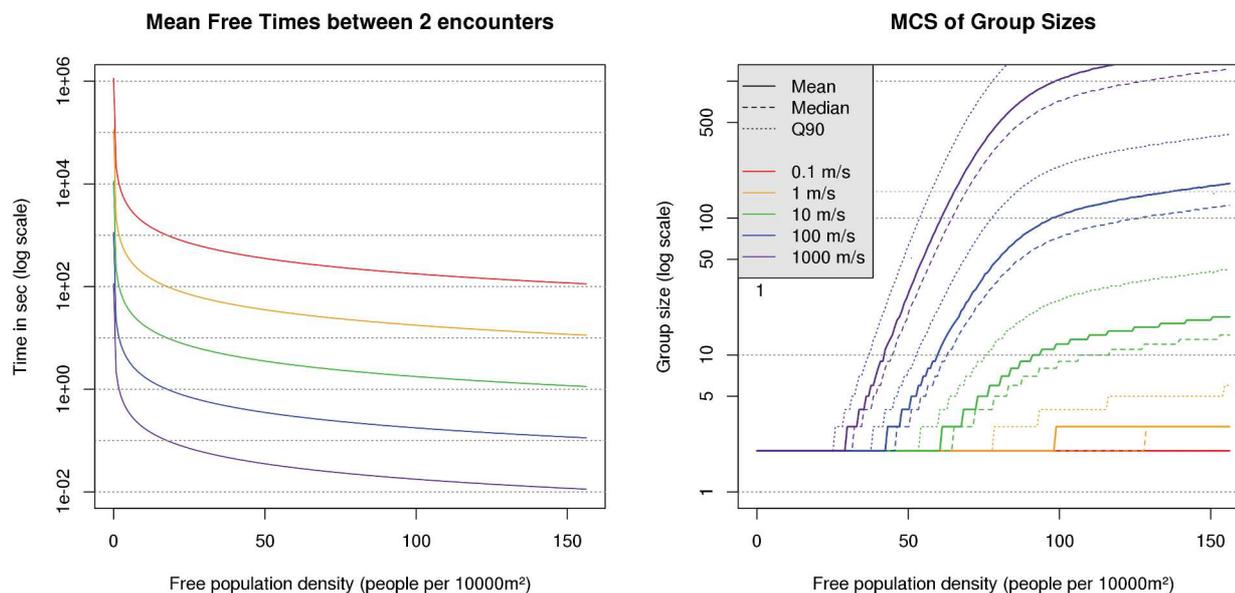
$$P(s) = \bar{\sigma} e^{-\bar{\sigma}s}$$

where  $\bar{\sigma}$  is calculated by :

$$\bar{\sigma} = \frac{B_{g_i}}{2} \frac{1}{1 + \left(\frac{F(t)}{N(t)}\right)^2}$$

this formulation shows how group size ( $S_{g_i}$ ) depends on the value of  $B_{g_i}$  and on the proportion of *free* individuals (*i.e.* the people available to form groups) in the total population (as expressed by a sigmoid function, here the Hill equation). This equation, a non-linear relation, clarifies the fact that by reducing the number of *free* individuals in a given population, the formation of new groups becomes more difficult. However these empirical formulations are

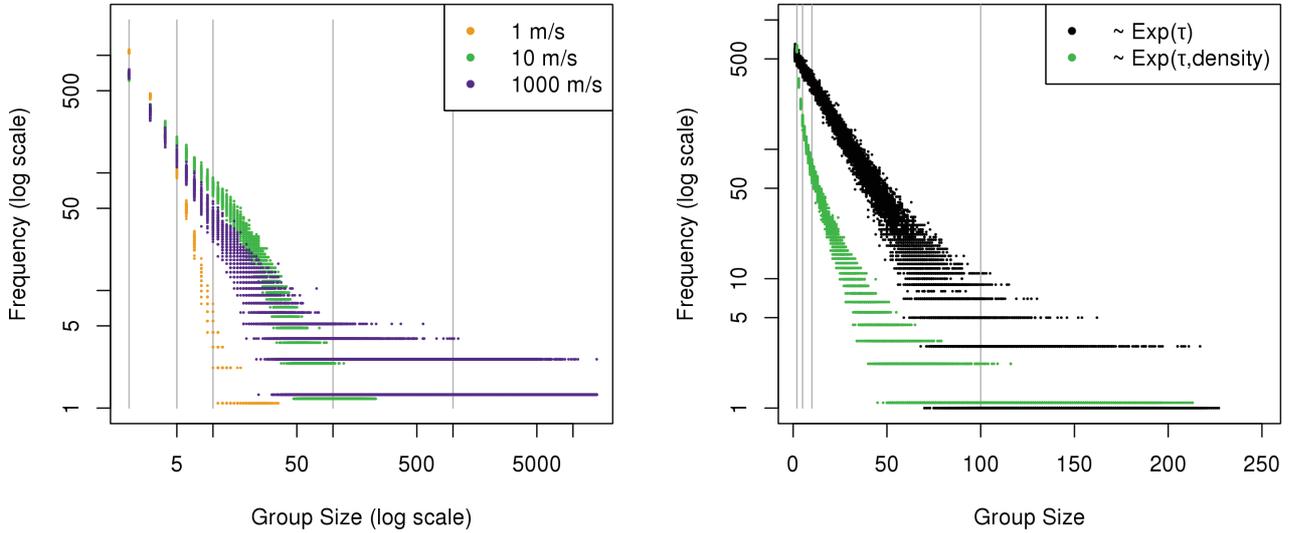
still debatable and need more investigation on this subject, but nevertheless it allows us to give a structure to a population in a homogenous model. The obtained dynamic behaviour seems reasonable in comparison to that obtained in multi-agent based systems (see chapter 3). At the individual level, groups are formed by encounters (unexpected or casual meetings of some individuals). Then it is important to verify that the density of *free* individuals and their speed have the expected impact on the mean time between two encounters and on the resulting group-sizes. The calculated effects are shown in figure 2.5. We can see notably (in right image) that when the density is decreasing, the group formation becomes difficult. Moreover, increasing the average speed shortens the mean free time to meet two individuals and allows the individuals to meet easily. Even with those who are far, leading potentially to the formation of big groups.



**Figure 2.5 Impacts of Free-population Density and Movement Speed on Mean-free-time and Group-size :** (left) The average free time between two encounters (here considered as particles that move randomly) is shown as a function of the average movement speed of individuals and free-population density. (right) A Monte Carlo simulation (MCS) of group-formation shows while the density is decreasing, the mean free time between two individuals encounters becomes long then group-formation is more difficult (lack of free (sole)-individual). By increasing in  $\bar{v}_{rel}$ , the groups are quickly formed by individuals even those that are separated by long distances  $w$ , which helps to form the big groups.

The impacts of the individual speed on group-sizes are compared (see figure 2.6-left) when increasing the speeds from  $1m/s$  to  $10m/s$  and up to an unrealistic speed of  $10^3m/s$ . The Log-Log plot shows that all possible sizes of groups can be formed when the individuals have a considerably high speed (m/s). The plot shape seems less linear at lower speed and the maximum size of groups does not exceed 50 to a few hundred people (speed  $1m/s$  to  $10m/s$ ). The group-size distributions (figure 2.6-right) illustrates the difference in behaviour when taking into account the density of the *free* individuals in the population as a sigmoid

function  $\left(\frac{1}{1+\left(\frac{F(t)}{N(t)}\right)^2}\right)$  or not, when calculating the exponential distribution parameter  $\bar{\sigma}$  (*i.e.* when the density term can be considered negligible  $\frac{F(t)}{N(t)} = 0$ ).



**Figure 2.6 Group Size Distribution.** (Left) Increasing the movement speed to an unrealistic speed of 1000( $m/s$ ) allows us to observe all possible group sizes in a given large population. At slower speed levels, the log-log plots lose their linearity, however the pattern remains linear (approximately) for all small groups while the groups are less than 50 individuals. At lower individual speeds the formation of considerably large groups is less likely than the small groups, in which case the big groups are considered as the rare events. (Right) The log-log plot of group size compares two exponential distributions of group-size, one (in green) when the effect of the free individuals density considered in calculation of parameter  $\bar{\sigma}$  and one when it is not considered (in black). As we can see, the occurrence of groups (including all different sizes) is less frequent when the density term is considered.

Once the group size is determined, its composition is obtained by a multivariate draw depending on the group size and on the availability of individuals  $S$ ,  $I$  and  $R$  in the population of free individuals (without physical interaction with the others). The multivariate draw is ensured by using a hypergeometric distributed variable.

In this section we have proposed a procedure to compute the group characteristics. The group-formation is not the result of individual decisions nor emergence. This will be different

in individual-based models (MAS or CA). In the latter one would expect a group formation by way of emergence, which is not something that is easy to obtain; in particular, in an MAS with random walking agents it would be difficult to obtain big groups. Moreover, one can ask the question, what are the boundaries that define groups ? In the next sub-section we will present a group-based multi-agent system where individuals can meet and form groups, can merge existing groups, and where the groups life is decided at the group level (and in which the SIR processes will occur).

### 2.2.2.2 Modelling the group-size in MAS-based Models

In multi-agent modelling, individuals have their own behaviours; here they simply move randomly. When individuals are sufficiently close together they can form a group. Here, we suggest a group formation procedure based on a random probability number ( $0 \leq \alpha < 1$ ) attributed to each agent (individual) that represents its grouping tendency. Let  $A_i = (x_i, y_i, \alpha_i)$  be an agent represented by its own (x,y) spatial coordinates and its grouping tendency. The variable  $\alpha$ , represents how likely an individual is to engage in a group with its own neighbours. Individuals with the distance of less than  $m$  meters from each other are considered in a neighbourhood. In our simulations we have chosen a neighbourhood of  $m = 4$ . Each individual explores its own neighbourhood and can join an existing group if the group tendency condition is still met:

$$d_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \leq m$$

The individuals are grouped if the average of group tendencies in a given neighbourhood with  $n$  individuals is greater than 50 percent:

$$\frac{\sum_i \alpha_i}{n} > 0.5$$

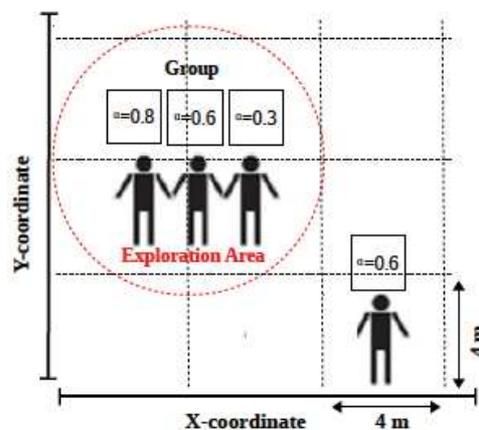


Figure 2.7 Group Formation of  $n = 3$  Individuals: The mean grouping tendency = 0.56. The individuals can move, by a given speed (1 m/s), in the simulated area to explore other individuals in their neighbourhood to form a new group or join to an existent one.

Simulations of structured populations (having a group formation) show that several groups form (or split out) at each  $t$  time (frequently). Group sizes vary from small groups (*e.g.* two colleagues at break-time or in a meeting room ...) to bigger groups (*e.g.* a group of students in a classroom ...). Figure 2.8 shows that in a multi-agent simulation, the formation of small groups is more frequent than for big groups. With these chosen neighbourhood, move speed and grouping tendency distribution we often observe a maximum size of 15 individuals.

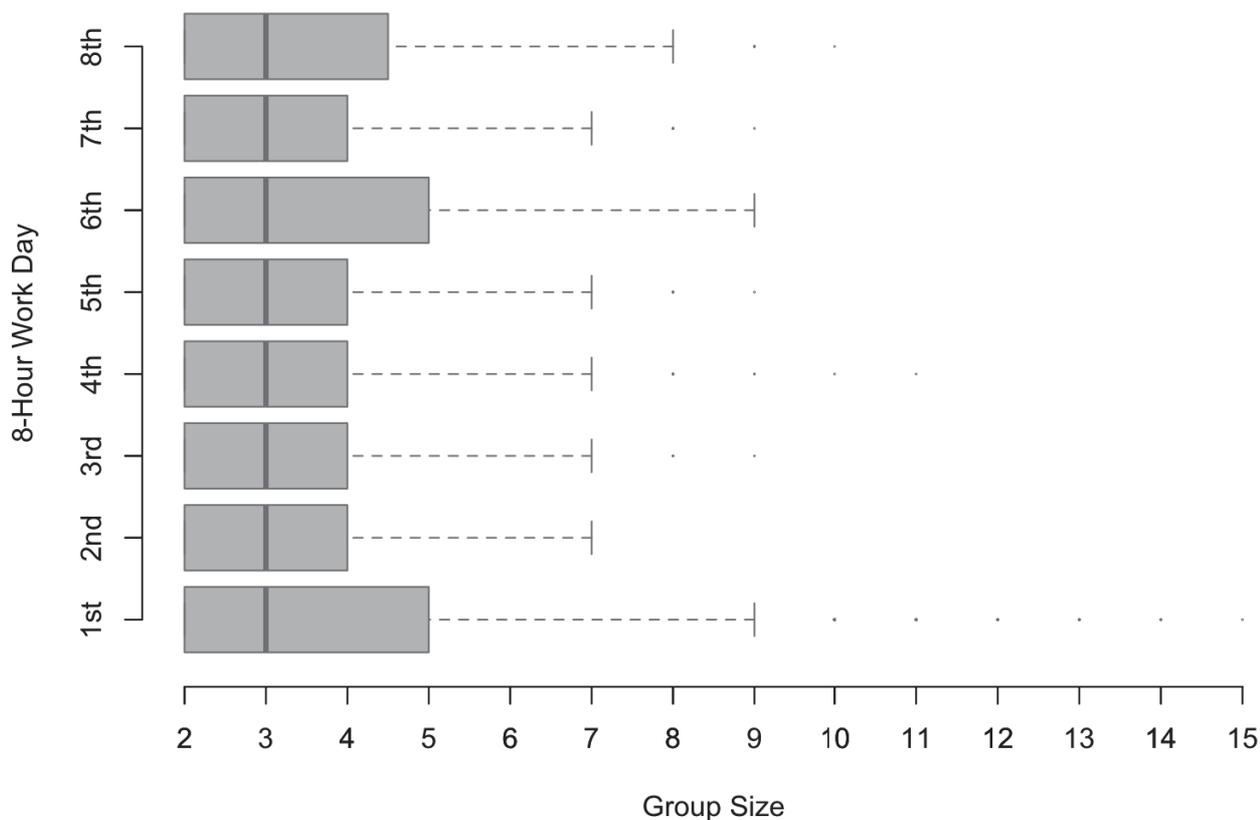


Figure 2.8 **Box Plot of Group-sizes During Simulated 8-hour Work Day:** Group-size varies between 2 to 15 during each one hour simulation time but the average group-size is approximately  $n = 3$  people. In this model we suppose that the groups are not merged and an agent's movement speed is  $1(m/s)$ .

In the MAS-based model, group formation (and consequently group-size) is an emerging characteristic. We expect that the group-size distribution is exponentially decreasing. The figure 2.9 shows four different exponential distributions that we intended to fit with group-size data observed in our MAS simulations. One can see that the best fits are obtained for values of  $\lambda$  between 0.5 and 0.8. We also did Monte Carlo simulations of group size distributions given an exponential model with a certain  $\lambda$  parameter, and compare them with the emerging group size distributions obtained in our MAS simulations. Figure 2.10 confirms that the best fit of the observed MAS simulation data and an exponential model is obtained for a

parameter  $\lambda$  close to 0.5. This exponential approximation seems more meaningful for small group sizes than for big groups. In section 2.2.2.1 we expressed the group size distribution by the probability  $P(s) = \bar{\sigma}e^{-\bar{\sigma}s}$ . Here the parameter  $\bar{\sigma}$  would then be equal to about 1/2. We also expressed this parameter by  $\bar{\sigma} = \frac{B_{g_i}}{2} \frac{1}{1+(\frac{F(t)}{N(t)})^2}$ . If we admit a fraction of  $F/N = 1$  then this would mean that  $B_{g_i}$  would be equal to about 2 seconds. This would correspond to a coherent average time between the formation of two groups of about 2 s in our MAS simulations. This is also coherent with the fact that the smaller groups are easier to form in such a time than big groups.

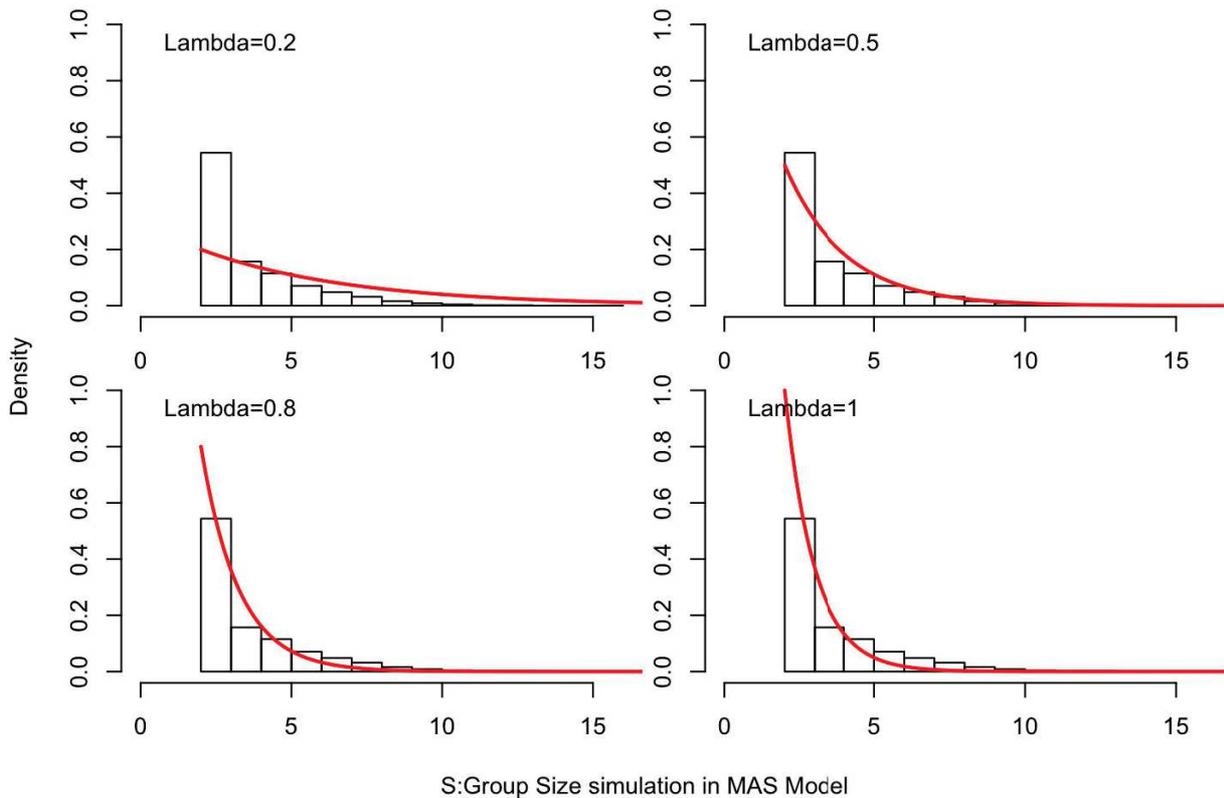


Figure 2.9 **Histogram of Group-size in MAS Model:** The exponential distributions (in red) with the rate parameter  $\lambda$  (0.2, 0.5, 0.8, 1) are traced to compare the best fit for group formation data. This shows an exponential distribution with  $\lambda = 0.5$  is almost the best fit.

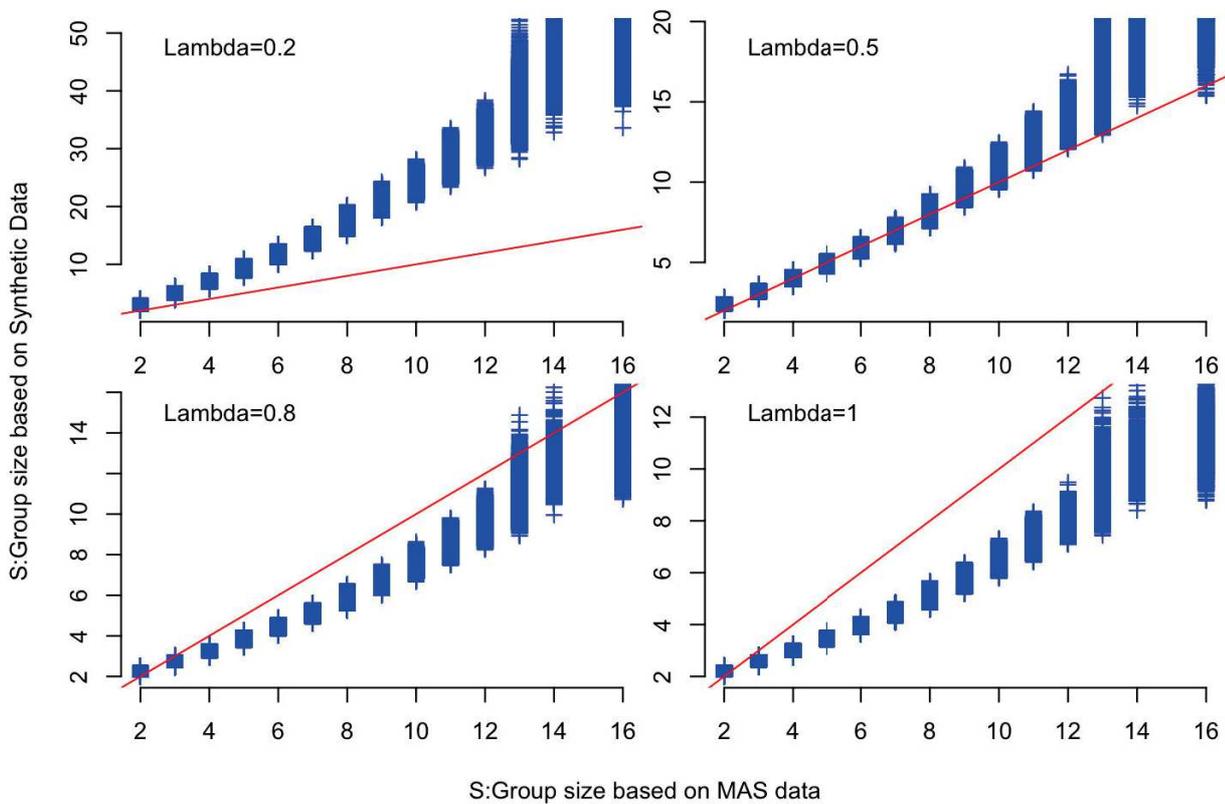


Figure 2.10 **The Goodness of Fit for MAS Data:** The synthetic data from the probability distribution of a given exponential distribution  $\lambda$  (0.2, 0.5, 0.8, 1) is compared with group-size distribution in MAS data. The goodness of fit is confirmed if the data from both synthetic and MAS model are located more and less on the red line ( $Y = X$ ). Bootstrap techniques are used with  $B = 1000$  replacement number.

### 2.2.2.3 Group-Lifetime

Group lifetime  $L_{g_i}$  is correlated with the group size, the size being often characteristic of a context (*e.g.* a concert, a couple, a meeting at work, a mathematics course class at school ...). Indeed, it is hardly conceivable that a concert consists of several thousand people lasting on average no more and not much less than 2 hours. However, there are many more opportunities in which groups of only 2 or 3 people can spend several days together or even more frequently they can just last for a small time (*e.g.* occasional meetings between

colleagues like when they shake hands, which takes only few seconds). In our models (SSA and MAS-based models) group lifetimes depend on group sizes. We suppose a gamma distribution with two parameters (shape and rate) for modelling the lifetime variable :

$$f(l; \alpha, \beta) = \frac{\beta^\alpha l^{\alpha-1} e^{-l\beta}}{\Gamma(\alpha)}$$

We propose the shape parameter to be  $\alpha = \text{group-size } (S_{g_i}) \times \text{coefficient } (A)$  and the rate parameter to be  $\beta = \text{group-size } (S_{g_i}) \times \text{coefficient } (B)$ . This means that the probability density function depends on the size of groups and that we can adjust the mean lifetime by the coefficients  $A$  and  $B$  such as:

$$E(L) = \frac{\beta}{\alpha} = \frac{A}{B}$$

This allows us to generate an appropriate distribution as an exponential distribution when group sizes are less than 20 and a gamma bell curve to a gaussian distribution for larger group sizes.

Then we modelled group lifetimes (both in SSA or MAS based simulations) by using a ratio  $A/B = 120 \text{ minutes} = 2 \text{ hours}$ , *i.e.* the average expected lifetime of a very large group (see figure 2.11).

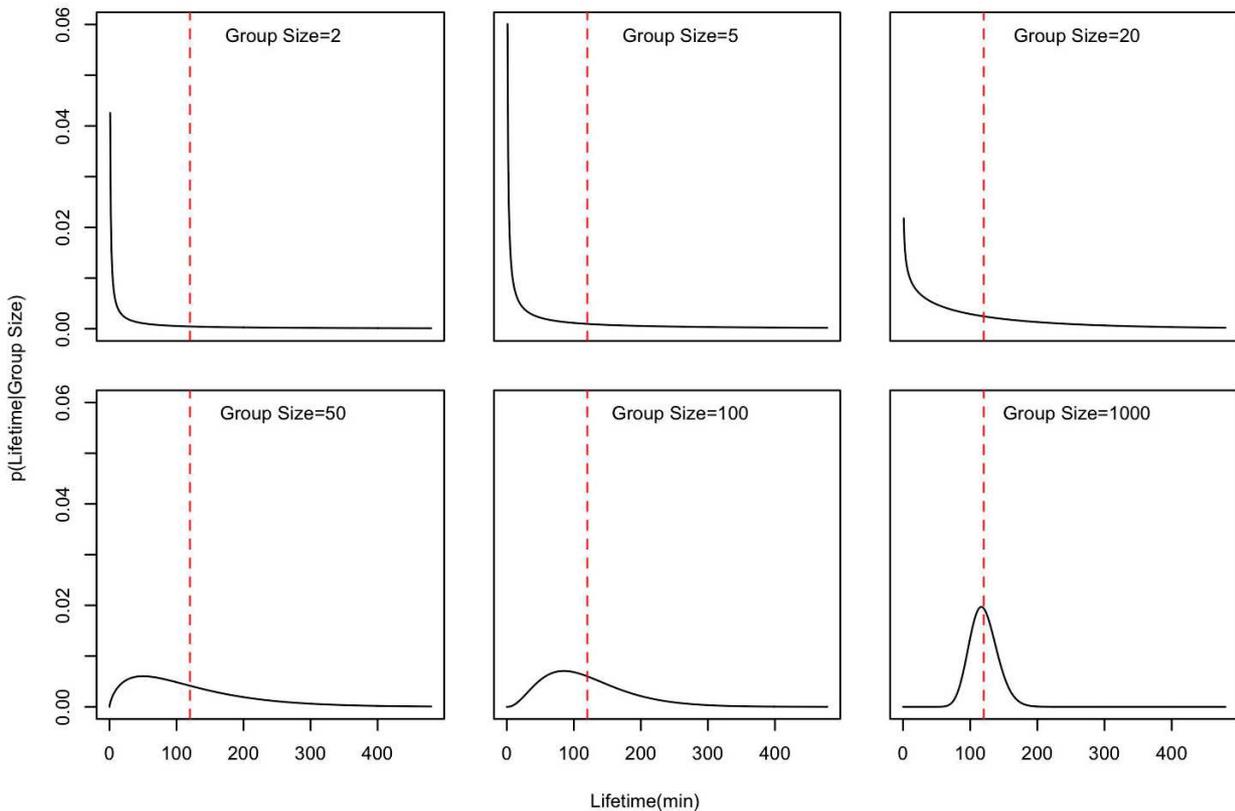


Figure 2.11 **The Gamma Probability Density Function for Different Group-size:** When groups have small sizes ( $< 5$ ), the distribution of their lifetime is identical to the exponential distribution ( $\alpha \approx 1$ ), this is the same for groups with average size (10 – 20 persons) but by increasing the lifetime variance. When groups reach an important size ( $> 30$ ), their distributions look like the bell shaped curves that localise more and less around the mean lifetime ( $L$ ), this signifies that groups with an important size ( $S = 1000$ ) last on average, approximately 2h (*e.g.* concerts etc.).

In the MAS model, a participatory procedure is used to compute lifetimes of each group as a collective decision of all members. In a given group with  $n$  members, a particular random variable  $\tau$  of the Gamma probability density function is attributed to each agent. These timing tendencies ( $\tau_1, \tau_2, \dots, \tau_n$ ) combine with the grouping tendencies ( $\alpha_1, \alpha_2, \dots, \alpha_n$ ) of

the  $n$  members of a group to allow us to compute the emerging (a consensus) group lifetime :

$$L = \frac{\sum_{i=1}^n \alpha_i \tau_i}{\sum_{i=1}^n \alpha_i}$$

This means that the agents with strong grouping tendencies will contribute more than the others to determine the group lifetime. In the real world, we are surrounded by a lot of constraints, which limits the duration of our activities, and each individual has its own inclinations and restrictions to be part of a group and spend time with the others. This weighted mean (emerging lifetime) allows us to compute the lifetime  $L_{g_i}$  of group  $g_i$  where all conceivable restrictions are represented behind probabilistic variables  $(\tau_i, \alpha_i)$ . In other terms, this manner of proceeding is a way to inject social features in the group forming consensus.

As an example the following group dynamics shown in figure 2.12 and figure 2.13 were obtained by simulations of a large city by an SSA-Group-based model (675000 people with a population density of 8700 people per  $km^2$ ). The difference between these two simulations is the moving rate of the individuals, which is low (0.1  $m/s$ ) in figure 2.12 and fast (1  $m/s$ ) in figure 2.13.

In figure 2.12, we can see that after about 1h30, the population reaches a kind of equilibrium between individuals in groups and *free* individuals, and that almost 50% of the population is grouped. We also observe that the majority of groups have a small size and that their lifetime behaviour as expected. The groups with big sizes have lifetimes of about 2h (as

supposed in our lifetime distribution law described before). These results highly depend on the chosen parameters such as the average movement speed of the individuals ( $\bar{v}$ ), for example in figure 2.13, where individuals move 10 times faster than in the first simulation. In this case we can observe that the proportion of grouped people in the population becomes much higher (over 90%). It has to be noted that increasing the movement speed causes an oscillatory phenomenon in the proportion of grouped-individuals in opposition to non grouped-individuals.

The simulation of structured populations (presenting a group formation) will allow us to simulate an epidemiological process in a more realistic way than if we considered a spatially homogeneous population. The infectious events will be simulated within groups by a stochastic SIR epidemic procedure (explained in part 2.3).

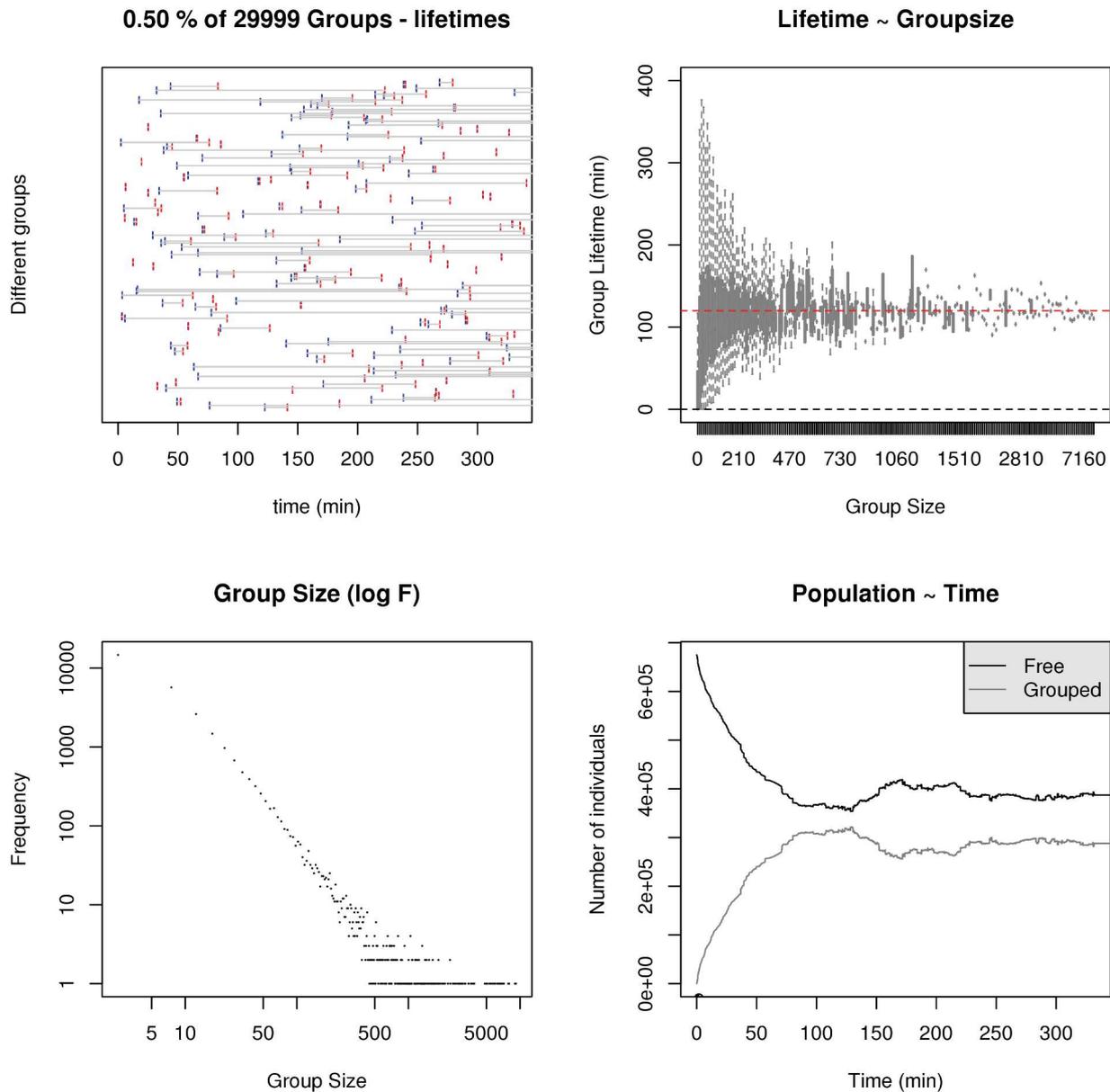
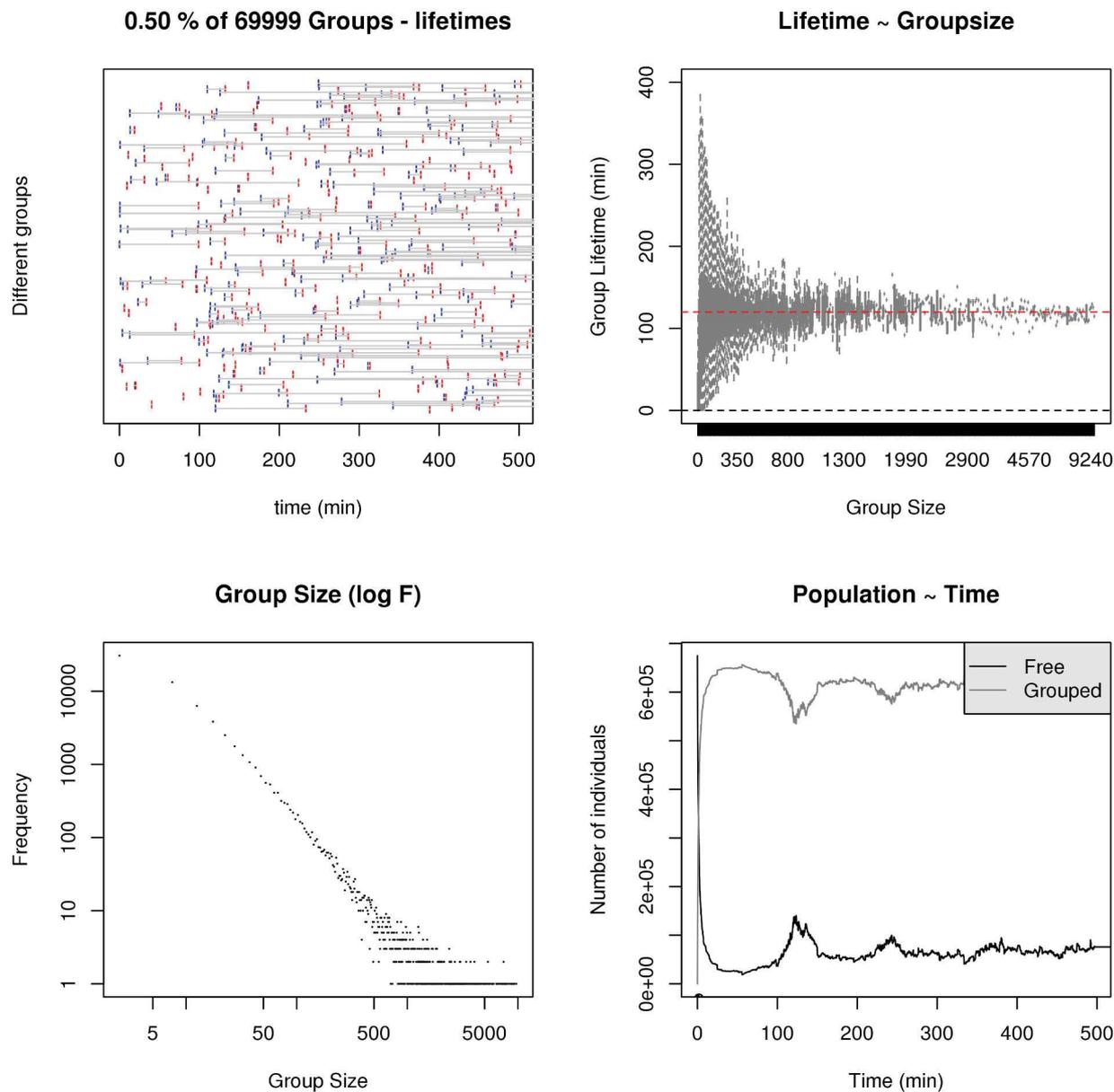


Figure 2.12 **Group Dynamics in a Simulated Big City (Slow Movement Speed)**: Almost 30000 groups are formed in the simulation of a big city (675000 inhabitants, density 8700 Pop /km<sup>2</sup>), during approximately 5 hours and 30 minutes (here just 50% of them and just 150 of groups lifetimes are displayed). Individuals move in average 0.1(m/s) and with an interaction radius of  $d = 0.5(m)$  as the average distance between two individuals for which they are considered in contact. Most of the observed groups have small sizes but some of them (rarely) reach considerably large sizes (here the largest group consists of 9120 people). The lifetime of small groups vary but most of the large group sizes last on average about 120 minutes.



**Figure 2.13 Group Dynamics in a Simulated Big City (Fast Movement Speed):** This simulation is similar to the previous one, but this time the average individual speed is  $1(m/s)$ . We observe that the group-formation rate increases considerably (just 30000 groups are formed during 3 hours 40 minutes). The proportion of the grouped-population is up to 90%. This also generates damped oscillations (with a period of 2 hours) which is clearly visible in the group-formation dynamic and explained by the synchronous formation of groups with sufficiently important sizes.

## 2.3 Computer Implementation of The SIR Models

Numerical simulation is an important part in our epidemiological study. Regardless of the dynamical models presented above, different numerical approaches were implemented to simulate epidemic dynamics or either group dynamics *e.g.* the differential equations of the epidemic process dynamics were numerically simulated either by means of discretised ODE or by homogenous stochastic simulations. The simulation, unlike the analytical computations, are not used to solve the system of equations, but are implemented to calculate the state changes of a system at any time.

There are several types of simulation algorithms that can be used. They are notably used to improve the understanding and at least observe the simulated behaviour of complex systems in different fields of science such as biology, physics and chemistry. For example, biology simulation approaches can be used to study the dynamics of ecological systems.

One of the simulation methods we will present is based on stochastic updates of the states of a system. Daniel T. Gillespie (1967) is known to be the first one that developed a stochastic simulation algorithm (SSA) to describe the dynamics of chemical reactions. Gillespie's algorithm uses a stochastic formulation described by a single differential-difference equation which is called the "master equation". The Gillespie's algorithm is a discrete and stochastic computational method that does not try to numerically solve the master equation for a given system; instead, it is a systematic computer-oriented procedure in which rigorously derived Monte Carlo techniques are employed to numerically simulate the very Markov process that the master equation describes analytically. Then the simulation algorithm is fully equiva-

lent to the master equation, even though the master equation itself is never explicitly used [42]. Since the stochastic temporal evolution of some molecular dynamics in chemically reacting systems seems analogous to population dynamics in epidemiological systems, then Gillespie's SSA can be used in the same way to generate the temporal evolution of a given population without having to deal with complicated analytical approaches [99, 50, 83].

Another manner of simulating and understanding a complex epidemic system is to use individual-based simulation methods. These are bottom-up methods that allow us to model heterogeneous populations (macroscopic level) in which each individual is represented and has different motivations and tendencies (microscopic level). Integrating the individuals' variability in such simulation methods allows the modelling of epidemic systems in which socio-cultural factors can play an essential role. Such computational modelling varies from elementary cellular automata (CA) in which only limited interactions (limited neighbourhood and states) are allowed, to more complex multi-agent systems (MAS) in which all the possible interactions between different agents can be considered.

There are several examples of the use of cellular automata based approaches in epidemiology, notably to simulate the spreading of an epidemic [110], or to study the impacts of population movements and the vaccination on epidemic propagation [102].

Studying an epidemic's dynamics in a given population by considering spatial characteristics through individuals characteristics, *e.g.* their movements, connections and aggregation, is important. In this context agent-based models are the most powerful to capture the disease spread patterns. MAS models were increasingly used in epidemiological contexts

[89, 65, 26, 76, 28].

In the following sections we will present the manner in which we implemented epidemic kinetics to perform numerical simulations. We will first discuss on the local structure and on the global spatialisation of populations in which we will perform epidemiological simulations. We will then describe several algorithms and numerical implementations, commencing with discretised ODE, then SSA based algorithms and finally individual based ones.

### 2.3.1 Different Models for Representing Local and Global Spatiality

In this thesis, we distinguish different numerical epidemic models depending on their level of spatiality. These models are categorised in two different manners, first when the space is structured (this will be called the *global level of spatiality*), and second when the population is structured independently to the manner space is structured itself (this will be called the *local level of spatiality*). Figure 2.14 illustrates the different levels of local and global spatialisation that can be considered and their corresponding numerical models.

Classical ODE models and their implementation are assumed to describe a homogeneous space (we will call them ODE-H). PDE models based on differential equations are partially derived in respect to space. Their numerical implementations use discretised environments, often in the form of spatial grids composed of squared elements (usually called *cells*), in a finite difference numerical scheme. This means that PDE-based models are globally spatial compared to ODE-H that are not spatialised globally (nor locally). As we showed in section 2.2.1.2 where we introduced a modified version of the SIR dynamics that include

the resistance correctly, a kind of population structure is also introduced in the form of *group-complexes*. So, even in the ODE and PDE, a very simple form of local structure can be taken into account, *i.e.* by using our modified SIR kinetics.

Contrary to ODE and PDE that use real numbers to represent the quantities (here,  $S, I, R$  are densities expressed in *individuals.km<sup>-2</sup>*) the stochastic simulations (SSA) allow us to compute the variable quantities in the form of single variations of +1 or -1 elements (*e.g.* individuals), *i.e.* by using integer numbers. In consequence, as we will see, it is easier to introduce a local structure of population in SSA-based models. The model called SSA-H corresponds to stochastic simulation algorithm-based models in homogeneous space (no global, nor local structure, except SIR group-complexes). One of its corresponding spatial versions is the SSA-CA so-called model that is defined as a stochastic simulation in a cellular automaton system. Another one could be a MAS where individuals would never form macroscopic identified entities like groups (a classical MAS system, not presented here). In the SSA-CA, each cell of the grid may contain 0 (empty cell), or 1 to several individuals (here, we implemented it in such a way that the states are only 0 or 1 individual). The contacts exist only in the close neighbourhood (a Moore neighbourhood is chosen, *i.e.* formed by the 8 adjacent cells) and are re-computed each time, so that one individual (in one grid cell) has no preferential relationship with the others in his neighbourhood: there is no notion of group in the SSA-CA numerical model but the observer can identify the emergence of temporary dense regions of individuals (where a disease can spread more rapidly) even if no rules were implemented here to force the individuals to move toward dense regions (a basic social model that could be implemented too). In the SSA-CA, there is no local structure but it is globally

spatialised.

Further, in increasing the local structure of populations, we developed two hybrid models of groups (see section 2.2.2) and SIR dynamics called SSA-G and MAS-G. The SSA-G (SSA-Group-based numerical model) is a globally homogeneous stochastic model having a group-based structure for its population (a local spatiality): there is no notion of positions of individuals nor groups, but groups form in which the SIR processes occur. The spatialised version of this numerical model is the so-called MAS-G. It is a multi-agent system in which each individual is represented in a spatial environment, move and behave independently to the others, but where groups form and become functional entities in which SIR processes occur.

Assuming the relative positions of individuals within groups is important, we could also imagine extensions to these two hybrid numerical models (SSA-G and MAS-G) in which a more precise local spatiality is introduced to describe the positions and movements of individuals within the groups, for example by simulating each group by a cellular automaton (so we would obtain the two corresponding numerical models SSA-CA-G and MAS-CA-G) or by another MAS. This could allow us to take into account the special behaviour of groups, like the group formed by the audience in a movie theatre where everybody is immobile, or, on the contrary, people at a drinks party that permanently mix together.

In this work, we particularly intensify our efforts on the parametrisation of the models (see table 2.1) in such a manner that the parameters have sense and the different numerical models can communicate, *i.e.* their parameters can be scaled from one model to another.

Changing notably (by scaling) the values of some parameters (*e.g.* the individual speeds, or the observation time) allow us to retrieve a similar behaviour of different models running at different levels. As an example, in individual-based models like SSA-CA, increasing the movement speed (or making observations over longer periods, or even simulating larger environments) leads us to simulate the same process by stochastic simulations in a homogeneous space (SSA-H) or even with differential equations (ODE-H). When the individuals move faster, the space becomes homogeneous since everyone can potentially meet everyone (*i.e.* the free collision time is reduced). The same situation happens between differential equation based models in homogeneous space (ODE-H models) and globally spatialised PDE (partial differential equations) based models by changing the diffusion rate.

---

<b>SIR Epidemic Related Parameters</b>	
$K$	Dissociation rate constant ( <i>individual.km<sup>2</sup></i> )
$k_{cat}$	Catalytic rate applied to IS group-complexes infection ( <i>day<sup>-1</sup></i> )
$k_1$	Forward group-complex formation rate constant association ( <i>km<sup>2</sup>.individual<sup>-1</sup>.day<sup>-1</sup></i> )
$k_{-1}$	Reverse group-complex formation rate constant dissociation ( <i>day<sup>-1</sup></i> )
$\beta$	First order Infection rate ( <i>day<sup>-1</sup></i> )
$\beta_{app}$	Effective area for sufficient transmissible contacts ( <i>km<sup>2</sup>.individual<sup>-1</sup>.day<sup>-1</sup></i> )
$\gamma$	first order recovery rate constant ( <i>day<sup>-1</sup></i> )
$1/\gamma$	Mean infectious period ( <i>day</i> )
$\lambda$	First order loss of immunity rate constant ( <i>day<sup>-1</sup></i> )
$1/\lambda$	Mean recovery period ( <i>day</i> )

---

<b>Population Related Global Parameters</b>	
$N$	Total population size ( <i>individual</i> )
$[N]$	Total population density ( <i>individual.km<sup>-2</sup></i> )
$\{[S]_0, [I]_0, [R]_0\}$	Density of S, I, R people at the beginning of the simulation.
$A$	Area in which the SIR dynamics are computed ( <i>km<sup>2</sup></i> )
$D$	Diffusion coefficient ( <i>km<sup>2</sup>.s<sup>-1</sup></i> )
$dt$	Time step ( <i>s</i> )

---

<b>Group and Individual Related (local) Parameters</b>	
$v$	individual speed ( <i>m.s<sup>-1</sup></i> )
$\lambda_e$	Exponential distribution parameter (group size drawings)
$\alpha_\Gamma$	Shape parameter of group lifetime's $\Gamma$ distribution
$\beta_\Gamma$	Rate parameter of group lifetime's $\Gamma$ distribution
$\alpha$	Grouping tendency of individuals $\in [0, 1]$

---

Table 2.1 Parameters used in our different numerical simulations

The globally spatialised models can be used to observe epidemic spreads in more realistic situations that allow us to take into account the aspects of an epidemic process linked to the environment's structure (*e.g.* social groups spatial structures like classes in schools, offices, stores ...) and heterogeneous aspects relative to individuals (*e.g.* degrees of vaccination of individuals *vs* high-risk individuals, propension of individuals to behave socially and form groups). In fact, there are many other types of model that can be used in epidemiological studies to integrate pre-topological information about population structures. One way is to hybridise different types of models ; this is the long term objective of our work. For example, one can study a Chickenpox (varicella) epidemic, or viral gastroenteritis infection dynamics, among Parisian primary students by simulating SIR dynamics in a hybrid model composed of a MAS-G model driven by a scripted social behaviour obtained from pre-topical information (who knows who, who eats there, who is in which classroom ...) about the number of primary schools and students in a given environment (school ...), coupled to an SSA-G numerical model that would describe the dynamics of the exterior of the simulated environment, and the exchanges between the interior and the exterior ensuring the development or the confinement of an epidemic.

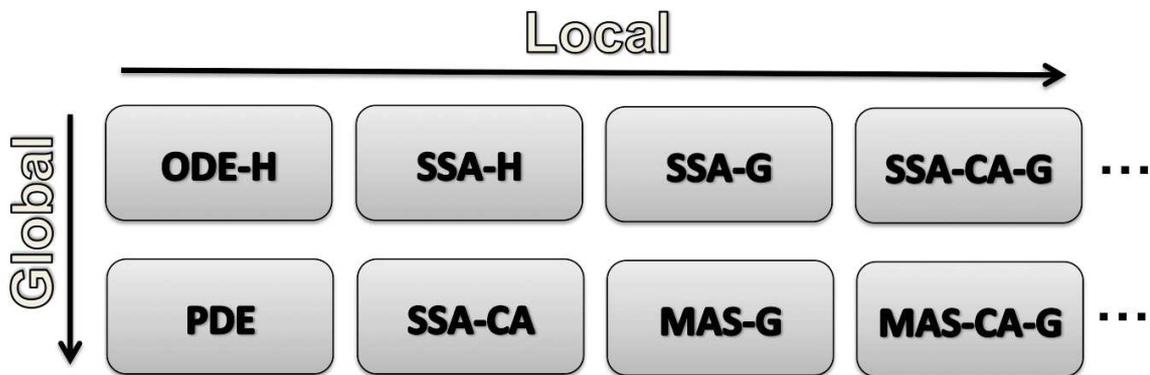


Figure 2.14 **Locally or Globally Spatialised Numerical Models** in which SIR processes can be computed.

All the models we used (some are shown in figure 2.14) can incorporate the different infectious dynamics presented in section 2.2.1 (classical SI, modified SIR without or with resistance) and implement them by using different simulation algorithms (SSA, finite differences ...). Finally, we must admit that we did not examine any real parameter study, the parameters used being mostly unchanged over all the simulations. A sensitivity analysis is now to be done in another study.

We will see now how the SIR dynamics were implemented in the different numerical models presented above.

### 2.3.2 Modified SIR in ODE (Homogeneous Space) and in PDE (Locally Homogeneous)

The numerical evolution of the density vector  $[X](t) = \{[S](t), [I](t), [R](t)\}$  is calculated in discrete time according to a given epidemiological reaction scheme described in section 2.2.1

and in the following system of equations:

$$\frac{d[X]}{dt} = \begin{cases} \frac{d[S]}{dt} = \lambda [R].A - \frac{k_{cat}}{K_M+[S]} [I][S] \\ \frac{d[I]}{dt} = \frac{k_{cat}}{K_M+[S]} [I][S] - \gamma [I].A \\ \frac{d[R]}{dt} = \gamma [I].A - \lambda [R].A \end{cases}$$

The quantities of each population-class ( $[S],[I],[R]$ ) are densities expressed in *people/km<sup>2</sup>* and  $[N] = [S] + [I] + [R]$  is the total population density in an area  $A$ . In the cases where the recovery (resistance) individuals  $R$  affect the number of available infected people  $I$ , the evolution of  $[I](t)$  can be affected by all the temporary group-complexes  $SR,SI,RI$  and  $SIR$  ; then the above modified SIR system is written as follows :

$$\frac{d[X]}{dt} = \begin{cases} \frac{d[S]}{dt} = \lambda [R].A - \frac{k_1 k_{cat}}{(k-1+k_{cat})+(1+\frac{[R]}{K_A}(1+\frac{[S]}{3K_A}+\frac{[S]}{3K_B}))+k_1[S](1+\frac{[R]}{3K_A})} [I][S] \\ \frac{d[I]}{dt} = \frac{k_1 k_{cat}}{(k-1+k_{cat})+(1+\frac{[R]}{K_A}(1+\frac{[S]}{3K_A}+\frac{[S]}{3K_B}))+k_1[S](1+\frac{[R]}{3K_A})} [I][S] - \gamma [I].A \\ \frac{d[R]}{dt} = \gamma [I].A - \lambda [R].A \end{cases}$$

We follow the Euler method with a Runge Kutta 4 approximation to compute the trajectories of  $[S](t)$ ,  $[I](t)$  and  $[R](t)$  in discrete time, with a typical time step of 10 minutes. This time step allows us to compute long periods (months or years) without a huge number of iterations, while giving a sufficient level of precision to the simulation, 10 minutes being a time comparable to the characteristic mean collision time to form groups, a time at which epidemiological reaction start to take sense.

In an ODE based system, the population of  $S$ ,  $I$ , and  $R$  people is uniformly distributed in space. There is no notion of distance between the individuals (*e.g.* the individuals can meet any other individual with the same probability regardless of whether a given individual is in the neighbourhood or not). The PDE-based numerical model is similar to the ODE-based numerical model, the difference coming from the fact that the density vector is a multivariate function  $[X](t, z) = \{[S](t, z), [I](t, z), [R](t, z)\}$  and represents the number of susceptible, infected and recovered individuals at time  $t$  and location  $z \in R^2$  (or  $\in R^3$  if one considers the individuals move in a 3-dimensional space). If  $D$  is the diffusion rate of the individuals (a value, assumed isotropic, as it can be evaluated from the average moving rate of individuals supposed isotropic too), the variation of the density vector is described by :

$$\frac{\delta}{\delta t}[X](t, z) = D \frac{\delta^2}{\delta z^2}[X](t, z)$$

The simulation area consists a 2-dimensional grid of squared cells in which the population  $[X](t, z)$  is supposed to be locally homogeneous. Diffusion is computed by using a finite difference numerical scheme of order 1 (Moore's neighbourhood).

In this manuscript, we only develop the work done on ODE-based numerical simulations and not the one based on PDE that give the same results in terms of global kinetics and that do not increased our knowledge more than other spatialised numerical models did (MAS-G and SSA-CA notably). In section 3.1, we show the particular influence of the dissociation con-

stant of group-complexes, of  $R$  people, and on the population density on epidemic dynamics by using an ODE-based numerical model using our modified SIR dynamics with or without resistance.

### 2.3.3 The SSA Algorithm - Application to SIR and Group Dynamics

The dynamics of the SIR epidemic process can be described by the exact *Gillespie's* simulation algorithm working in continuous time or by its discrete time version (*tau-leaping* version). This algorithm is used to simulate the evolution of a given variable state vector  $X(t)$ , from its initial values  $X_0 = X(0)$ , by generating random pairs of instants and indices,  $(\tau, j)$ , that respectively correspond to times  $\tau$  at which reactive events cause the values of  $X$  to change, and to the number  $j$  of the reactions implied. The vector  $X(t) = (X_1(t), X_2(t) \dots X_{N_c}(t))$  represents the state vector, where  $X_i(t)$ ,  $i = 1 \dots N_c$  classes, is the  $i^{th}$  quantity in the state vector  $X$ , for example, the number of individuals belonging to the  $i^{th}$  class (*e.g.*  $X(t) = (S(t), I(t), R(t))$  and  $X_1(t) = S(t)$ , ..., in an *SIR* model) at time  $t$ . The number of individuals in each class varies due to the occurrence of  $N_p$  available transitions processes (*i.e.* reactions in a kinetic systems). A reaction is a process that causes a change in the numbers of individuals in a given population. Each  $R_j$ ,  $j = 1 \dots N_p$  reaction process is characterised by its propensity function  $a_j(t)$ ,  $j = 1 \dots N_p$  that is the average probability of occurrence of a reaction  $R_j$  in the next  $\tau$  time, *i.e.* in the time interval  $[t, t + \tau]$ .

The state-change vector  $v_j = [v_{1,j}, v_{2,j}, \dots, v_{N_c,j}]$  gives all the instantaneous changes caused by the reaction  $R_j$  over all compartmental classes. The state-change vector  $v_{i,j}$  describes

the change of the  $i^{th}$  class of a population by a reaction  $R_j$ . For example in an *SIRS* model (refer to section 2.2.1.1), the propensity function  $\alpha_j(t)$   $j = 1 \dots 3$  is described respectively by  $a_1(t) = \beta_{app}[S][I]$ ,  $a_2(t) = \gamma[I] \times A$  and  $a_3(t) = \lambda[R] \times A$ .

According to Gillespie's algorithm, the probability that the reaction  $R_j$  occurs exactly at instant  $\tau$  is  $P(\tau, j) = P(\tau) \cdot P(j/\tau)$  where  $P(j/\tau) = P(\tau, j)/P(\tau) = a_j/a_0$  and  $a_0 = \sum_{j=1}^{N_p} a_j$  then  $P(\tau, j) = a_j e^{-a_0 \tau}$  and  $P(\tau) = \sum_{j=1}^{N_p} P(\tau, j) = a_0 e^{-a_0 \tau}$ .

In the continuous time version of the algorithm, an update of the state-change vector is computed at each iteration of the simulation by choosing two uniform random numbers  $r_1$  and  $r_2 \in [0, 1]$  to generate the couple of variables  $(\tau, j)$ . First the instant of the next reaction event is determined by:

$$\tau = \frac{1}{a_0} \ln\left(\frac{1}{r_1}\right) \quad (2.10)$$

Once an event has decided to occur, the  $j^{th}$  reaction ( $R_j$ ) is chosen as the "next" reaction in the infinitesimal time interval,  $[t, t + \tau]$  as the smallest integer  $j$  satisfying the equation below:

$$\sum_{i=1}^j a_i > r_2 a_0 \quad (2.11)$$

Once  $(\tau, j)$  is computed, the population state vector  $X(t)$ , the current time  $t$  of the simulation (and number of the current iteration), but also the state-change vector  $\mathbf{v}$  are updated.[42, 83].

Finally, one can say that the SSA-based model is numerically solving (*i.e.* generating

some possible solutions) a stochastic master equation corresponding to a classical model based on the propensity function ( $a_i$  where  $i \in [1, N_c]$ ) which is the probability of changing the state vector due to the occurrence of one of the  $N_p$  possible reactions (*i.e.* actions resulting in transition of individuals from actual states to other states (reaction) or from actual places to other places (diffusion, moves)):

$$\frac{\delta}{\delta t} P(X, t) = \sum_{j=1}^{N_p} a_j(x - v_j) P(x - v_j | x_0, t_0) - a_j P(x, t | x_0, t_0)$$

Another version of Gillespie's algorithm is known as the  $\tau$ -leaping version and works in discrete time instead of continuous time. In continuous time, only one discrete event occurs at time  $t + \tau$  (*e.g.* a variation of +1 or -1 individual in a population). In spatialised implementations of Gillespie's SSA notably, it is useful to run the simulation in discrete time in such a way, the changes are evaluated at a fixed - small - time step  $dt$ . Then, one must evaluate the number of events that have been produced during this time step  $dt$  in a certain region of space (*e.g.* a cell in a square grid). For each reaction  $R_j$ ,  $j \in [1, p]$ , one computes the associated propensity  $a_j$ . then, the characteristic time at which the reaction  $j$  occurs is  $\tau = 1/a_j$ . The number of reaction events of type  $R_j$  occurring during  $dt$  is obtained by drawing according to a Poisson law of parameter  $\lambda = dt/\tau$ .

Gillespie's simulation algorithm has already been applied to different epidemiological mod-

els, such as an SIR model to study the seasonal influenza epidemic [27] or a temporal SEIR model to evaluate the outbreaks of Ebola Hemorrhagic Fever [73] and a multiple-group model (SIGR) to study vector-borne diseases such as malaria[50].

In the present work, we used Gillespie's algorithm in different contexts. The SSA-H numerical model uses the continuous time version of the algorithm to simulate the evolution of SIR kinetics. In the SSA-G numerical model, its continuous time version is used to simulate the group dynamics while the discrete time version is used to simulate the modified SIR kinetics within the groups. In the SMA-G numerical model, the discrete time version is used to simulate the SIR kinetics within the groups (formed by an emerging process of collision and merging). Finally, the SSA-CA numerical model also uses a particular implementation of the discrete time algorithm for the same purpose (SIR kinetics) and the classical discrete time version to compute the individual moves.

### 2.3.4 Implementation of SIR Dynamics in SSA-Group Based Models

Our SSA-Group numerical model works either in discrete and continuous time in homogeneous space. It has been developed to integrate a group structure to populations in order to simulate SIR dynamics within. The global population of  $N$  individuals is separated into two distinct sub-populations : the *free* population of individuals that are not engaged in any group, and the *in-group* population that is a heterogeneous collection of  $N_g$  different  $g_k$  groups of  $S_{g_k}$  ( $k \in [1, N_g]$ ) individuals (grouped individuals). Groups are characterised by two features : their lifetime and their size, all described by probabilistic laws (see section 2.2.2). It is reasonably assumed that no event of any type (group formation, group death, SIR event ...)

occurs at *exactly* the same time as another one. Then, it is acceptable to modelise the group dynamics as a succession of events; by using an SSA in continuous time to describe new group formation, two groups can never form at exactly the same time, but, depending on the density of the population of free individuals and on the moving rate of individuals, very short times can separate these events. Within a group, the space is considered homogeneous and the discrete time version of Gillespie's algorithm is applied to compute one of the three possible reactions of a given SIR model (we used the modified SIR models with or without resistance). Infection reactions are bimolecular type reactions ( $S + I \rightarrow I + I$ ), so they can only occur in the context of groups on the contrary to mono-molecular reactions (recovery  $I \rightarrow R$  and loss of immunity  $R \rightarrow S$ ) that might occur in both the *free* population or within groups. Figure 2.15 illustrates how several SIR reactions would take place in a given population. We also assumed that the formation of groups is independent to the class to which the free available individuals belong: infected people  $I$  have the same probability of being included in a group as the others ( $S$  and  $R$ ), that's to say there is no discrimination in our model, nor effects of retention due to illness (*i.e.* someone sick stays at home).

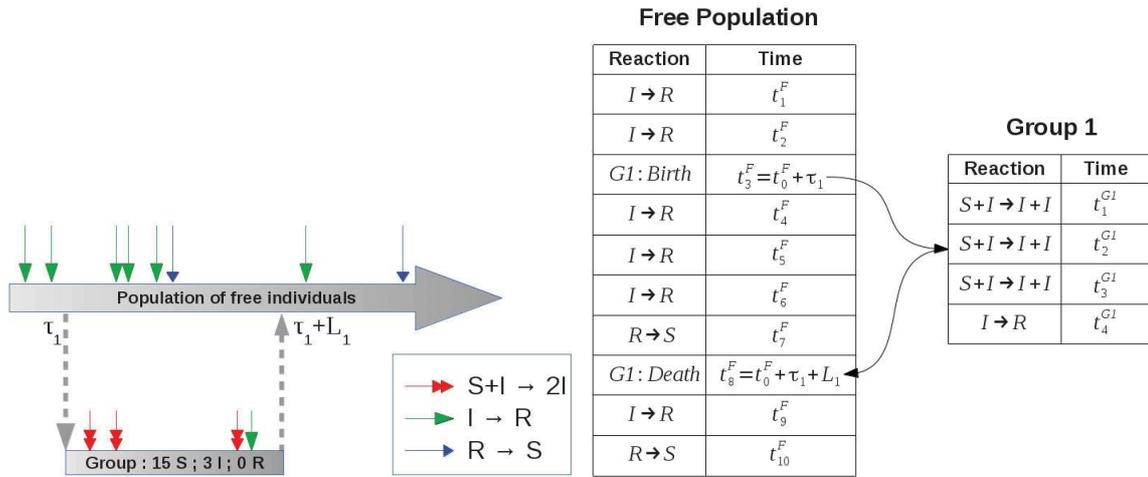


Figure 2.15 **SIR infection process in a group-based SSA model.** (Left) Vertical dotted arrows indicate the formation and the dissociation of a group. The groups forms at time  $\tau_1$  and has the lifetime  $L_1$ . In this example, this group is composed of 15 susceptible and 3 infected people (and no recovered immune people). The 3 different reactions of a classical SIR model are represented by 3 different arrows. In a group-based SSA model, groups form from a collection of independent individuals. In a group, space is considered homogeneous and Gillespie’s algorithm is applied to compute one of the 3 possible reactions of the SIR model. Since infection reactions need contacts, they can only occur in the context of groups. Healing (formation of recovered individuals R from infected subjects) and loss of immunity (formation of susceptible subjects S from R subjects) processes do not need, however, to occur in groups; they can occur within groups or in the main population of independent subjects. (Right) The table shows the respective event lists of both the free population and the first imaginary group-formation.

The reason for which we chose a continuous time SSA for modelling group kinetics and a discrete time SSA for the SIR kinetics have to be explained. If we had proceeded like this (*i.e.* if we had used a continuous time SSA for both group and SIR kinetics) we would have encountered (and we effectively did) a serious problem of time paradox that would have cause bad evaluations of the SIR kinetics.

As discussed before, the time  $t + \tau$  at which the next reaction will occur is computed by using a continuous time Gillespie’s algorithm. Once determined, a chosen reaction causes one change in the population state vector at instant  $t + \tau$ . However, before that, to choose the next

reaction to occur, we need to compute the propensity function which depends on the state vector at time  $t$ . This means that all changes in the *free* and in the *in-group* population must be considered. Then in continuous-time evolution, this causes some intractable implementation issues related to the interactions between events: recovery or loss of immunity events on the one hand, *i.e.* mono-molecular reactions that would occur in one part of the population (in the free population, or on the contrary in the in-group population), and infectious events on the other hand, *i.e.* bimolecular reaction that would occur in the other part of the population (respectively in the in-group population, or on the contrary in the free population). As shown in figure 2.16, the evaluation of the time of the next reaction, given a certain type of reaction (mono-molecular or bimolecular) in one part of the population can depend on the time at which another reaction of the other type (resp. bimolecular or mono-molecular) had occurred. These paradoxical problems in the SSA-G model can easily be solved when the simulation of SIR dynamics is realised in discrete time.

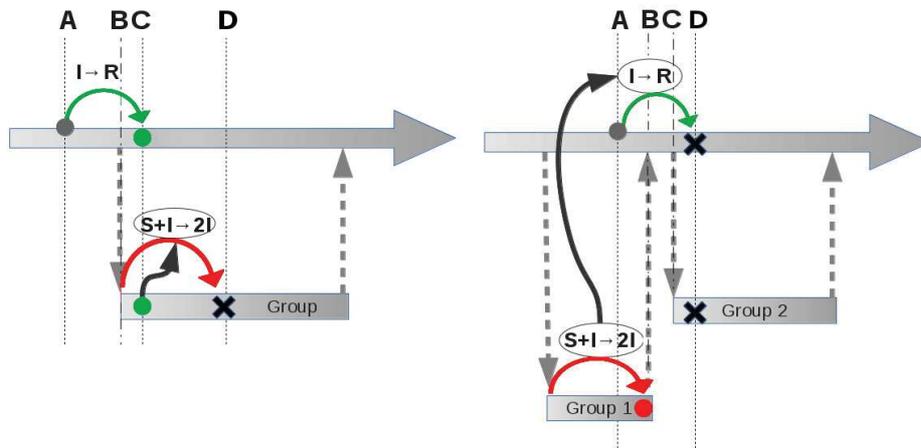


Figure 2.16 **Problems relating to interactions between bimolecular and monomolecular reactions-events in the SSA-G model in continuous time** (Left) The time of the next reaction event in a group (here a bimolecular one in  $D$  identified by the red arrow) depends on the composition of the group at time  $B$ , *i.e.* when it is evaluated. However, if some mono-molecular events (green arrow :  $I \leftarrow R$  or  $R \leftarrow S$  reactions) predicted before in the population (here in  $A$ ) occurs (here in  $C$ , green points) after the time the event in the group is evaluated (here, when the group forms) it can change the composition if they effectively occur in the same group. The event in the group is incorrectly evaluated since it will occur at a probability and time based on a different composition to the one at time  $D$ , *i.e.* when the event is effectively produced. (Right) Similarly, if a reaction occurs in a group (in  $A$ ) after the time a reaction that belongs to the population is evaluated  $A$  and before it occurs  $D$ , the latter will also not be correctly evaluated.

The SSA-G numerical model is implemented as follows (see table 2.2) :

The simulation starts after the initialisation of the state vector at time  $t_0 = 0$  with no groups and a number of  $N$  individuals in the *free* population. The computation starts by generating the instant at which the first group will be formed (birth-time  $B_{g_1}$ ): an exponentially distributed number,  $\tau_{g_1} = \tau_1$ , is randomly chosen to determine the moment  $B_{g_1} = t_0 + \tau_1$  of birth of the first group in the timeline. Its birth-time  $B_{g_1}$  is registered in the event list of the population. As mentioned before, during this time  $[t, t + \tau_{g_i}]$  (here, for the first group  $i = 1$ ), no groups can form (model assumption). Once the group formed, other information about this given group, such as group-size  $S_{g_i}$  and group-lifetime  $L_{g_i}$ , are computed. Each instant  $t$  (*i.e.*

corresponding to both group-birth and group-death events) is registered in a long list of events. It has to be noted that the event list contains all the group birth and death events and constitutes consequently the structure of the population over the timeline. Although the SIR dynamics could be computed afterwards, *i.e.* after the entire population structure over time is determined, we chose to compute the SIR dynamics within groups and the free population as the simulation advances so as to display easily the SIR results (but this is just a pragmatic consideration).

Let us consider that we are at time  $t + \tau^{g_i}$  just after a new group  $g_i$  has formed. Before anything else, we compute the time  $\tau^{g_{i+1}}$  after which the next group  $g_{i+1}$  will form, *i.e.* its birth-time is  $S_{g_{i+1}} = t + \tau^{g_i} + \tau^{g_{i+1}}$ . Once the next group birth-time  $S_{g_{i+1}}$  is known, we know that no other group can form between  $t + \tau^{g_i}$  and  $t + \tau^{g_i} + \tau^{g_{i+1}}$ . However, possible existing groups can disintegrate during this time, thus changing the composition of the free population. Then, we can evaluate the SIR events between  $t + \tau^{g_i}$  and the next group-related event (birth  $S_{g_{i+1}}$  of the next group or death of another group  $g_k$  at time  $t + \tau^{g_k} + L_{g_k}$ ,  $L_{g_k}$  being its lifetime). Different SIR reactions can occur and their implementation depends on whether they belong to the free population, *i.e.* mono-molecular events only, or to the in-group population, *i.e.* both mono-molecular and bimolecular events. First, using Gillespie's continuous time SSA, we evaluate the list of  $I \leftarrow R$  or  $R \leftarrow S$  mono-molecular reactions that occur in the free population during the period  $\tau^{F_i} \in [t + \tau^{g_i}, t + \tau^{g_i} + \tau^{g_{i+1}}] \cap [t + \tau^{g_i}, t + L_{g_k} [$  and we add these events to the event list, each one at its own instant  $\tau^{F, event_j} \in \tau^{F_i}$ . All these event are computed and the free population is updated. Then, within the current group (formed at  $t + \tau^{g_i}$ ), we can also compute the SIR events that may occur before this group

disintegrates (dies) at time  $t + \tau^{g_i} + L_{g_i}$ . This is done by using the discrete time version of the SSA with a time step  $dt$  of 2 seconds (*NB* : we impose a minimal lifetime of 2 seconds to our groups, considering that there's no group context for lower lifetime values).

Figure 2.15 shows an example in which a list of 10 events are to occur at instants  $t_1^F$  to  $t_{10}^F$ . The event #2 is the last event to occur before a group forms at  $\tau_1$  and the event #3, corresponds to the formation of this group  $G1$ . At time  $t_3^F = t_0^F + \tau_1$ . When this group  $G1$  is formed, its size and lifetime are randomly chosen according to dynamics described in section 2.2.2, as well as its SIR composition that is computed at time  $t_3^F$  by a multivariate hypergeometric drawing depending on the group size and on the availability of individuals  $S$ ,  $I$  and  $R$  in the population of free individuals at this time.

Once the group is formed and during its lifetime, the list of SIR events (all reactions including the infection process) are evaluated by an independent simulation procedure (discrete time SSA). In figure 2.15 the three infectious events ( $S + I \rightarrow I + I$ ) and one recovery event ( $I \rightarrow R$ ) respectively occurred at times  $t_1^{G1}$  to  $t_4^{G1}$ . The other transition events take place in the *free* population at times  $t_4^F$  to  $t_7^F$ , before the next group-related event occurs at time  $t_8^F$ . The event at  $t_8^F$  corresponds to the moment when group  $G1$  splits out and each of its  $S$ ,  $I$  and  $R$  individuals return back to the *free population*. The composition of the free population changes. Then other mono molecular events can be computed at times  $t_9^F$  to  $t_{10}^F$  before the composition of the free population changes again (*i.e.* before any group -related event occurs). The table below (Table 2.2) describes the implementation of the SSA-G numerical model for a given population of global size  $N$ :

```

Step(0) Initialisation
-Set SIR kinetic algorithm parameters
-Set Gillespie algorithm parameters
-Set group dynamic parameters (Exponential and Gamma distributions)
-Set simulation start time  $t_0 = 0$ ,  $\#iteration_0 = 0$ 
-Set  $dt$  = fixed time interval and  $t_{max}$  the duration of the simulation
-Initialise the free population SIR vector  $X_0^F = [S_0^F, I_0^F, R_0^F]$ 
-Initialise the list of in-group population SIR vectors
-Initialise the list of in-group population  $(S_g, B_g, L_g)$  parameter vectors
-Initialise an empty event list  $E$ 

Step(1) Compute the first group formation and SIR kinetics
-Calculate and register  $B_{g_1} = t_0 + \tau_{g_1}$  next group  $(g_1)$  birth-time  $\sim Exp$ 
-Calculate and register  $S_{g_1}$  group size  $\sim Exp$ 
-Calculate and register  $L_{g_1}$  group lifetime  $\sim \Gamma$ 
-Register  $B_{g_1}$  and  $B_{g_1} + L_{g_1}$  in the event list  $E$ 
-Evaluate the occurrence of a SIR reaction during  $dt = 2S$  within group  $g_1$ 
-Update the group  $g_1$  SIR state vector
-Increase the in-group time by  $dt$  (iteration  $j$ )
-If  $B_{g_1} + j \times dt < B_{g_1} + L_{g_1}$ 
     $\Rightarrow$  Return to step(1)

2.A - Compute next group formation
-Calculate and register  $B_{g_k} = t + \tau_{g_k}$  next group  $(g_k)$  birth-time  $\sim Exp$ 
-Calculate and register  $S_{g_k}$  group size  $\sim Exp$ 
-Calculate and register  $L_{g_k}$  group lifetime  $\sim \Gamma$ 
-Register  $B_{g_k}$  in the event list  $E$ 

2.B - Compute free population SIR events
-Compute  $t_i^F$  the time of next SIR mono-molecular reaction in free population
     $\Rightarrow$  Jump to Step(2-C)
-Choose next SIR mono-molecular reaction in free population
-Update the free population state vector
-Return to step(2.B)

2.C - Check group events in  $E$ 
-If next event in  $E$  is the death of group  $g_{k-1}$ 
     $\Rightarrow$  Update the free population SIR vector with SIR from group  $g_{k-1}$ 
     $\Rightarrow$  Return to step(2.B)

2.D - Compute in-group population SIR events
-Evaluate the occurrence of a SIR reaction during  $dt = 2S$  within group  $g_k$ 
-Update the group  $g_k$  SIR state vector
-Increase the in-group time by  $dt$  (iteration  $j$ )
-If  $B_{g_i} + j \times dt < B_{g_i} + L_{g_i}$  go back to step(2.D)
-Go to step(2.A)

```

Table 2.2 SSA-G numerical model implementation

### 2.3.5 Implementation of SIR Dynamics in Stochastic Cellular Automata (SSA-CA) Based Models

The SSA-CA numerical model corresponds to the implementation of SIR dynamics by a discrete time stochastic algorithm within a cellular automaton system. The simulation environment is a grid of square cells containing at the most one individual per cell (otherwise the cell is empty). The simulated area consists of  $1 \text{ km}^2$  represented by  $100 \times 100$  cells of  $10\text{m} \times 10\text{m}$  each. This gives a population density of about 10000 inhabitants per  $\text{km}^2$  (if each cell contains one individual).

The individuals can move by jumping stochastically from one cell to another. Their average movement speed is  $36\text{m}$  per iteration time  $dt = 10\text{min} = 360\text{s}$ , meaning  $0.1\text{m}\cdot\text{s}^{-1} \simeq 0.4\text{km}\cdot\text{h}^{-1}$ . Although this moving rate seems a bit slow compared to usual human walking speeds ( $1\text{m}\cdot\text{s}^{-1} = 3.6\text{km}\cdot\text{h}^{-1}$ ), the average movement (walking) speed during a day for an individual might be between  $10^{-3}\text{m}\cdot\text{s}^{-1} \simeq 4 \cdot 10^{-4}\text{km}\cdot\text{h}^{-1}$  and  $1\text{m}\cdot\text{s}^{-1} \simeq 0.4\text{km}\cdot\text{h}^{-1}$ , and largely slower than  $1\text{m}\cdot\text{s}^{-1} = 3.6\text{km}\cdot\text{h}^{-1}$ . It must be remembered that during a day a given individual is often immobile (*e.g.* sitting in front of a computer and working to his Ph.D. thesis, waiting for the copy machine or the printer, spending time in the lift, having a coffee with colleagues around a table, ...). Both move and SIR dynamics are computed as discrete time stochastic dynamics. The time step  $dt = 10\text{min}$  is chosen to be large enough to observe some epidemic events and small enough to allow the individuals to move frequently from one  $10 \times 10\text{m}^2$  cell to another.

The simulation algorithm is divided into two phases, first the displacement of individuals and second the computation of the SIR reactions in each cell. Both computations that change the cell state vector (presence or absence of an individual of nature S, I or R) are updated in parallel, *i.e.* two computational grids containing S, I, R or no individual (empty cell) are used, the one being the *active grid* during the time  $dt$  from which the old states are taken and used in the computation, the other being a *temporary grid* in which the new states (computed from the old ones) are written; and reciprocally at the next time step the two grids permute).

Both displacements and SIR reactions depend on the neighbourhood of cells. The neighbourhood chosen is a Moore's neighbourhood composed of the 8 adjacent cells. Displacements depend on whether empty cells are present in the neighbourhood of the considered cells. If there is at least one empty cell in the neighbourhood of a given cell, then a probabilistic variable ( $\alpha_1$ ) is evaluated to assess the likelihood of movement to that empty place; in the case of more than one empty space being available in the neighbourhood, the movement is more probable and is chosen randomly to one of the empty neighbouring cells. If there is no empty space around a given cell (*i.e.*  $\alpha_1 = 0$ ) then a probabilistic variable  $\alpha_2$  is drawn from a joint probability distribution to let two neighbours swap their cells. In practice the simulation algorithm evaluates in the same time the probability of displacement from one filled cell to an empty cell (if there is one) and the joint probability of swapping the cell with another individual in the neighbourhood.

The SIR epidemic events (reactions) are computed according to a particular implementation of the discrete time SSA (see figure 2.17). Indeed, only one single reaction can

exist that concerns a given individual in its cell: if an individual is  $I$ , the only reaction it is concerned with is the recovery  $I \rightarrow R$  ; if the individual is  $R$ , the only reaction it is concerned with is the loss of immunity  $R \rightarrow S$  ; then, if the individual is  $S$  and if there are  $I$  in its Moore's neighbourhood, the only reaction it is concerned with is an infection  $S_{cell} + I_{neighbourhood} \rightarrow I_{cell} + I_{neighbourhood}$ . In the last case, if there are  $N_I \leq 8$  infected persons in the neighbourhood, evaluating this bi-molecular reaction is equal to evaluating  $N_I$  times a mono-molecular reaction  $S_{cell} \rightarrow I_{cell}$  with a fixed probability.

In consequence, increasing the number of infected neighbours increases the probability that a given susceptible individual is infected, even if there are also recovered (resistant *R e.g.* vaccinated) individuals in the same neighbourhood. Hence the SSA-CA numerical model does not consider the resistance to have an effect on the SIR kinetics themselves : in consequence, we use the equations of the modified SIR kinetics without resistance in this numerical implementation. However resistant individuals play an emergent spatial role by blocking like rocks in an avalanche the spreading of the disease. When a given susceptible individual is surrounded by numerous resistant individuals, the probability of being in contact with an infected individual is considerably reduced. At large scales, percolation effects will be observed depending on the density of resistants and empty spaces (*i.e.* the density of people in the environment).

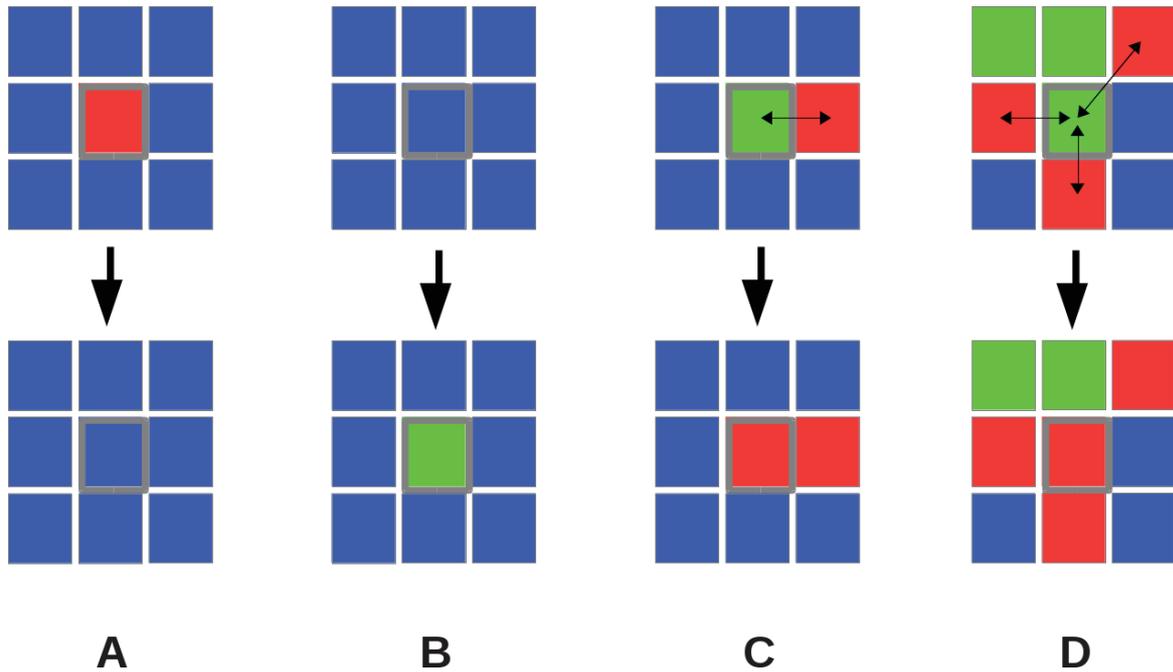


Figure 2.17 **Stochastic Simulation Algorithm based on Cellular Automaton (SSA-CA Model)**. These examples show the behaviour of the cellular automaton system, (here 9 cells). The green, red and blue cells respectively represent individuals (one per cell) in population class  $S$  susceptible,  $I$  infected and  $R$  recovered respectively; (A) Recovery event: mono-molecular reaction (no interaction with neighbours), the central cell turns into blue; (B) Loss of immunity event: mono-molecular reaction (no interaction with neighbours), the central cell becomes green; (C) Infection event: bi-molecular reaction, interaction occurs between a susceptible individual in (central cell) and an infected individual (left cell), the arrow showing the interaction between the two individuals; (D) same type of reaction (infection event), but involving a more substantial infected neighbourhood. We also see that in (C), the other individuals ( $R, S$ ) surrounding the given infected individual can make an isolated environment to prevent propagation to the rest of the population.

### 2.3.6 Multi-Agent System Modelling of Group Structure and SIR Dynamics (MAS-G)

**2.3.6.0.1 Group dynamics in the MAS-G model.** We developed our multi-agent system in the language *NetLogo*<sup>2</sup>. The environment is composed of a two dimensional grid

<sup>2</sup>a programmable multi-agent modelling environment.

of  $200 \times 200$  cells. Each cell of the grid is a square  $16(m^2) = 16 \cdot 10^{-6} km^2$  that contains a NetLogo agent that may corresponds to a single individual or a collection of individuals (*e.g.* a group). The population density of a simulated area containing  $N = 10000$  individual-agents, is equal to  $10^4 / (200 \times 200 \times 16 \cdot 10^{-6}) = 15625$  inhabitants per  $km^2$ . Each individual explores (moves randomly at a given fixed speed, chosen equal to  $1m.s^{-1}$  in this model) the area in its own neighbourhood and can join an existing group if it is sufficiently close to the group, or form a new group with other individuals in the same way. In this MAS-G model we assume that the existing groups can not merge between each others (only single individuals can join existing groups). Once formed, a group obeys the same law as those described for the SSA-G model. Details about group dynamics has been discussed in section 2.2.2. Since a communicable infection occurs during close contacts between individuals who have already formed a group, in our MAS-G model the effective population density actually depends on the number of in-group individuals, a group having its own area computed from the single surfaces occupied by its components (the individuals).

Following is the algorithm we used. First, the algorithm implemented in the MAS-G numerical model is based on two main classes : *individual* and *group*. Each of these classes, with their member attributes and functions, are described in details in figure 2.18.

Individual	Group
<ul style="list-style-type: none"> <li>- xcor(int)</li> <li>- ycor(int)</li> <li>- susceptible?(boolean)</li> <li>- infected?(boolean)</li> <li>- recovered?(boolean)</li> <li>- grouped?(boolean)</li> <li>- grouping-tendency(float)</li> <li>- in-group-with(int set)</li> </ul>	<ul style="list-style-type: none"> <li>- xcor(int)</li> <li>- ycor(int)</li> <li>- group-size(int&gt;2)</li> <li>- group-members(int set)</li> <li>- group-position(int)</li> <li>- group-start-time(int)</li> <li>- group-end-time(int)</li> <li>- group-lifeTime(continuous)</li> <li>- # new-cases(int)</li> </ul>
<ul style="list-style-type: none"> <li>- assign-grouping-tendency()</li> <li>- assign-status()</li> <li>+ move()</li> <li>+ grouping()</li> <li>+ end-infectious-period()</li> <li>+ end-recovery-period()</li> </ul>	<ul style="list-style-type: none"> <li>+ choose-a-S-I-R-reaction()</li> <li>+ ungroup()</li> <li>+ compute-LifeTime()</li> <li>+ export-SIRSpopulationData()</li> <li>+ export-IncidenceData()</li> <li>- export-SIRSgroupsData()</li> </ul>

Figure 2.18 **Individual and Group classes of MAS-G Model:** The algorithm's variables include x-coordinate (y-coordinate) of all individuals and groups, the SIRS compartmental classes, the information about whether individuals are involved in a group (or not), and how likely they are to engage in a group, the set of individuals that are in group with a given individual, number of members in each group, the set of group's members, instants at which a group-agent appears or disappears in simulation, how long a group lasts during simulation and the number new entering individuals in I-class at group level.

At a global level, the operational function *assign-grouping-tendency* attributes a floating number  $\alpha \in [0, 1]$  to each individual-agent. The function is recalled once each time before a group forms. Based on the fact that an individual, depending on its interests, might have a variable tendency to join a group, the computed grouping tendency of a given agent can vary from one group to another group. At the level of individuals, the *assign-status*, assigns individuals to one of the susceptible, infectious or recovered classes of the SIR model. This function is called once at the initial time and is recalled each time one of the SIR transition processes occur during the simulation. The *move* function makes the individuals walk around

randomly. The direction of movement of an individual is chosen randomly, then the next location that it will reach along this direction is computed according to its speed as the distance it will cross from its current location. This function is called only for ungrouped individuals since we assume that the individuals in a group are less likely to move ( $v \approx 0$ ). The *grouping* function consists in two sub-operations, *i.e.* the *form-a-group* and *join-a-group* functions. Each individual is potentially willing to form a group or to join an existing one in its neighbourhood. The neighbourhood includes all individuals whose distances from the concerned individual is less than or equal to a certain number, the proximity number (in our simulations, we chose a radius of  $4m$ ). This leads to the surface area around individuals of  $\pi \cdot 4^2 \approx 50m^2$ . Whenever the mean of grouping-tendencies of ungrouped individuals in a neighbourhood (or of an ungrouped individual with other individuals belonging to an existing group) is greater than 50%, a group forms (resp. a new member joins the existing group). Groups are distinct objects that inherit their characteristics from individuals, notably the group lifetimes that are calculated by *compute-LifeTime* as combinations of individual grouping tendencies and times  $\tau$  drawn from a Gamma distribution as already described, such as  $L = \sum_{i=1}^n \alpha_i \tau_i / \sum_{i=1}^n \alpha_i$ . Of course, when the lifetime of a group is reached, the group disassembles and its individuals are released into the environment by the function *ungroup*. The fact that groups are now identifiable entities (group-agents) distinguishes group-based simulations from individual-based simulations: SIR simulations deal now with a lower number of entities, the group-agents, instead of a greater number of individual-agents.

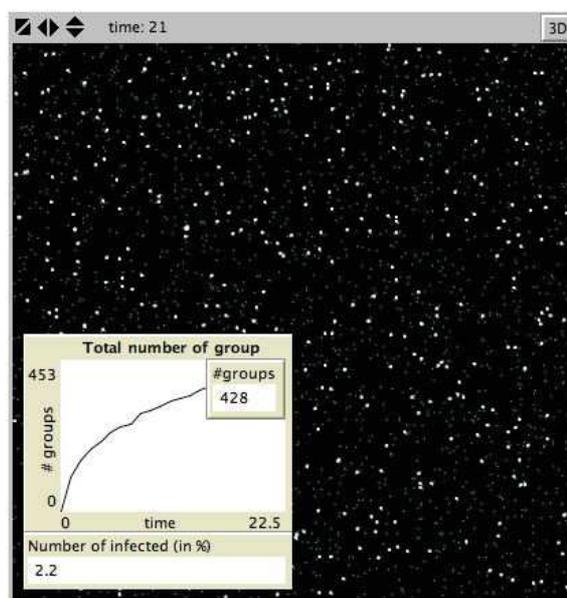


Figure 2.19 **Multi Agent Simulation in Group-based System** This model is an individual-based system conducted in *NetLogo* (Java-based) environment. A random probability number,  $0 \leq \alpha < 1$ , attributed to each agent (individual) as their grouping tendency. Individuals explore in the area of their own neighbourhood; they can form a group and make decision to leave a group (when groups split up).

**2.3.6.0.2 SIR dynamics in the MAS-G model - dynamics.** Since, given the densities and total number of individuals used here, and given the manner in which a group forms in this model (random encounters and mergers), it is not very probable that large groups will form. We mainly observe groups of a size smaller than 10 individuals, in which no sub-groups (*i.e.* group-complexes) can form or play a significant role. Then, we assume that the effect of resistance within the groups have not be taken into account by way of modified SIR dynamics like in the SSA-G model (in which large groups can form). In consequence, the compartmental SIR model used here is the classical SIR model with a temporary immunity, which is appropriate for various communicable diseases viewed at the scale of individuals (spread from person-to-person) e.g.: viral agents like measles, mumps, influenza, smallpox,

cholera, typhoid fever and diarrhoea.

The classical SIR model consists of three compartmental categories: susceptible  $S$ , infectious  $I$  and recovered or temporal immune classes  $R$ . The process is through a directly transmitted infection (DTI) process, meaning that individuals  $S$  might become infected by being in direct contact with another individual in the  $I$  class in such a way  $S + I \xrightarrow{\phi} I + I$ . The force of infection ( $\phi$ ), known also as horizontal incidence [47], is described by the rate of contacts  $c$ , the probability  $p$  that a contact is effectively realised with an infectious individual ( $p \simeq \frac{I}{N}$  known as the prevalence of infection where  $I$  is the number of infected individuals and  $N$  is total number of individuals in the population), and the probability  $v$  that a contact between an  $I$  individual and a  $S$  individual results in a transmission ( $\phi \propto c \times p \times v$  where  $v \propto \beta$  and  $\beta$  is the infection rate) [11]. For the others,  $I$  and  $R$ , the individuals are respectively transformed into  $R$  ( $I \xrightarrow{\gamma} R$ ) when their infectious period is finished, and into  $S$  ( $R \xrightarrow{\lambda} S$ ) when they lose their immunity. The average infectious and recovery periods are determined respectively by  $\frac{1}{\gamma}$  (where  $\gamma$  is the recovery constant rate) and  $\frac{1}{\lambda}$  (where  $\lambda$  is the loss of immunity constant rate) [47]. The amount of change for state vector  $X(t) = [S(t), I(t), R(t)]$  is formulated by the following system of equations:

$$\frac{dX}{dt} = \begin{cases} \frac{dS}{dt} = \lambda R - S\phi = \lambda R - Sc\frac{I}{N}v \\ \frac{dI}{dt} = Sc\frac{I}{N}v - \gamma I \\ \frac{dR}{dt} = \gamma I - \lambda R \end{cases}$$

For a given compartmental model, the characteristic features of a particular DTI disease, such

as the adequate contact number (*e.g.* a couple of people for HIV), disease duration, recovery period, ..., can be translated into model-dependent parameters by changing the values of the transition rates. For example in the SIR model presented here, transmission, recovery and loss of immunity rates are represented by a set of parameters,  $\beta$ ,  $\gamma$  and  $\lambda$  respectively. The recovery and loss of immunity rates shouldn't be influenced by group features described above such as group size or lifetime: recovery and loss of immunity transition terms for each individual depend only on disease features regardless the fact that a given individual belongs to a group or not. However, the transmission terms depend on the rate of contacts, which increases directly with the population density ( $c \propto D = \frac{N}{A}$  where  $A$  is the total surface area occupied by the population) ; it is known as a density-dependent transmission process [11, 91]. The changes in number of  $I(t)$  are formulated by  $\frac{dI}{dt} = \beta S \frac{I}{A} - \gamma I$ . At the population level, the global area  $A$  is assumed constant, then  $\beta$  is equal to  $\beta^*$  of the mass action transition equation  $\frac{dI}{dt} = \beta^* SI - \gamma I$  [11] and the effect of density can be neglected in such a model (*NB.* remember that we can , however, express the importance of density by considering the effective density in which the individuals live; see appendix B). In density-dependent transmission,  $\beta$  expressed in  $area.individuals^{-1}.time^{-1}$  (*e.g.*  $km^2.individuals^{-1}.h^{-1}$ ) is defined as the effective area over which a S-individual makes effective infectious contacts per unit time. At the group level, since the group size and its induced occupied surface area is often reduced, the density has a non-negligible impact on the effective number of contacts that occur within a group. The individuals in more densely populated groups are more likely to make more intra-group contacts, increasing the probability of infection. Then, at the group level, the changes of the state vector for a group  $G_i$  is  $X_{G_i}(t) = (S_{G_i}(t), I_{G_i}(t), R_{G_i}(t))$

is investigated through the following system of differential equations:

$$\frac{dX_{G_i}}{dt} = \begin{cases} \frac{dS_{G_i}}{dt} = \lambda R_{G_i} - \beta S_{G_i} \frac{I_{G_i}}{A_{G_i}} \\ \frac{dI_{G_i}}{dt} = \beta S_{G_i} \frac{I_{G_i}}{A_{G_i}} - \gamma I_{G_i} \\ \frac{dR_{G_i}}{dt} = \gamma I_{G_i} - \lambda R_{G_i} \end{cases}$$

where  $A_{G_i}$  represents the surface occupied by the group  $G_i$ . The changes of the state vector  $X_G(t)$  for in-group people observed at the population level is determined by the sum of all the changes within the groups,  $\frac{dX_G}{dt} = \sum_{i=1}^{n_G} \frac{dX_{G_i}}{dt} = [\sum_{i=1}^{n_G} \frac{dS_{G_i}}{dt}, \sum_{i=1}^{n_G} \frac{dI_{G_i}}{dt}, \sum_{i=1}^{n_G} \frac{dR_{G_i}}{dt}]$ , where  $n_G$  is the number of groups existing at time  $t$  in a given population. There are also changes in the population state vector that occur among the individuals not engaged in any group, isolated people designated before as the free population. These changes happen when the  $I$  or  $R$  individuals are transformed into  $R$  and  $S$  individuals respectively after a recovery (resp. loss of immunity) transition process. Then the vector of state changes that concern only the free population (independent to any group-related process) is formulated by :

$$\frac{dX_F}{dt} = \begin{cases} \frac{dS_F}{dt} = \lambda R_F \\ \frac{dI_F}{dt} = -\gamma I_F \\ \frac{dR_F}{dt} = \gamma I_F - \lambda R_F \end{cases}$$

Then the whole change in the population state vector is calculated by summing all changes from the groups and from the free population :

$$\frac{dX}{dt} = \sum_{i=1}^n \frac{dX_{G_i}}{dt} + \frac{dX_F}{dt}$$

**2.3.6.0.3 SIR dynamics in the MAS-G model - algorithm.** Within a group, a Gillespie's SSA is implemented by the function *choose-an-S-I-R-reaction* to compute the list of SIR epidemic events that will occur in the group. The function is recalled each time a new individual becomes a member of the group and changes its composition, *i.e.* so the propensities are updated and the list of SIR events too. The *export-SIRSpopulationData*, *export-SIRSgroupsData* and *export-SIRSgroupsData* functions are called to record the population dynamics, anytime a change occurs in the SIR individual classes of a group.

In populations composed of both groups and individuals (free population), recovery and loss of immunity transition processes can occur among the free individuals or within groups. In the last case, these reactions appear in the list of SIR events evoked above. In the first case, an independent stochastic algorithm is used to compute the transitions from *I* to *R* ( $I \rightarrow R$ ) or from *R* to *S* ( $R \rightarrow S$ ). These two transition processes correspond to exponentially distributed waiting times in a given class, *e.g.* for a transition from *I* to *R*, the change  $\gamma I$  corresponds to  $p(t) = 1/\gamma e^{-\gamma t}$  distribution, describing the fraction that is still in the infectious class a time *t* after having entered into this class [47].

The table 2.3 describes the SSA-based SIR implementation in a given group *k*.

<p>Step(0) Initialisation</p> <ul style="list-style-type: none"> <li>-Input parameters <math>\beta, \gamma, \lambda</math></li> <li>-Input state-change vector <math>v_1 = [-1, 1, 0], v_2 = [0, -1, 1], v_3 = [1, 0, -1]</math></li> <li>-Input initial state vector : <math>X_{G_k}^0 = [S_{G_k}^0, I_{G_k}^0, R_{G_k}^0]</math></li> <li>-Set <math>t = \text{group start time}, \# \text{ iteration} = 0</math></li> </ul> <p>Step(1) Compute Propensity Function</p> <ul style="list-style-type: none"> <li>-Calculate <math>a_j(t) = [\beta S_{G_k} \frac{I_{G_k}}{A_{G_k}}, \gamma I_{G_k}, \lambda R_{G_k}] j \in [1 : 3]</math></li> <li>-Calculate <math>a_0 = \sum_{j=1}^3 a_j</math></li> </ul> <p>Step(2) Generate <math>(\tau, j)</math></p> <ul style="list-style-type: none"> <li>-Pick <math>r_1, r_2 \in [0, 1]</math></li> <li>-Compute <math>\tau = 1/a_0 \ln(1/r_1)</math></li> <li>-Choose the smallest integer <math>j</math> satisfying <math>j = \sum_{i=1}^3 a_i &gt; r_2 a_0</math></li> </ul> <p>Step(3) Update</p> <ul style="list-style-type: none"> <li>-Update the state vector <math>X_{G_k} \rightarrow X_{G_k}(t) + v_j</math></li> <li>-Update time <math>t \rightarrow t + \tau</math></li> <li>-Go to step(1)</li> </ul>
--

Table 2.3 Gillespie algorithm implementation for a given group  $k$ 

Following is a schematic diagram that illustrates the simulation path of the grouping and SIR dynamics in a MAS-G numerical model.

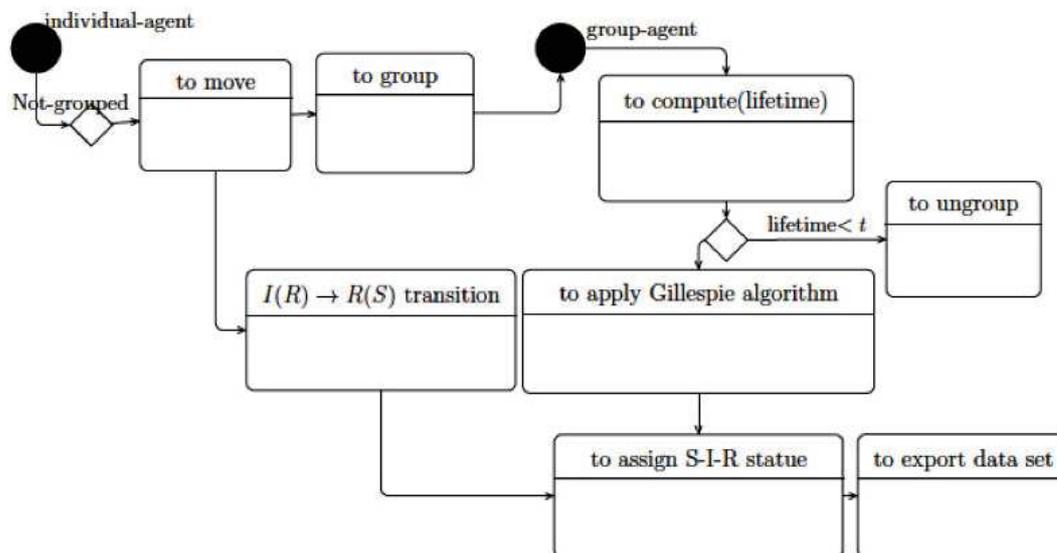


Figure 2.20 **Group-Forming and SIRS-epidemic Simulation Diagram**

To study the influence of groups on the epidemic dynamics of a communicable disease, we also elaborated an MAS numerical model, which was also based on SSA-based SIR dynamics, but that does not allow group formation. This time, infection reactions are computed each time  $I$  and  $S$  individuals are in the same neighbourhood. This allows us to compare two SIR dynamics, with or without group dynamics. In both cases, populations (with or without groups) are considered homogeneous, but differently. At the group level, individuals can be considered as homogeneous populations. The role of groups here is to provide dense regions composed of numerous individuals, *i.e.* to increase the probabilities of infectious contacts.

It is also important to remember that groups form locally from individuals having a certain moving rate. One can expect that spatial heterogeneities in the SIR distribution can appear when the moving rate of individuals is sufficiently limited, so that they don't mix together

sufficiently. Then the MAS-G model should converge with an SSA-CA numerical model in which the moving rates are limited. On the contrary, when their rates are important (*e.g.* infinite), the dynamics obtained in a MAS-G would not correspond to a simulation using a homogenous SSA-based numerical model, nor an EDO based one with classical SIR dynamics, because groups would nevertheless play a role.

## 2.4 Measurements

### 2.4.1 Global Measurements

First, in epidemiological studies at the macroscopic level (*i.e.* at the population level such as the population of a country, a region or a big city), one can observe the number of changes in the number of  $S$ ,  $I$  and  $R$  individuals. In the context of public health surveillance, these measures are mainly limited to the estimated number of infected people at the population level: the number of people infected by a specific disease (*e.g.* Flu, gastroenteritis ...) are declared by local monitoring healthcare personals (*e.g.* generalist doctors, responsible persons in the hospitals or analysis laboratories ...). Then the information is treated by statistical methods (*e.g.* interpolation technics to correct missing data) to provide a robust estimation of this value (number of infected people). However, technically speaking, the estimation of  $I$  individuals or of other global information like epidemic thresholds (*e.g.*  $R_0$ ) don't give access to many types of information, such as the proportion of resistant  $R$  and susceptible  $S$  individuals (*NB.* and of course of the different degrees of immunity and susceptibility against multiple diseases and the variations) in a given population. Access to this information may

allow us to extract a real knowledge of an epidemic process at a global scale, especially for those in which the individuals can be highly immunised naturally or by vaccination.

Monitoring the evolution of the maximum infection speed  $V_{max}$ , of the Michaelis constant  $K_M$ , of the  $\beta_{app}$ , and of the  $R_0$ , helps to understand the different effects of  $R$  individuals on the epidemic process since all these variables depend on non-competitive inhibition due to resistant people  $R$  (see section 2.2.1.2). The reproduction number  $R_0$ , usually thought constant, is defined as the product of infection rate (which depends on the resistance) and on the average infection period  $1/\gamma$  such that  $R_0 = \beta_{app}[S] \times 1/\gamma$ . The reproduction number  $R_0$  reaches a maximum if  $\beta_{app}$  is maximum; this happens when the infection starts to develop and reaches its maximum speed, *i.e.* when there is no resistance yet (assuming there are also no vaccinated people). Its maximal level can be compared to  $R_0$  that is known as the *basic reproduction number*. This value is however not constant and starts to decrease when  $R$  increases. The strength of the variations of  $R_0$  is strongly linked to the density of  $R$  but also to the parameters that control group dynamics (*e.g.* the group association/dissociation constants  $k_1$  and  $k_{-1}$  relative to the catalytic constant  $k_{cat}$ , or in other terms to the  $K_M$ ). Figure 3.2 in the next chapter (section 3.1 of chapter 3), a series of SIR kinetic plots obtained by 4 different ODE-based simulations using the modified SIR dynamics, illustrates the impact of  $R$  individuals on these different variables that could be partially measured. The first row corresponds to a model using modified SIR dynamics without resistance as described in section 2.2.1.1. The second, third and last rows correspond to simulations using modified SIR dynamics with resistance (see section 2.2.1.2) in different conditions. These SIR kinetic

plots show i) the variations of  $S$ ,  $I$  and  $R$  individuals, ii) the evolution of  $\beta_{app}$  and the  $R_0$ , and finally iii) the  $K_M$  and the  $V_{max}$  as functions of during  $t = 100$  simulated days. The effect of the conjugated effect of the resistance  $R$  and the formation of group-complexes (inhibited complexes) is very clear (second and third rows) when compared with the kinetics without resistance (first row) or when group-complexes can't form easily because of an important dissociation (last row). The effect is stronger when  $R$  are present from the beginning. The consequences of taking into account the resistance and group dynamics are particularly obvious when comparing the values of  $R_0$  and  $\beta_{app}$  in their corresponding plots.

### 2.4.2 Epidemic outbreaks and criticality analysis

The observal of random fluctuations in epidemic outbreaks asks the question of the possible use, in such epidemiological studies, of concepts from critical systems, notably the self-organised criticality phenomena and their associated tools. Criticality is a characteristic of macroscopic systems at thermodynamic equilibrium that, when shifted away from this equilibrium by regular amounts of energy or matter, relax by abrupt variations of their potential, *i.e.* towards their return to the thermodynamic equilibrium. They are composed of numerous elements in interaction that sometimes behave together (during the relaxation) as a whole and in this way macroscopic behaviours appear. In many physical systems, the distribution of size and the occurrence of such variations follow power-laws (exponential). Self-Organised Criticality (SOC), known for a long time by physicists, appears in several natural phenomena like earthquakes, avalanches, biological extinctions, phenomena with natural fractal structures. The famous sandpile model introduced by Bak et al. is a simple example which describes

the criticality mechanism and shows how most of the variabilities we observe all around us follow the concept of criticality [9, 62]. In living systems (in the large sense), criticality has been observed in economical studies and also in some social-political mechanisms, *e.g.* S. Galam [32, 33] has shown how criticality and related concepts may apply to understand the dynamics of some social issues in our complex psycho-sociological life. This concept also captured epidemiologists' attention because they give a multi-scale view of epidemiological phenomena that is not allowed in usual mean field approaches [29, 40, 56, 85, 99, 55].

There has been several attempts to interpret the distributions of epidemic sizes and their durations, and to show how much these phenomena composed of numerous isolated events (which we will call micro-events) follow power law distributions that are signatures of a system (in our case a disease and its spreading) at a macroscopic level [86, 58, 59]. The behaviour of epidemic processes can be studied like catastrophic systems (earthquakes, avalanches ...). For example, according to a study carried out by Rhodes et al.(1997) [85], the outbreaks of measles epidemics in the Faroe Islands, observed for 100 years, show critical dynamics as it can be observed in natural phenomena (see Figure 2.21). The probability distribution of the monthly incidence of a measles virus infection (an epidemic event size  $s$ ) in the Faroe Islands obey a sort of scaling distribution; here, a power law. A straight line of slope  $-\tau + 2$  (where  $\tau$  is the scaling exponent) fits this behaviour. As the slope is negative, the large epidemic events are less likely to occur than the small events and we can estimate

this probability with the equation below.

$$p(s) \equiv \text{prob}(\text{Epidemic.Size} \geq s) = \frac{\sum_{s'=s}^{\infty} s'}{\sum_{s'=1}^{\infty} s'} \propto s^{(\tau-2)}$$

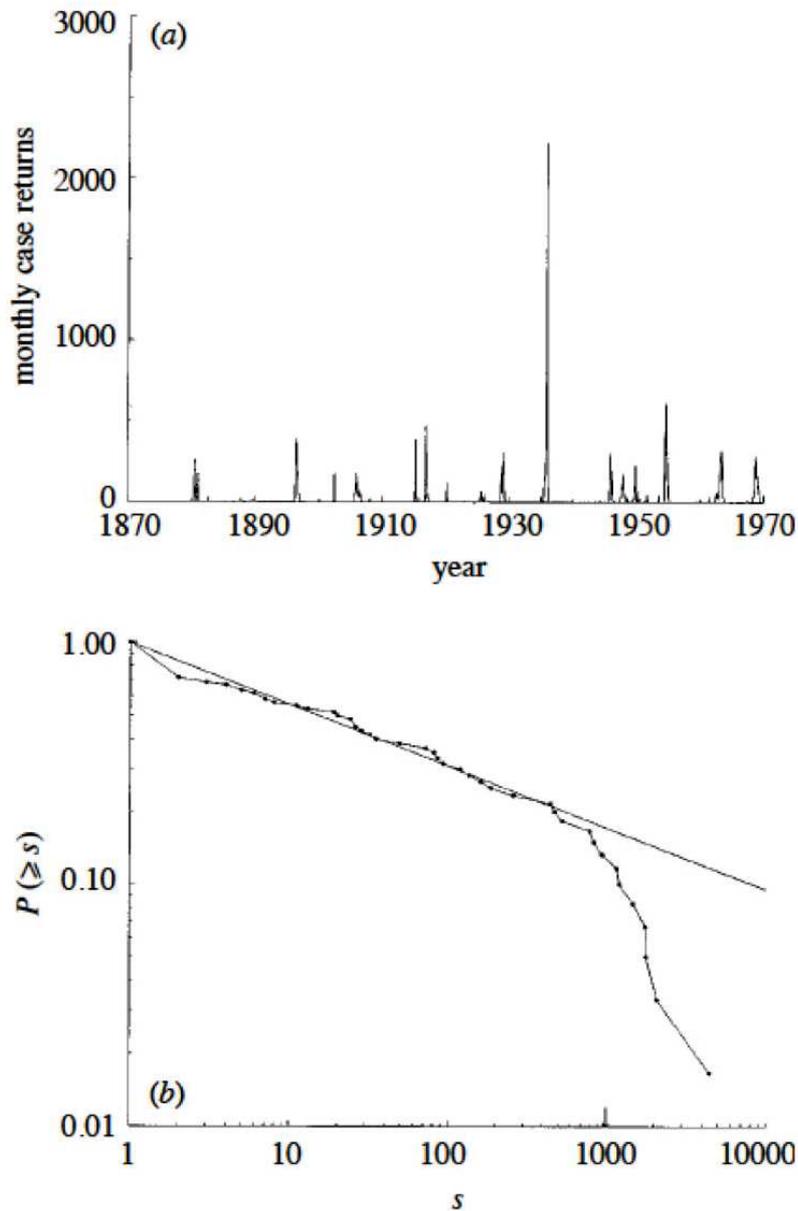


Figure 2.21 **Critical Dynamics:** (a) The monthly measles incidence (epidemic events) in the Faroe Islands (population ca. 25 000) during a time span of 100 years (b) Epidemic size probability distribution for the Faroe Island measles data (log-log plot). The best fit line with a slope  $-\tau + 2$  estimates  $\tau = 2.265 \pm 0.014$  (95 % confidence interval) see Rhodes et al. [85] for more details.

As the models based on homogeneous SSA, MAS or CA (deterministic or stochastic), *i.e.* models in which the notion of individual is present, give fine-grain information about the

number of  $S$ ,  $I$  or  $R$  individuals, this allows us to define two types of event associated to each class of individual, the *catastrophe* and the *rescue* events. For example, if we consider the  $I$  individuals, *i.e.* from the infection point of view, the number of new infected cases during a  $\delta t$  period of time before observing any new recovered (rescued) individual among the infected population is considered to be *one rescue event* (it is a rescue for infected people since their number grows) and the number of new recovered  $R$  individuals during a continuous  $\delta t'$  period of time is *one catastrophe event* ( $I$  individuals are lost). On the contrary, if we consider now the  $R$  individuals, *i.e.* from the recovery point of view, the number of newly recovered cases during a  $\delta t$  period of time before observing any new loss of immunity events is considered to be *one rescue event* (it is a rescue for recovered people since their number grows) and the number of new recovered  $S$  individuals (loss of immunity) during a continuous  $\delta t'$  period of time is *one catastrophe event* ( $R$  individuals are lost and transformed into  $S$ ). This is the same for  $S$  people, the catastrophes of the one, being the rescues of the other. In the simulations, the periods of time during which there is a continuous variation of new cases are not necessarily equal, on the contrary, that is only observed when the periods are fixed, (*e.g.* during weekly observations. The epidemic duration distribution behaves the same way to probability distribution of epidemic size: the probability of observing an epidemic event with a long duration is less than that of events with short durations. This suggests that we should introduce a measure that could cover both the epidemic size and its duration at the same time. To do that, we introduce the *epidemic charges*: infection charge, recovery charge and health charge. As we will see, both our individual-based and homogeneous SSA-based models provide fine-grain information (a microscopical resolution of time, space and SIR

population) and allow us to compute the epidemic charges that could be used to understand and identify epidemics. This could allow us to extract the essence of real epidemiological systems and provide a new method to identify a characteristic signature of different types of epidemic processes and of the factors that affect their dynamics. These signatures are identified from the probability density distributions of macro-events (temporal contiguous microscopic events, *i.e.* catastrophes and rescues). In the simulations, they might depend on model parameters such as infection or recovery rates, vaccination strategies and spatial or social-cultural factors. This might be the same for real epidemiological systems.

### 2.4.3 Micro and Macro Events – Fine to Coarse-Grain Observations

To clarify the concept of *micro* and *macro* events, let us consider an epidemic process in which there is only one type of infection that is transmitted by an initial infected individual in a given population. The graph below ( 2.22) illustrates this process. Infectious contacts occur either directly by infecting agents (in red) or indirectly by healthy agents becoming infected (in blue). The numbers (labels) associated to the agents indicate the new generation becoming infected by previously infected individuals during a time interval of  $d$  days. The transmission of an infection to a new person is an epidemic event described at the *microscopic level*. However, observing this population as a whole over a given period of time  $T$  allows for an understanding of the process at the *macroscopic level* : this process that is continuous and concentrated in time, constitutes an epidemic event viewed at the macroscopic level, *i.e.* it's a *macro-event*. This kind of interaction between susceptible and infected individuals (by a communicable infection) happens in all epidemic processes, in whatever manner they are

observed, whatever models are used. However, depending on the model (*e.g.* the ODE and PDE models), or in the manner that the real process is observed (*e.g.* when observations are yielded once a week) these events are hidden, since one considers only the population level as a global macro-event and individual levels are ignored. In meta-population models where sub-populations (groups) are considered as micro-environments where different macro-event epidemic processes occur, one can choose to observe the changes at a global scale. However, they also provide *micro-event* related information that we think is very useful to fully understand an epidemic process. For example, a simulation of an SIR model by SSA methods in homogeneous space results in small stochastic variations of each population class of  $S$ ,  $I$  and  $R$ . Each variation (+1 or  $-1$  individual) constitutes a *micro-event*. Depending on the population class these micro-events can be designed specifically as  $S$ -,  $I$ - or  $R$ -micro-events, and a contiguous series of related micro-events is a  $S$ -,  $I$ - or  $R$ -macro-event respectively. An example of the variations of infected people obtained by a homogenous space SSA simulation is given in Figure 2.24 (Top). The epidemic  $I$  rescue macro-events are associated with a contiguous increase in the number of  $I$ , while  $I$  catastrophic macro-events are associated to a contiguous decrease in  $I$ .

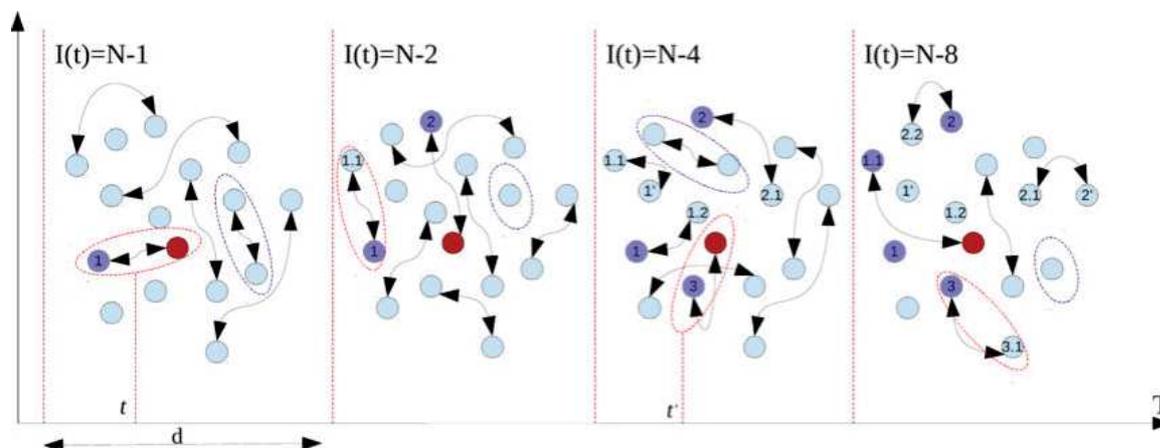


Figure 2.22 **Micro & Macro Events at the Individual Level** : Infectious contacts occur either directly by the infecting agent at departure (in red) or indirectly by healthy agent becoming infected (blue) in the following steps. The numbers that appear to label agents, indicate the new generation becoming infected by previously infected individuals during time interval  $d$ . Transmission of an infection to a new person is an epidemic event at micro level, however observing this population all over a given period of time  $T$  (here  $T = 4d$ ), represents an epidemic event at the macro level.

The sum of absolute values of these variations  $\varepsilon_{\mu_i^t}$ , ( $i > 0$ ) in each class, during a given time interval of  $\Delta t$  months, days, or hours, represent the *macro-events* that have different sizes (in total number of individuals)  $S_M^{\Delta t}$ .

$$S_M^{\Delta t} = \sum_i \text{Micro}_{Events}^i(t \in [t_0, t_0 + \Delta t]) = \sum_i \varepsilon_{\mu_i^t}$$

This is the sum of all  $i$ -occurred *micro-event* at instant  $t$  during a  $\Delta t$  interval,  $t \in [t_0, t_0 + \Delta]$ .

The sign of the macro-event indicates a negative or a positive flow toward or from a given class, *i.e.* a positive flow of  $I$ -micro-events is an infection rescue, while a negative flow of  $I$ -micro-events is an infection catastrophe. Figure 2.23 illustrates the flows associated to each class (S, I, R) in a SIRS epidemic model during time  $T$ .

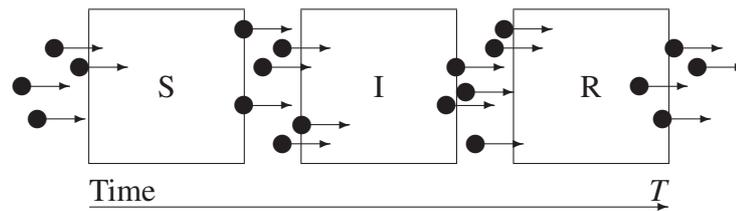


Figure 2.23 **SIRS Micro & Macro Events:** The graphical pattern of flows for each class of the population (S, I, R) in a SIRS epidemic model during time  $T$ .  $I$ -micro-event for positive flows entering the class  $I$  are infectious flows and the ones that are leaving as negatives flows are rescued  $I$ -micro-event flows

The global negative flows of each class as well as the positive flows exert a feedforward effect on an epidemic process. If a given macro-event, occurring during the period  $T$ , consists of a set of recovered- $I$ -micro-events (formation of  $R$  from  $I$ ), more than infected- $I$ -micro-events (formation of new  $I$ ), this will lead (in the future) to a reduction of the effective number of adequate available contacts for the transmission of an infection between susceptible and infected individual. Consequently it will have an impact on the size of an outbreak during or after the period  $T$ . This is to say that the concepts of rescue and catastrophe events may be debatable since the macro-events are in fact not independent, for example the healing process naturally depends on disease recovery rate (that depends itself on heterogeneous individual immune systems). Figure 2.24 (Top) shows a time series of the number of infected individuals ( $I(t)$ ), simulated by a SSA-based model of classical SIR dynamics in homogeneous space. Each elementary variation of  $I(t)$  (+1 or -1 individual) at each instant represents a micro-event. A set of contiguous micro-events of the same type (increase of  $I$  or on the contrary decrease of  $I$ ) constitutes a macro-event. They are shown in Figure 2.24 (Bottom) and correspond to the macro-event that occurred between the iterations (time) 0 and 150, *i.e.* the region in grey in Figure 2.24 (Top). Rescues are shown in red while catastrophes

are in blue. In this example, there are many important rescues (important formation of new  $I$ ) while the catastrophes (recovery events) are less important (few new  $R$  form).

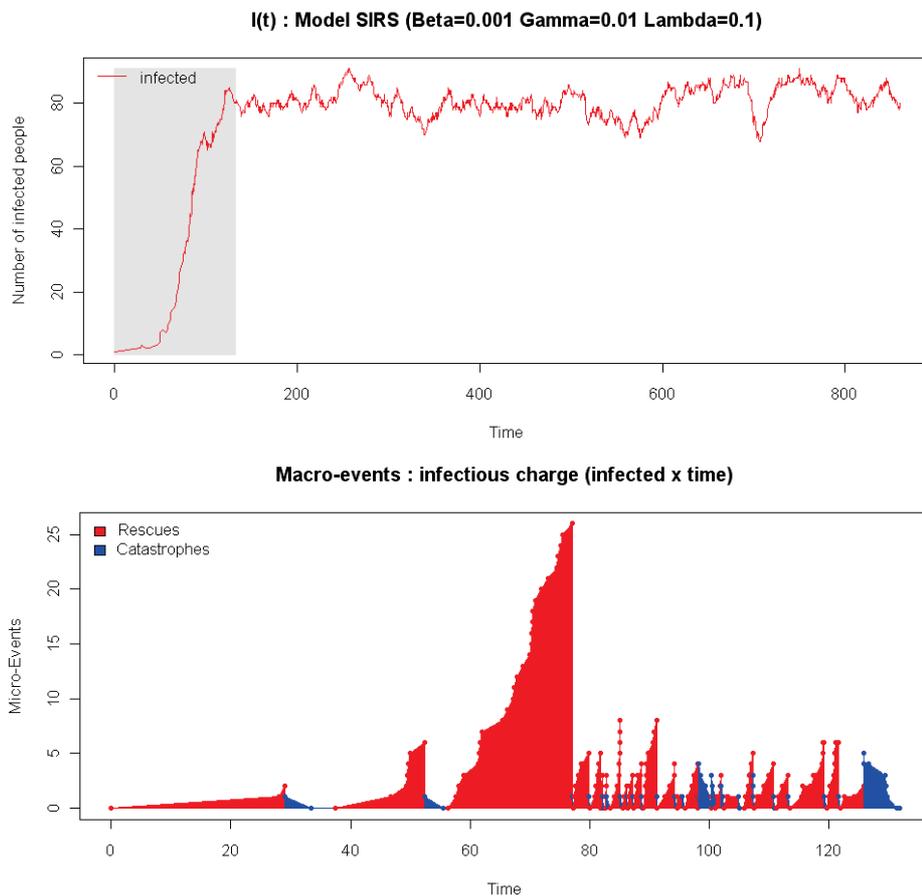


Figure 2.24 **Stochastic SIR model: micro and macro events.** (Top) Time series of infected people  $I$  in a homogeneous SSA simulation show unit variations ( $+1$  or  $-1$  infected subject) that we call *micro-events* (Bottom) Contiguous variations of the same sign form *macro-events*. Catastrophes (negative macro variations) and rescues (positive ones) correspond to the region filled in grey on the time series. In this simulation  $N = 10000$ ,  $I_{int} = 1$ ,  $R_{int} = 0$ .

The distribution of event sizes  $S_M^d$ , in terms of numbers of individuals, can be studied as for critical phenomena (discussed in 2.4.2) then for observed *micro-events* with size  $s_M$  ( $s_M = +1$  or  $s_M = -1$  when computed by the exact SSA) during an epidemic process :

$$p(s_M) \equiv \text{prob}(S_M \geq s_M) \propto c \times s_M^{-\alpha}$$

where  $\alpha$  is the scaling parameter (exponent parameter) and  $c$  is a normalisation constant.

Here, the idea is not to estimate this exponent  $\alpha$  so as to fit the power law distribution with a straight regression line. We are more interested by the notion of criticality as it has been studied in natural phenomena to measure the size and intensity of a given catastrophic event. For an example in seismic catastrophic events, the frequency, intensity (size), type and distribution of earthquakes during a given period of time and in a given area, are referred to as seismicity. As shown in [9] an earthquake is considered to be a self-organised critical phenomenon. The earthquakes of large magnitude have a lower likelihood of occurring than imperceptible seismic tremors, which can occur almost continuously during a given time. In terms of liberated energy, a sum of numerous small tremors can be similar to a big one (but the amount and the time-density of this liberated energy determines the effect of the earthquake). When not liberated as a sum of small tremors, energy accumulates and this may lead to big catastrophic earthquakes when a critical threshold is reached. This kind of phenomenon can be seen in an epidemiological context. Usually, the number of secondary new infected people in a given population are used to measure the importance of an epidemic (*e.g.*  $R_0$ ). The importance of an epidemic process is an issue that should be clarified.

Is an epidemic important when the infection germs (*e.g.* virus, bacteria,... ) are dangerous for their host (diseases such as Hepatitis, HIV, ebola), or when a disease spreads very fast in a given population (like ebola, influenza)? The notion of danger includes a psychological aspect : no one in Europe wants to be infected by HIV, although no one dies from HIV in Europe nowadays. An epidemic can also be declared important in relation to

its size (number of new infected in a given population during a given time period). The economic impact caused by an epidemic can also be considered to be essential to considering an epidemic process as important. In 2013, a lack of personal hygiene cost French companies a staggering 14.5 billion euros. This was 13.7 billion euros and 12.6 billion euros in UK and Germany respectively <sup>3</sup>. Epidemic of common cold, skin infections, food poisoning, respiratory diseases, gastro-enteritis and urinary tract infections are diseases that can easily spread because of lack of hygiene. A minimum of hygiene could, however, eliminate many viruses, bacteria and other germs. The same could be said about emerging vectors such as *Aedes aegypti* mosquitoes that vehiculate the arboviruses that causes Chikungunya or Zika diseases: reducing their habitat (*i.e.* water ponds, reserve of water in flower display cases ...) helps to control their impact.

The importance of an epidemic outbreak can also be seen at different microscopic to macroscopic levels. The notion of epidemic importance will therefore depend on the level at which epidemic outbreaks are observed. The importance of an epidemic at the microscopic level consists of both economic and health impacts on individuals or on small groups of individuals (in the above example, the lack of hygiene has a cost of about 900 euros per worker per year for the companies). The fees caused by a disease are also energy and time wasting, with various consequences on individuals, families, companies and governments. As an example, the importance of a gastroenteritis epidemic can vary from the individual level (... family level, small company level, ..., city level) to national population level.

---

<sup>3</sup>Reference: Centre for Economics and Business Research

During such an epidemic, only a few cases are reported: there are about 100 to 300 infected cases  $I$  per 100000 inhabitants, *i.e.* 6000 to 20000 infected cases and a respective density of  $0.0094 I/km^2$  to  $0.031 I/km^2$  in a country (France in this example) of 66 million inhabitants of about  $640000km^2$  and of average density  $103 inhabitants/km^2$ ). However, the densities of infected cases are in fact largely greater when viewed locally in the places where the disease spreads.

Therefore it seems necessary to understand the impact of an epidemic process at each different level (at all the levels at the same time).

A macro-event can consist of a set of numerous micro-events during a very short period of time or, on the contrary, in very rare micro-events during a very long period of time (*i.e.* in this case, for a macro-event to exist, no micro-events of different natures may occur, *e.g.* if the micro-events correspond to infections, the recovery rate must be very low). The first case (an intense and short macro-event) corresponds to a virulent contagious germ (*e.g.* ebola, influenza), while the second is not virulent but persistent in the infected organisms (*e.g.* HIV, EBV). In the first case the macro-event is short : it spreads rapidly but might be easily stopped. This might be contrary in the second case. Considering together the event sizes and their durations in a combined descriptor allows a more precise description of the nature of an infection and its importance. In this way, one can use a quantity that indicates the importance of an epidemic event and that combines the information of how big an event is and how long this event lasts. Hence, in the long term (coarse grain observation) this quantity, in the two cases evoked before (intense and short epidemic events vs long

non-virulent epidemic periods), might be the same. This quantity is analogous to the total liberated energy in the case of earthquakes: a part of the earth's crust deforms either by a long-term series of small events (weak energy) or by a unique big and short earthquake; the energy implied in the deformation is the same at the end and in any case the crust has changed.

Two macro-events might have the same degree of importance. As a quantity that measures both the information on the number of micro-events of the same nature and on the duration of the associate macro-event, we propose to define epidemic charges as the integration of series of micro-events over time. In an SIR model, three epidemic charges can be defined : the *infection charge*  $C_I$ , the *healthcare charge*  $C_R$ , and the *reserve charge*  $C_S$ . All are normalised by the environment surface  $s$  in which epidemic processes are observed. They are defined as follows :

The *infection charge* gives a measure of the importance of a disease, in terms of spreading (related to virulence) and of duration of the associated epidemics.

$$C_I = 1/s \int_t^{t+d} I(\tau) d\tau$$

By extension, the *healthcare charge* gives a measure of the manner a system (a population) resists the infection, including the formation of recovered people  $R$  by natural immunity or by vaccination.

$$C_R = 1/s \int_t^{t+d} R(\tau) d\tau$$

Finally, the *reserve charge* measures the – increasing – reservoir of susceptible people available for new epidemics. This could include the formation of susceptible people  $S$  by loss of immunity, but also due to birth or to non-immune cured people.

$$C_S = 1/s \int_t^{t+d} S(\tau) d\tau$$

These charges cannot be measured or understood in the context mean-field ODE/PDE models whereas they are naturally obtained in SSA-based and MAS-based models that consider the individual levels. On the other hand, these charges are based on both incoming (positive) and outgoing (negative) flows in each compartmental class, meaning that obtaining the same measurements in real systems, we need microscopic information that is not easily accessible. As an example, when measuring the infection charge, to determine the catastrophe events we need to know the exact instants at which the infected individuals become cured (remembering that a sequential occurrence of micro-events of a same nature corresponds to a macro-event called "rescue" or "catastrophe" depending on its sign and in regard to a particular class  $S$ ,  $I$  or  $R$ ) otherwise we are not able to define the duration of a given macro-event. This also gives rise to the problem of the filiation of micro-events (that concern individuals) to particular macro-events, notably when the spatiality is considered. In fact, when a disease spreads from one individual to another, it is extremely difficult to track exactly the disease incubation, infection or recovery period of each individual, or to know the exact trajectory of the disease

from person to person. Moreover, even if it was known, it would be difficult to attribute boundaries to such macro-events apart from by using spatial and temporal thresholds that limit the events in space and time. Although measuring charges is difficult, such a challenge is the possibility of using newly developed technological devices like smart-phones, smart-watches . . . .

Criticality analysis brings attention to the manner in which an epidemic event can be seen at different levels, from microscopic (*i.e.* at a fine-grain scale: individuals or small groups level (see figure ( 2.22 )), to macroscopic (*i.e.* at the population scale). The potential of this approach points out the importance of studying epidemiological models in different ways such as natural critical phenomena (avalanches, earthquakes, species extinctions *dots*). This notably provides fine-grain information such as a microscopical resolution of time, space and population, that could constitute additional and probably more useful information than the synthetic thresholds, which are conventionally used in epidemiology. Application of these charges is still a theoretical concept, however we show how charges are obtained in our simulations (Section 2.4.4) and how this concept could be used to control epidemics.

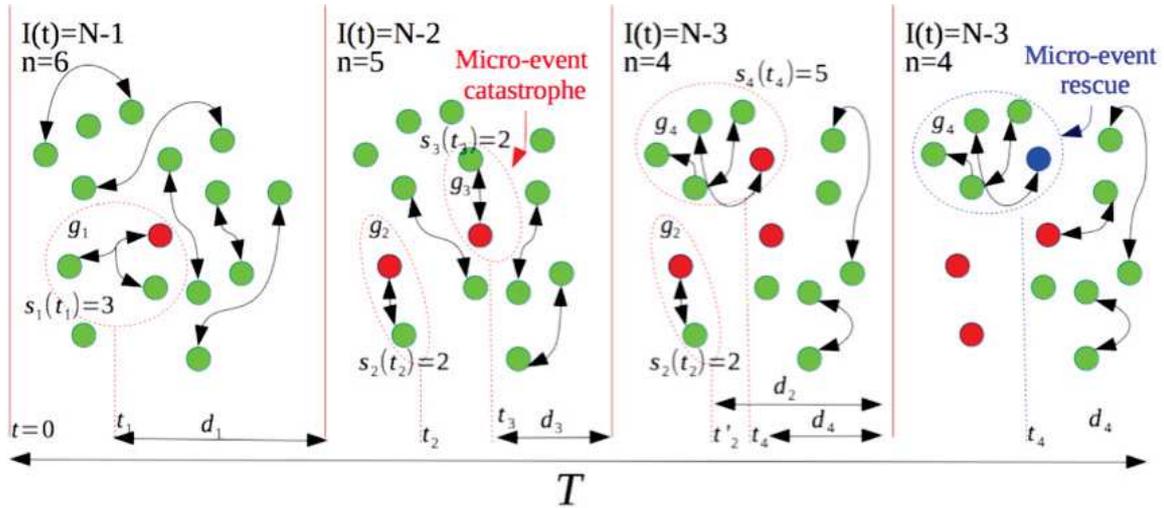


Figure 2.25 **Micro & Macro Events at Population Level** : Schematic diagram of an epidemic process in *micro-event level*:  $N$  is population size,  $n$  is number of groups formed during  $d$ , group-size at instant  $t$  is assigned by  $s_i(t)$  where  $i \in [2, N]$ ,  $d_i$  indicates a group-lifetime.

#### 2.4.4 Infection Charge as Signatures of Diseases

As explained in the previous chapter 2.4.3, the epidemic charges derived from *micro* and *macro* events can not be calculated by ODE or PDE-based simulation approaches: since fine-grain information is not available, it is not possible to measure epidemic events and consequently epidemic charges. Stochastic simulation based and multi-agent (including cellular automata) based simulations can somehow provide information on the distribution of these charges. The shape of epidemic charge curves might be a source of information in epidemiological studies. Their shape might indeed be affected by infection/recovery parameters, by the spatial characteristics of the environment, or even by the group-formation dynamics. One could expect to find characteristic sets of curves associated to certain diseases in certain environmental and social contexts, in such a way these distributions could correspond to signatures of diseases in their "natural" environment. The spreading of disease

is not the same in every environment, *e.g.* due to a different development of mosquitoes, chikungunya spreads differently in Europe and in the Caribbean ; the same could be said about the spreading of other diseases like HIV or Ebola over the different continents due to different environments and social contexts.

Such signatures would be helpful to identify which factor plays an important role in a given epidemic process and at which scale: some factors could play a role in small groups (and the associated limited epidemic charges) but not in important groups, producing charge curves far from being power laws. In our models, we can show the difference of epidemic charge distributions (here infection and healthcare charges) obtained from an SSA-based simulation and a Group MAS-based simulation of a classical SIR model using the same kinetic parameters. The normalised probability distribution of infection charge (catastrophe) and healthcare charge (rescues) shown in figure 2.26, points out the effect of taking into account the notion of groups (that locally increases the density of individuals and favours eventual infection processes) and the spatiality (that keeps away individuals in such a way that some interaction may never occur). This information is interesting to know how to reduce big epidemic events while managing the occurrence of small events, for example to stop big outbreaks by finding target sub-populations to apply prevention strategies. An explicit example is to control a common cold (caused by viruses like influenza, parainfluenza and rhinovirus): among the employees of a company, one could target sub-populations of employees that have children (a new study reveals that people living with children are infected with viruses up to 45 weeks per year, and this varies depending on the number of

children in the family [19]), and then by applying specific prevention strategies on these subsets, one could expect to control important outbreaks. The economic impact of epidemic outbreaks on public health at different levels (individual, family, societies, national and even international level) justifies the importance of managing the epidemics at the microscopic level to control them at the macroscopic level.

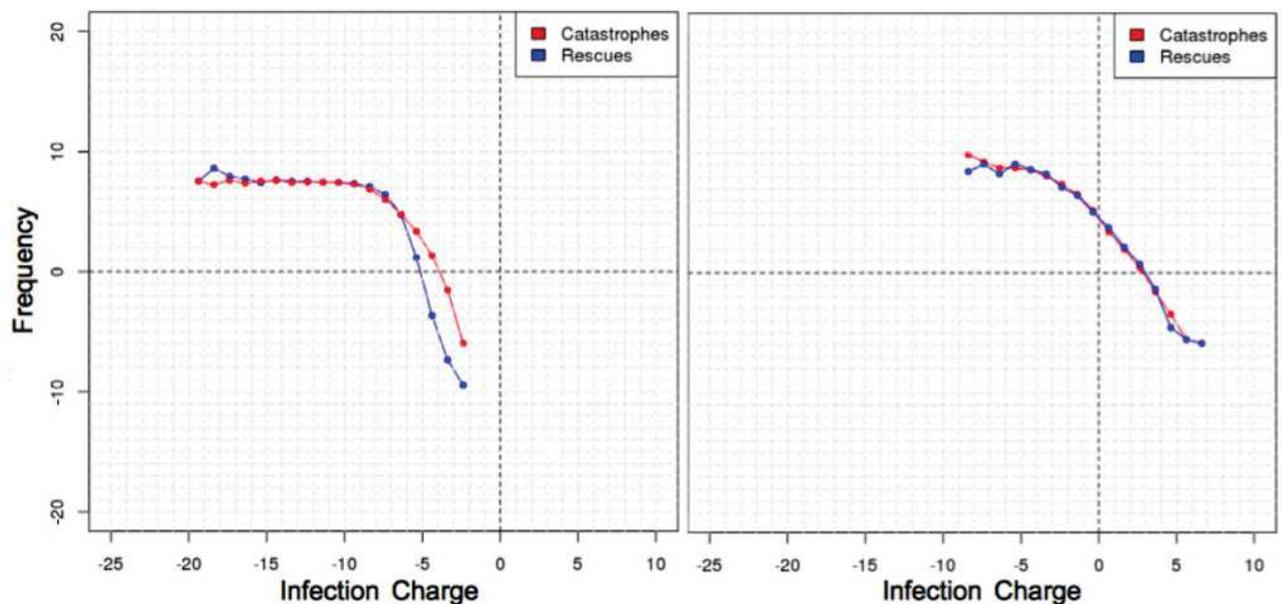


Figure 2.26 **Probability Distribution of Infection Charge** (in log-log plot) : SIRS epidemic process with parameters  $\beta = 0.1$   $\gamma = 0.01$  and  $\lambda = 0.1$  is simulated by SSA-based model in homogenous population (Left) and MAS-based with group structured population (Right). The characteristic shape of each cure might be considered as a signature for this SIRS epidemic process which depends on population structure.



# Chapter 3

## Results of our Epidemiological Simulations

### 3.1 SIR Epidemic process in ODE-based Model: Continuous models

#### 3.1.1 Modified SIR dynamics without resistance

In some epidemiological models (including ODE-based), the individuals in a given population (human, animal or even vegetal) can be considered homogenous (identical) or they can even be considered as being homogeneously distributed in their environment (space). In reality, the population density is very heterogeneous. A given environment might consist of areas with different population density, some of them empty of individuals or, on the contrary, very highly populated. Individuals in extremely dense populations have a higher likelihood of

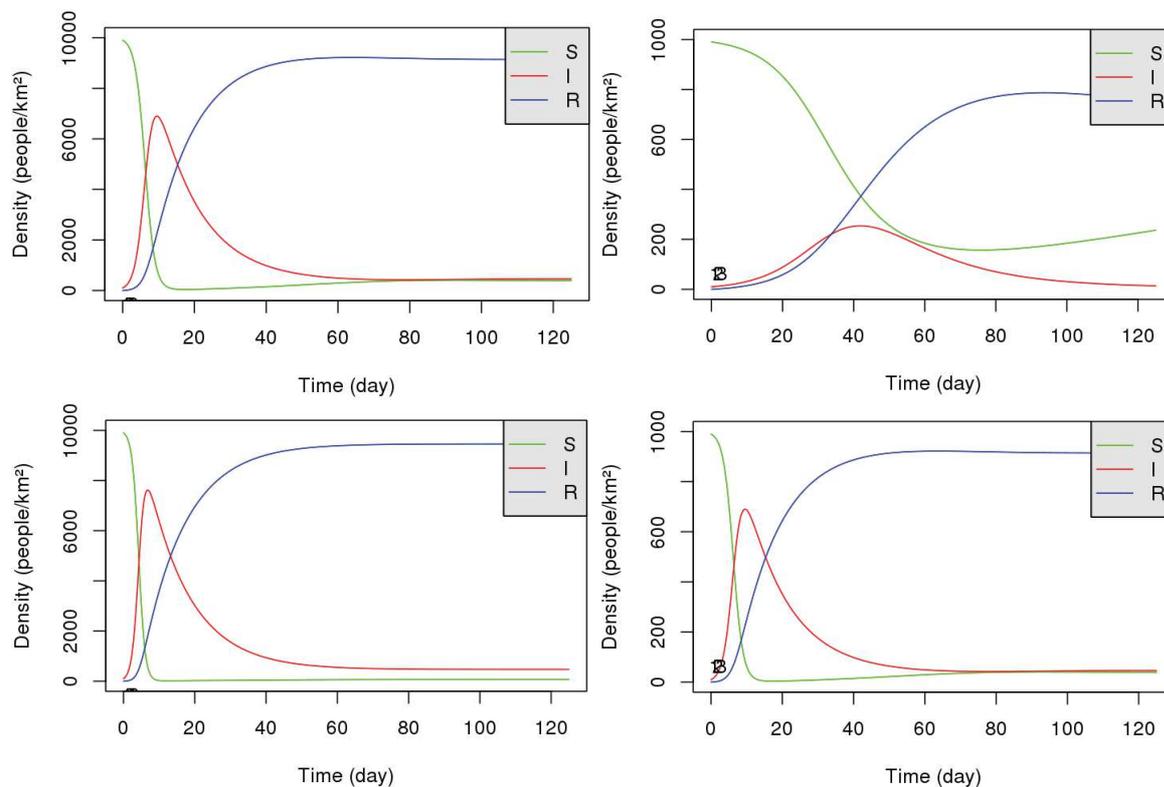
being in contact with infection agents, but this definitely depends on population structures. In this case, population density is not the only factor that can have an extreme impact on epidemic dynamics.

Here we show how population structures (the way individuals can be in contact with each other) can affect an epidemic process. For this reason, we use the kinetic SIR epidemic process (refer to section 2.2.1) in an ODE-based numerical model to study whether only changing the population densities affects the epidemic dynamics or whether any changes in population structures lead to more important impacts. In our modified SIR ODE-based model without *resistance* the only group-complexes that are allowed to form are *SI* ones. Their formation, a limited structuration of the population, is controlled by the dissociation constant,  $K$ . Changing the value of  $K$  (*i.e.* through  $k_1$  and  $k_{-1}$ ) causes a change in the quantity of *SI* group-complexes<sup>1</sup> (associations of *S* and *I*-individuals) in the global population. As an example, when  $K$  is decreasing, more *SI* groups might be formed and consequently many infectious events will occur. Here in the modified SIR kinetic equations, the dissociation constant  $K = k_{-1}/k_1$  has a more active role than the population density in epidemic dynamics. Both the balance and the individual role of  $k_{-1}$  and  $k_1$  are important: an important value of  $k_1$  means that lot of group-complexes form, but it can be counterbalanced by an even value of  $k_{-1}$ , which means that group-complexes disassemble quickly, thus giving a turn-over quantified by  $K$ . When the value of  $K$  is high (when  $k_{-1}$  is important and  $k_1$  low), the effect of the population density is important. On the contrary, when  $K$  is low (when *SI* form and/or are

---

<sup>1</sup>Reminder: Group notion in ODE-based model is not the same as in SSA or MAS-based model. Here we don't intend to study the same group dynamics, including group sizes and lifetimes, that can be seen in the MAS model. Here, the concept of a group is limited to kinetic complexes that appear as temporary chemical species and that provide connections between *I,S* and eventually *R* individuals.

stable), the role of the density is not really important. Indeed, this constant rate  $K$  affects the number of  $SI$  group-complexes that form, groups in which the infection transitions might be observed. This means that when there is a higher tendency to provide  $SI$ -groups (low values of  $K$ ), the population becomes a group-structured population and epidemic reactions happen regardless of the population density (see figure 3.1). On the contrary, when the population does not tend to a group-structured population (when  $K$  is important), good conditions for epidemic transmission can also be obtained by an increase of the density.



**Figure 3.1 Dissociation constant & Population Density Impacts on SIR Epidemic Dynamics :** These four simulations based on ODE illustrate the SIR epidemic model without resistance. The initial population in each class are:  $S_{int} = 99,000(\text{indiv.})$ ,  $I_{int} = 1000(\text{indiv.})$ ,  $R_{int} = 0(\text{indiv})$  and kinetics parameters are  $k_{cat} = 1.2 \text{ day}^{-1}$ ,  $\gamma = 0.08 \text{ day}^{-1}$ ,  $\lambda = 0.004 \text{ day}^{-1}$ . These parameters might define a communicable disease such as measles. The two plots on the left side of the figures (top and bottom) correspond to populations with a density of 10,000 Pop. per  $\text{km}^2$ , 10 times more than the population density corresponding to plots on the right block (1000 Pop. per  $\text{km}^2$ ). The dissociation constant in two plots on the top is high ( $K=5,000 \text{ indiv.km}^{-2}$  with  $k_1 = 5 \text{ km}^2 \cdot \text{indiv}^{-1} \cdot \text{day}^{-1}$  and  $k_{-1}=25,000 \text{ day}^{-1}$ ). These are not suitable conditions to form  $SI$ -groups. Then the infectious peak is reached very late when the density is low (top right) compared to when the density is high (top left). With a lower (10 times) dissociation constant ( $K=500 \text{ indiv.km}^{-2}$  with  $k_1=5 \text{ km}^2 \cdot \text{indiv}^{-1} \cdot \text{day}^{-1}$  and  $k_{-1}=25000 \text{ day}^{-1}$ ) as in the two plots on the bottom, the conditions are suitable for  $SI$ -groups formation. Then we can see (from left to right plot) that changing the population density has no impact on epidemic dynamics. Here, the structuration of the population (group-complexes) rules.

Therefore, the population density does not affect the epidemic dynamics when  $SI$  group-complexes are more likely to be formed. Otherwise, the model behaves as in SIR classical models in which the population is considered homogenous.

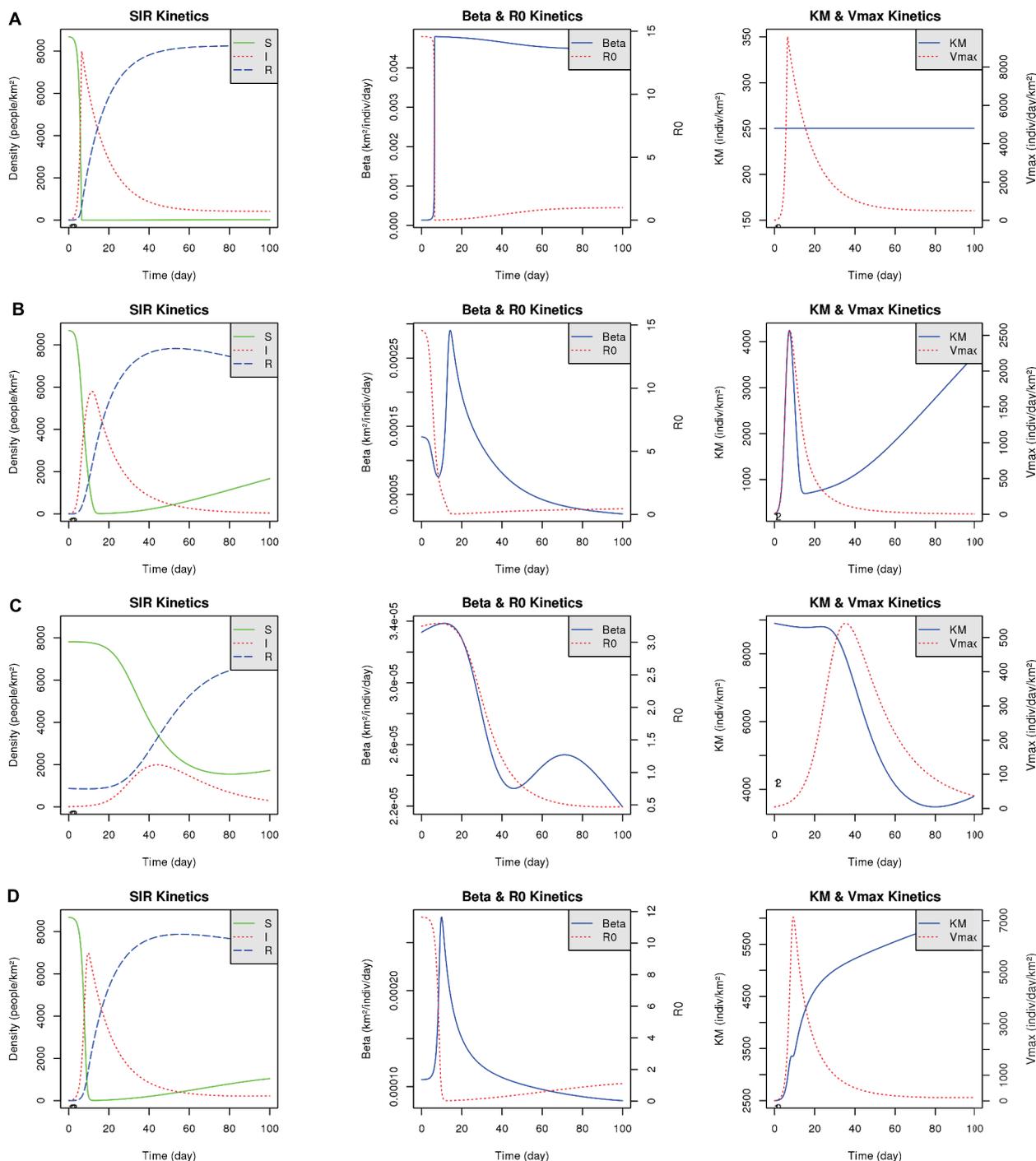
### 3.1.2 Modified SIR dynamics with resistance

This behaviour would be different if  $R$  individuals would play their role of resistants in the population. The inhibiting effect of the resistance appears by means of the formation of other group-complexes such as  $SR$  and  $IR$  or even  $SIR$  (wherein no infection transition can occur). We show here that the presence of resistant individuals has an inhibitory effect on newly infected individual dynamics. This is due to a disequilibrium in the formation of group-complexes in favour of  $SR$ ,  $IR$  and  $SIR$  group-complexes in which the infection is not possible, instead of  $SI$  group-complexes. Due to a non-competitive inhibition, a pumping effect from free individuals to  $SR$ ,  $IR$  and  $SIR$  group-complexes produces a lack of free  $I$  or  $S$  individuals in the population, thus preventing the formation of  $SI$  groups and the subsequent interaction between these two sub-populations. In an epidemiological context,  $SR$ ,  $IR$  and  $SIR$  group-complexes are inactive groups (non-infectious). These can be considered as temporary quarantine environments that isolate  $I$  and  $R$  individuals from direct interactions. The dissociation constant  $K$  is the parameter that controls the non-competitive inhibition process as well as the formation of  $SI$  group-complexes involved in the infection transmission. In fact,  $K$  is not related to the epidemiological process but its role is limited to control the formation of group-complexes, whatever they are. It should be noted that  $K$  could be distinguished respectively as  $K_A$  and  $K_B$ , which correspond to the respective dissociation constants of  $SR$  group-complexes on the one hand, and the three other group-complexes  $IR$ ,  $IS$  and  $SIR$  on the other hand (see Appendix A.2). A high dissociation constant  $K$  leads to a situation that is not suitable for forming any  $SI$  complex, which does not facilitate the disease spread nor the formation of  $SR$ ,  $IR$  and  $SIR$  group-complexes. Then, in such conditions of

high dissociation, because very few group-complexes form, the resistant individuals have a very limited impact on epidemic dynamics and the modified SIR ODE-based model behaves as seen in the bottom-right plot of figure 3.1 (modified SIR ODE-model without resistance under weak population density), *i.e.* as for a classical SI model. On the contrary, with lower values of  $K$ , as well  $SI$  as other group-complexes  $IR$ ,  $SR$  and  $SIR$  can form in the same time depending on the respective amounts of free  $S$ ,  $I$  and  $R$  people. As the number of  $R$  individuals increases in a population, more inactive group-complexes  $SR$ ,  $IR$  and  $SIR$  form in such a way that the conditions become less favourable to form  $SI$  group-complexes (more likely for infection transition) thus inhibiting the infection process. This confirms the role of prevention strategies (*e.g.* vaccination campaigns, prevention at TV, *etc.*) to inhibit the disease propagation.

Monitoring the evolution of maximum apparent speed ( $V_{MAX}$ , the maximum rate in enzymatic reactions), the Michaelis constant ( $K_M$ ) and the  $R_0$ , helps to understand the effects of  $R$  individuals in a population as inhibitors. The reproduction number,  $R_0$  is defined as the product of infection constant (of first order) and mean infection period such that  $R_0 = \beta_{app}[S] \times 1/\gamma$ . The  $R_0$  reaches a maximum when  $\beta_{app}$  is maximal and this happens when an infection is starting to spread. This value is not constant and, at its maximal level, it can be compared to the  $R_0$  that is known as the basic reproduction number (as measured in classical SI models). Figure 3.2 illustrates the impact of  $R$  individuals and its dependance on the structuration of populations. The plot of SIR kinetics in the first row (Figure 3.2 A) corresponds to a modified SIR model without resistance. The plots in the second, third and last rows (Figure 3.2 B, C, D) show kinetics of modified SIR models with resistance. These SIR kinetic plots (first

column) are accompanied by the evolution of  $\beta_{app}$ ,  $R_0$  (in the second column), and  $K_M$  and  $V_{MAX}$  (in the third column) during a time of  $t = 100$  simulated days.



**Figure 3.2 Effects of  $R$ -individuals on SIR Epidemic Dynamics.** The first simulation (A) uses a modified SIR model without resistance, while the three other simulations (in B, C, D) are based on a modified SIR epidemic model with resistance, *i.e.* by considering  $R$  individuals as inhibitors in a noncompetitive inhibition process. The initial population in each class are:  $S_{int} = 675,000$  individuals,  $I_{int} = 675$  individuals,  $R_{int} = 0$  individuals in the simulations A,B,D, but  $R_{int}=6750$  individuals (10 times more) in the simulation shown in C. The population density is assumed constant in all simulations and is equal to 8,700 individuals per  $km^2$ . The common kinetics constants are:  $k_{cat} = 1.2 \text{ day}^{-1}$ ,  $\gamma = 0.08 \text{ day}^{-1}$ ,  $\lambda = 0.004 \text{ day}^{-1}$ . The three first simulations (A,B and C) use low dissociation constants ( $K=250 \text{ indiv.} km^{-2}$  with  $k_1=5 \text{ km}^2 \cdot \text{indiv}^{-1} \cdot \text{day}^{-1}$  and  $k_{-1}=1,250 \text{ day}^{-1}$ ), on the contrary to the last simulation (D) in which the dissociation constant is 10 times higher than in the other simulations ( $K=2,500 \text{ indiv.} km^{-2}$  with  $k_1=5 \text{ km}^2 \cdot \text{indiv}^{-1} \cdot \text{day}^{-1}$  and  $k_{-1}=12,500 \text{ day}^{-1}$ ).

The three simulations (A, B, C) with a low value of  $K$  ( $K=250\text{indiv.km}^{-2}$  with  $k_1=5\text{km}^2.\text{indiv}^{-1}.\text{day}^{-1}$  and  $k_{-1}=1,250\text{day}^{-1}$ ) are favourable to the formation of group-complexes: with a density of  $N = 10000\text{indiv.km}^{-2}$  the effective association would be of  $N \times k_1 = 50000\text{day}^{-1}$  (50000 events of group-complex formation per day) while there would be only  $1250\text{day}^{-1}$  dissociations of group-complexes, giving a ratio of  $N \times k_1/k_{-1} = 50000/1250 = 40$  times more associations than dissociations. On the contrary, the last simulation (D) is less favourable to the formation of group-complexes compared to the other simulations : dissociations is 10 times higher, giving a very small advantage to associations compared to dissociations (4 times only) for the same population density.

The effect of  $R$ -individuals is very clear when comparing the SIR kinetics curves in A (modified SIR model without resistance) and B (modified SIR model with resistance). The consequences are particularly obvious by comparing the values of  $R_0$  and  $\beta_{app}$  in their corresponding plots. These effects are even more obvious when comparing conditions in which there is few resistance at the beginning ( $R_{init}=1\%$  of the population) and other in which many resistants are already present ( $R_{init}=10\%$  of the population) (comparison of simulations B and C). One can see that resistance have an effect on both the maximum number of infected people (about  $6000\text{I.km}^{-2}$  in B but only about  $2000\text{I.km}^{-2}$  in C) and at the time at which this maximum is reached (the maximum of infection ; about  $10\text{day}$  in B compared to about  $40\text{day}$  in C). This is expected since, due to the non-competitive inhibition,  $R$  individuals play a role on both the  $K_M$  and the  $V_{MAX}$  of the infection process.

Finally, one can compare to conditions in which the formation of group-complexes is very limited due to a strong dissociation constant  $K=2,500\text{indiv.km}^{-2}$ , *i.e.* conditions in which

encounters between  $S$  and  $I$  individuals become rare (simulation D). This time, the SIR kinetics obtained are very similar to those obtained in the absence of resistance (simulation A) ; even the  $R_0$  and the  $V_{MAX}$  can be compared, but one can nevertheless observe strong differences in the behaviour of  $\beta_{app}$  and  $K_M$ . These parameters are indeed no more negligible when the density of  $R$  becomes very important, *i.e.* after the infection peak. This produces a long term effect : infection is inhibited and the  $S$  reservoir can fill again ; it is not possible in simulation A where the resistance is absent: the remaining  $S$  individuals are constantly consumed by remaining  $I$  individuals.

The dissociation constant  $K = k_1/k_{-1}$  is explained by different combinations of  $k_1$  (association) and  $k_{-1}$  (dissociation) constants. Then, high values of  $k_1$  and  $k_{-1}$  explain the formation of many group-complexes (*i.e.* thus increasing the infection probability) that live only during short periods of time since they also dissociate very fast. When both these constants are low, some groups (many inactive complexes among them) are formed for long periods of time (low dissociation rate). As seen in figure 3.1 showing the impact of population densities, the effect of the resistant population on epidemic dynamics are also derived from the role of the  $K$  constant; the average turnover of formation/dissociation of group-complexes. One can observe the same effects in SSA-G, SSA-CA and Group-MAS based models (next section), by changing the simulation parameters such as the individual's speeds or the global population densities, *i.e.* those that influence the individual-based group formation dynamics.

However, the results obtained here show that even by using an ODE-based model with our

modified SIR kinetics, that is to say a type of model that mostly concerns the population level (they are not about individual interactions), one can study the effects of factors that occur at an individual level, and their consequences at the population level. The formation (or dissociation) of group-complexes in a given population doesn't allow us to study the epidemic process in heterogeneous structured population (*e.g.* epidemics within groups of individuals). However, in such a model, information at the individual and at population level remain connected. For example, adjusting the value of  $K$  affects the structure of individuals' contacts as surely adjusting the individuals' moving rates and social tendencies would do. The constant  $K$  may be written as an expression depending on individual parameters like speeds and social tendencies. Finally, one can say that inactive group-complexes represent the average proportions of immune-groups in a population, *i.e.* groups in which infection processes are very rare. This is useful for understanding by way of such models the influence of vaccination campaigns (more  $IR$ ,  $SR$  or  $SIR$  complexes in populations can be considered as a consequence of vaccination campaigns).

## **3.2 SIR Epidemic process in SSA & MAS-based Models- Discrete models**

Here, we study the epidemic dynamics in structured populations (*e.g.* populations having groups of heterogeneous sizes, composition and lifetimes, populations in which individuals have social behaviours, preferential places and habits, *etc.*), by models based on stochastic simulation algorithms (SSA) or multi agent simulation (MAS), particularly with 3 models

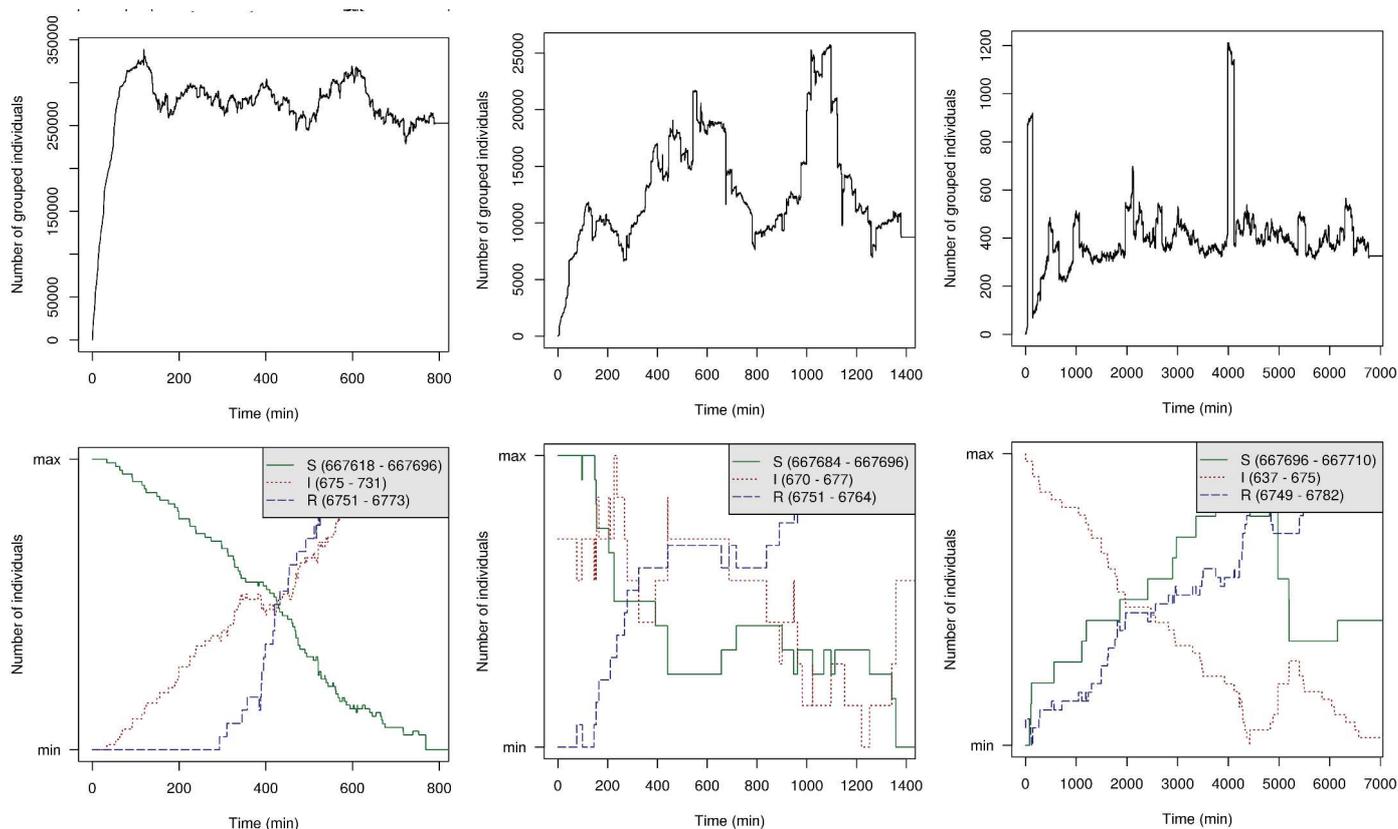
: a stochastic simulation model in homogenous space but with a group structure (SSA-G), an individual-based stochastic spatial model using a cellular automaton (SSA-CA), and a multi-agent system having a group structure (MAS-G), all presented in Chapter 2.

### 3.2.1 SSA-G Model

It is reasonable to assume that dense populations or groups of individuals are more likely to give rise to faster disease outbreaks. On the contrary, in the populations or groups in which individuals are far away from each others or where they have less chances of being in contact with each-other, the SIR dynamics will behave differently: infection events should be more rare. This assumption can be admitted according to the manner that diseases propagate during catastrophic epidemics in the real world. Figure 3.3 shows how a decrease of the global population density can affect the infectious process. These three simulations are realised with the SSA-G model where group-formation is controlled (as seen in section 2.2.2.1) by individual speed  $\bar{v}$  and average interaction radius  $d$  but also by the available free population and its effective density, a value that depends on the effective surface in which the population live. The graphic shows three different epidemic dynamics in the same environment depending on various levels of population density (from more populated to less densely population). The density is adjusted by the effective surface proportion parameter: given a global average surface in which the population of 682425 individuals (including 675000  $S$ , 675  $I$  and 6750  $R$  at the beginning of the simulations) live, we define a proportion of this surface, the effective surface that the population effectively occupies. In these three simulations, the proportion corresponds to 0.1%, 1% and 10% of the global surface. The first plots (Figure 3.3 Left)

correspond to a very dense population: about 680000 people in 0.1% of a global surface of  $68km^2$  (density of  $10000indiv.km^{-2}$ ), as to say an effective density of about  $10^6indiv.km^{-2}$  ! In such a dense environment, many groups form and favours the propagation of the disease. In the two other simulations, the situation is completely different. When the effective surface is multiplied by 10 (1% of the global surface), the average number of groups that form is reduced by a factor of 10 (Figure 3.3 Middle). In these conditions, the infection does not develop; it is quite stable (maintained stable over a day). Finally, the conditions of the last simulation (effective surface: 10% of the global surface) do not allow to maintain the infection (Figure 3.3 Right). The environment is too much diluted and the infection epidemics are almost stopped by the presence of an increasing number of resistant individuals.

We clearly show here the importance of having a group structure in the population and on the importance of spatial parameters, notably the effective surface. Other parameters play a similar role like the individuals' speeds and the interaction radius attributed to each individual.



**Figure 3.3 Population Density & Stochastic Simulation Algorithm of Group-based Model (SSA-G)** These three simulations use a SSA-G model to simulate modified SIR Kinetics with resistance at the scale of big population (*e.g.* big cities). The initial population in each class and common parameters are:  $S_{init} = 675,000$  individuals,  $I_{init} = 675$  individuals,  $R_{init} = 6750$  individuals,  $k_{cat} = 1.2 \text{ day}^{-1}$ ,  $\gamma = 0.08 \text{ day}^{-1}$ ,  $\lambda = 0.004 \text{ day}^{-1}$ . These parameters might correspond to the dynamic behaviour of diseases like measles. The dissociation constants used are low ( $K=150 \text{ indiv.km}^{-2}$  with  $k_1=5 \text{ km}^2.\text{indiv}^{-1}.\text{day}^{-1}$  and  $k_{-1}=750 \text{ day}^{-1}$ ), which favours the formation of  $SI$ ,  $SR$ ,  $IR$  and  $SIR$  group-complexes in the population. The mean individual speed is  $\bar{v} = 0.1 \text{ m.s}^{-1}$  and the average interaction radius  $d = 0.5 \text{ m}$ . In the simulations, from the left to the right, the effective surface proportion varies: (Left) 0.1% of the global surface, (Middle) 1% and (Right) 10%.

(Left) In the left plot SIR dynamic corresponds to those that would occur in a dense population wherein many groups form (almost 40% of total population). A very fast infection process is observed (more than 60 new  $I$ -individuals in almost 13 hours). (Middle) In the second simulation, the population is less dense than in the previous one. Obviously less groups form (only 2% of the total population). The number of  $I$ -individuals is quite stable, maintained at a value close to the initial value for a duration of about 24 hours. (Right) The situation of the 3<sup>rd</sup> simulation becomes more extreme. Very few groups form, which disfavours the encounters between  $S$  and  $I$  individuals. moreover, the infection epidemic is expected to vanish as it is progressively stopped by an increasing number of  $R$ -individuals in the population.

### 3.2.2 SSA-CA Model

The individual's move speed can influence the way people of different – epidemiological – nature are mixed within groups. This affects the role of  $R$ -individuals. Assuming that there are some parts of the environment where sub-population are organised in a way there is always less  $R$  than other people ( $R$  people staying in a limited part of the environment), then when an epidemic starts in that population, these parts of the environment might be subject to many catastrophic and important macro-events and few rescue events. This would lead to a strong increase of disease outbreaks in those parts since the importance of  $R$  sub-populations would be negligible. This is particularly pronounced when there is no mixing of individuals (no moves) between two such regions (one without resistance and the other saturated by resistant people), producing epidemic outbreaks that are more severe in the area without resistant individuals than within the parts with more resistant individuals. This is the case at the scale of big environments (regions and countries) where immune people are well represented in some places (villages) but not in the surrounding places; in this case, the epidemic spreads around without encountering any resistance. At a smaller scale, when people are surrounded by resistant individuals, the disease can not spread easily and often fails to spread. This is the case when a significant part of the population is vaccinated. All of this addresses average situations, viewed at large scales (big populations) and over long periods, but since epidemics are the expression of a sum of many micro-events, one must consider the importance of resistance at the level of few individuals during short periods.

Vaccinated individuals might indeed be considered as local obstacles to prevent the local propagation within limited areas. This depends on a ratio between the mixing rate (individual

speeds) and the infection rate, that is to say between a kind of diffusion (at least a transport of matter: the individuals) and a reaction. When the ratio is in favour to the diffusion, the reaction becomes homogeneous over space and in the case of epidemics, this favours the infection. When the diffusion is limited (limited moves), the reaction occurs in very small areas (heterogeneities in space) and the disease does not spread easily. In real life, there are many situations in which people mix very fast (market, school playgrounds, *etc.*), or on the contrary when they don't mix (school classes, cinemas, queues in an administration or at the supermarket, *etc.*).

In the case where the mixing of sub-populations is very fast then  $I$  individuals have the same chance as  $R$  individuals of meeting  $S$  individuals. We evoked a well known examples (markets, *etc.*), but there are some facilitators that change the rules of people mixing in our societies. An example for this can be the effects of new technologies like the Internet (social networks where people can fix a date) as facilitators to meet people. For sexual experiences this can increase the number of sexual partners in a population and consequently lead to an epidemic of sexually transmitted diseases (STD). In this case, some 'hookup applications'<sup>2</sup>, *i.e.* geolocated social applications on smartphones (like the well known Tinder or Grindr apps) play a considerable role in the rise of STD in the population. A New York University study in 2013 ([20]) shows an increase of 16% of HIV cases in the USA (across 33 states) between 1999 and 2008 that were caused by dates organised through *Craigslist*<sup>3</sup>. Another example comes from the *Rhode Island Department of Health* in the US, that showed that between 2013 and 2014, the cases of syphilis grew by 79%, HIV infections were up 33%

---

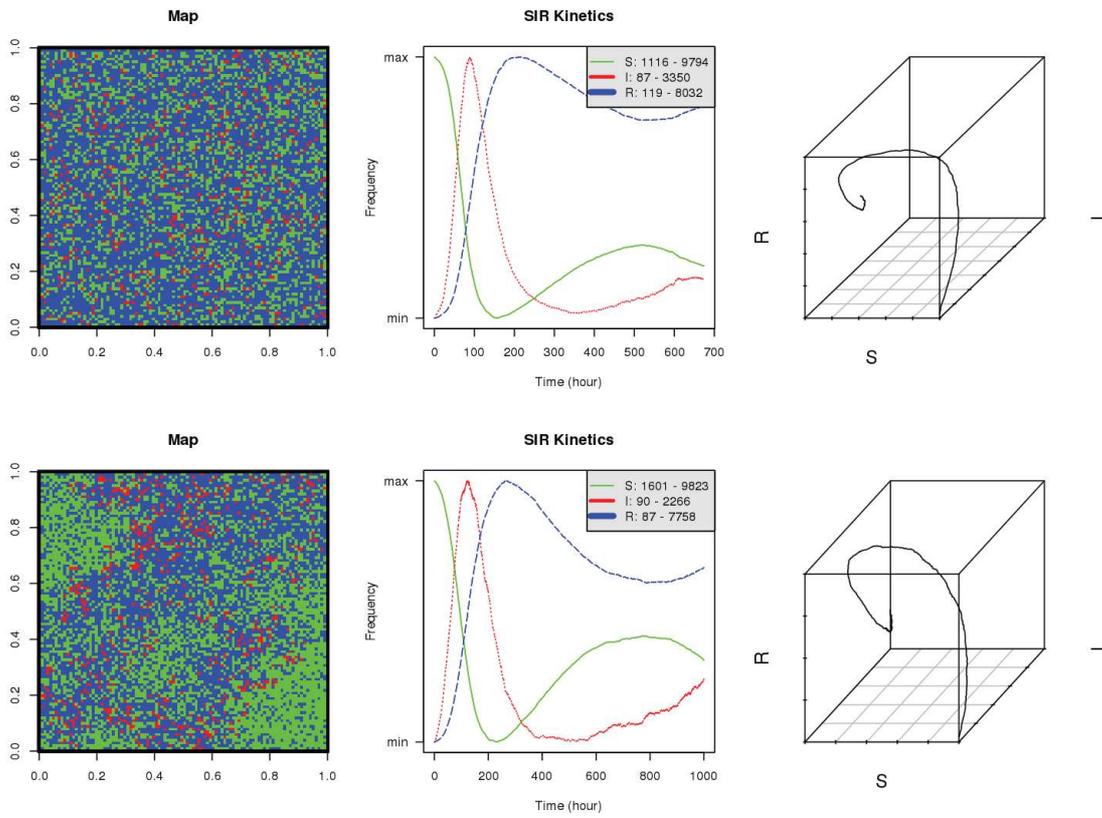
<sup>2</sup>Mobile phone applications used to meet people for casual dates

<sup>3</sup>Classified advertisements website

and gonorrhoea cases increased by 30%, partly due to such applications. STD cases in young adult populations are growing at a faster rate than the rest of the population, particularly because of the increasing use of "hookup apps".

The effect of resistance can be clearly explained by stochastic cellular automata simulations. In the SSA-CA model, the propagation of the disease is due to direct contacts in the close neighbourhood, from individuals to their nearest neighbours, and the SIR dynamics obtained show a stochastic behaviour. When the population is important (10000 individuals in a  $100 \times 100$  cells of  $100 \text{ m}^2$ ), the stochastic fluctuations are averaged and we obtain curves similar to those obtained by deterministic continuous ODE simulations. Figure 3.4 illustrates the negative effect of a high move speed that mixes individuals. The simulation shows that when individuals move quickly ( $0.1 \text{ m.s}^{-1}$ ), the infection peak reached early (at the  $3.75^{\text{th}}$  day), the proportion of infected individuals is about 33% of the population. When individuals move slowly ( $0.0001 \text{ m.s}^{-1}$ ), the infection peak is reached later (at the  $6.25^{\text{th}}$  day) and the proportion of infected individual is lower (only 22% of the population).

In this figure, the nature of the individuals is distinguished by different colours in the environment ( $S$  in green,  $I$  in red,  $R$  in blue). When the individuals move quickly, the environment is permanently mixed at a large scale and appears homogeneous (Figure 3.4 Top Left). On the contrary, one can observe large areas where only one type of individual, mainly  $S$  and  $R$ , dominate the others, the environment appears heterogeneous (Figure 3.4 Bottom Left). One can also observe that  $I$  individuals are mainly located at the boundaries between  $S$  and  $R$  regions.



**Figure 3.4 Impact of resistance and moving rates on SIR Dynamics:** These two simulations use a stochastic model of modified SIR kinetics implemented in a cellular automaton with common parameters  $\beta_{app} = 0.03 \text{ indivi}^{-1} \cdot \text{day}^{-1}$ ,  $\gamma = 0.08 \text{ day}^{-1}$  and  $\lambda = 0.004 \text{ day}^{-1}$ . The population consists of 10,000 individuals (or 10,000 cells) and the environment covers an area of  $1 \text{ km}^2$  ( $100 \times 100$  cells of  $10 \times 10 \text{ m}$ ). On the left is shown the simulation environments in which the individuals are identified by colours : green for  $S$ , red for  $I$  and blue for  $R$ . On the middle are shown the oscillatory SIR dynamics as a function of time, and on the right SIR dynamics are displayed in a state diagram that shows the beginning of the spiral of damped oscillations. The move speeds of individuals are different in the two simulations : (Top)  $0.1 \text{ m.s}^{-1}$  ) vs. (Bottom)  $0.0001 \text{ m.s}^{-1}$ . In the first case with a lower move speed, the epidemic infected curve peak is reached after only 90 hours (3.75 days) on the contrary to the second case that occurs after 150 hours (6.25 day).

In fact, these are infectious propagation fronts. These fronts of  $I$  between  $S$  and  $R$  regions are clearly visible in figure 3.5 that shows SIR kinetic dynamics in a simulation environment map during 12 days (4 days between each picture). Fronts of individuals form due to a lateral inhibition phenomenon, a kind of scorched earth policy: a small core of  $I$  individuals first

forms (first map of figure 3.5), the injections spread by increasing laterally (radially) the number of  $I$  individuals and by consuming  $S$  individuals. The front of  $I$  advances but as the simulation advances,  $I$  individuals may be converted (stochastically) in their turn into new  $R$  individuals. The more  $I$  individuals in a region, the more  $R$  individuals will form. Since the disease can not be transmitted to  $R$  individuals, the only way it can spread is toward regions densely populated in  $S$  individuals, causing the front to advance. New cores of  $R$  individuals form in place of  $I$  dense regions. Then, the process regenerates  $S$  dense regions since  $R$  individuals may become susceptible another time (loss of immunity). At large scales, this produces waves of  $I$  individuals, as observed in many excitable media like the Belousov-Zhabotinski reaction [41].

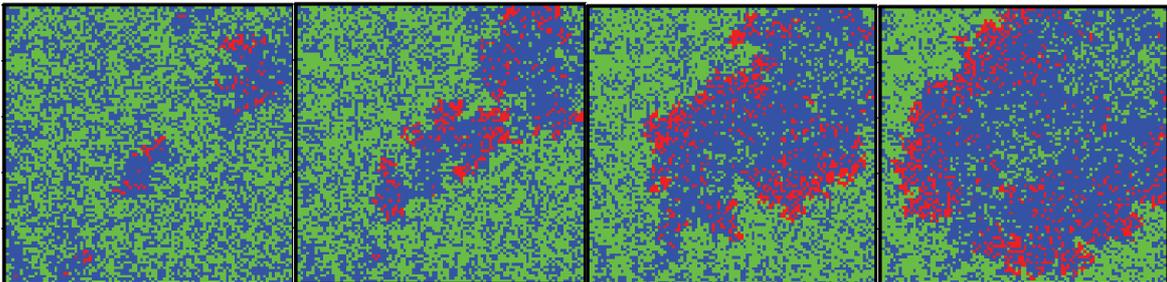


Figure 3.5 **SIR Kinetics Propagation Fronts** : The environment maps shown here are a snapshot of a SIR kinetic dynamics with the same parameters as those given in previous simulation except for  $\beta_{app} = 0.2 \text{ indivi}^{-1} \cdot \text{day}^{-1}$ . Increasing the infection rate allow us to highlight the formation and propagation of infection fronts (from left to right in above images). This simulation illustrates the propagation fronts of  $I$ -individuals and how  $R$ -individuals play the inhibitor role to prevent disease spread.

### 3.2.3 MAS-G Models

The population density can be considered at two different levels: global and local. As shown in previous models based on ODEs, the population density at the global level corresponds

to the number of total individuals in a given population, per unit area. As shown by those models, density at the global level can influence the epidemic process and group-formation dynamics at the same time. However, how about the densities at group levels and their impact on the epidemic process?

In a structured population (*e.g.* a population with a group structure), the density at each level can separately affect the epidemic dynamics. In this case it is necessary to study coupled group formation and epidemic processes by models in which group dynamics are based on "intelligent individual interactions" and on "social aspects". This is to say that all interactions between individuals that produce groups of a certain size, lifetime and effective surface, depend on each individual decision. Multi-agent systems are the most adapted and more reliable than other modelling approaches to study the epidemic dynamics in a given population at the microscopic scale. As described in Chapter 2, a multi-agent and group-based model (MAS-G) is used to study group formation impacts on SIR epidemic dynamics. We compare here the simulation results obtained by such an MAS-G model with another MAS model in which there is no group formation.

In this MAS-G model, as mentioned before (see section 2.2.2), we attribute a random probability number,  $0 \leq \alpha < 1$ , to each agent (individual) to represent their own grouping tendencies. This causes individuals to decide to form their own group or to join existing groups or not. Therefore there is no external control on the manner groups form (*i.e.* on the quantity *group size*). Group lifetime is computed by a concerted procedure in which all the group's members participate and that is based also on a probabilistic distribution (refer to section 2.2.2.3).

The simulation environment consists of a grid of about  $200 \times 200$  cells, each cell being a square of  $4 \times 4$  m. A number  $N = 10^4$  agents (individuals) simulated in such an area represents a population density of 15,625 inhabitants per kilometre squared, a value between the average densities of Grenoble with  $8,837 \text{ indiv.km}^{-2}$  and Paris with  $21,498 \text{ indiv.km}^{-2}$ . The time step of simulations is 3 seconds (similar to the shorter group lifetimes) and we simulate about 10 to 20 days. *NetLogo* is a very flexible programming environment, well adapted to such MAS, but it is not well optimised for large scale simulations, so our runs used weeks to months of real time for a few tens of days of simulated time.

Once the simulation starts, groups form and a SIR epidemic process based on Tau-leaping SSA (discrete time version of Gillespie's algorithm) is applied at both levels, population and groups. Then the actual population density depends on the number of individuals already in a group per square unit of a given area ( $m^2$ ). The density-dependent transmission algorithm is not affected by the number of solitary individuals that have not participated in group formation. The fact that an individual is present in a given area without any contact with the rest of the individuals can't cause any infection transmission and such an individual can only be considered as a quarantined individual.

The total number of groups that form, during the 432000 steps of simulation time (15 days), is equal to about 90000. In a characteristic simulation, we obtained 92737 groups including 82519 small groups ( $n \leq 5$ ), 10136 middle sized groups and 82 larger groups ( $n > 10$ ). As expected, but to be confirmed, our algorithm works well, producing size distributions in accordance with the global probabilistic law used in the SSA-G model: group

sizes are exponentially distributed. Since all groups are not supposed to have at least one  $I$  or  $R$  individuals among their members, they can't be considered as active groups in which the epidemic process algorithm provides at least one of the possible reactions (infection, immunisation, loss of immunity). An example of simulation (after 15 days) with 1990 groups (1238 small, 743 middle size and 9 bigger groups) is plotted in figure 3.6. Population densities (grouped individual density) are displayed for each  $80 \times 80 m$  blocks (dotted area) in grayscale that represent areas from the less populated (white) to very densely populated ones (black). This illustrates how groups are heterogeneously distributed in space.

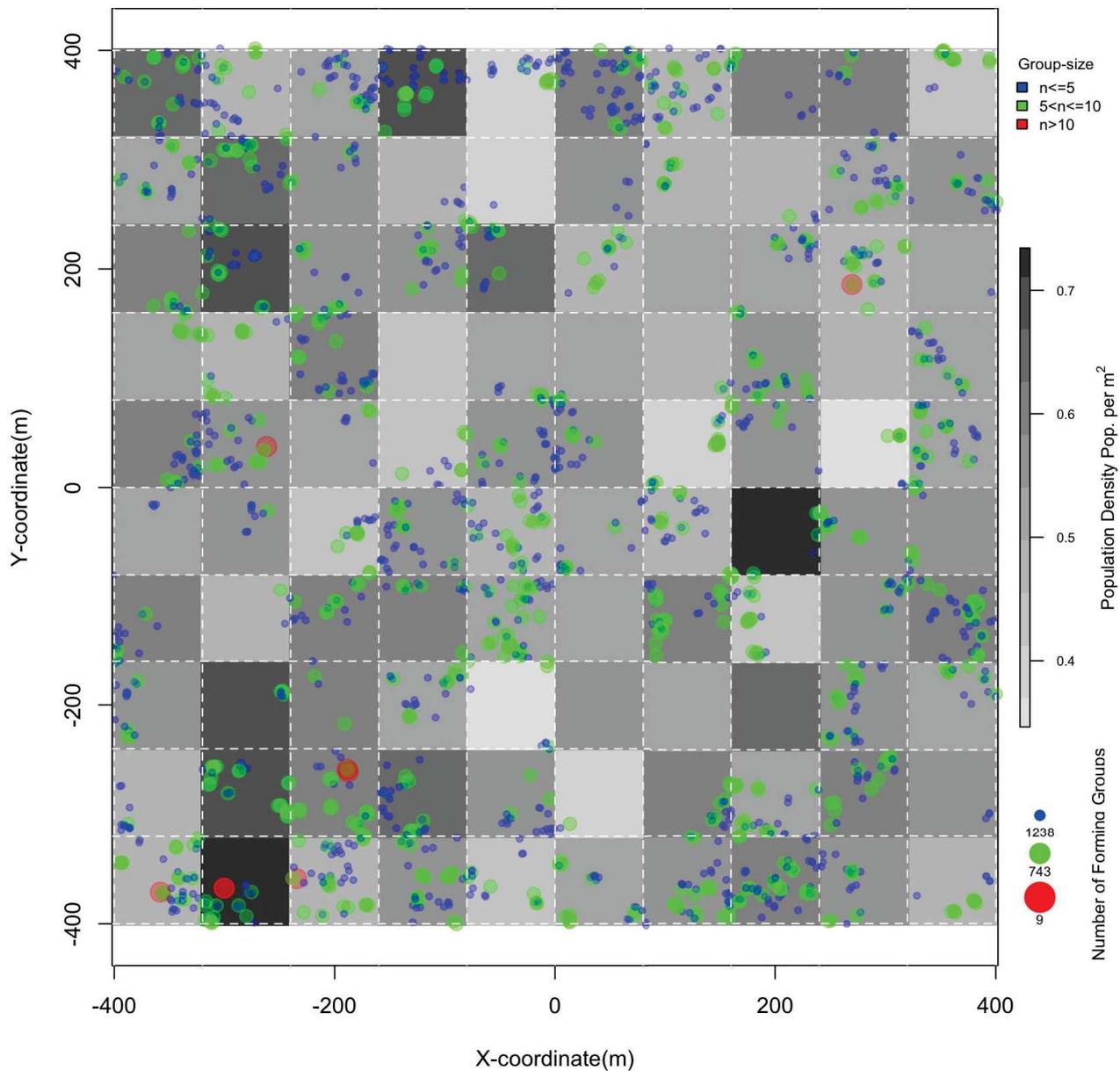


Figure 3.6 **Group Formation & Population Density in the MAS-G Model:** Density is calculated as the number of total grouped individuals (active and inactive groups) per square meter in each area block (dotted area) of  $80 \times 80$  m. The number of active groups is 1990 (including 1238 small, 743 middle size and 9 bigger groups) after 432000 simulation steps (15 days of simulated time). They are plotted on the density map.

In figure 3.7, the active groups are plotted in the same way but, this time, the map shows the number of infection outbreaks per block. The intensity of infection outbreaks is displayed

by a white, then yellow to red palette to represent the different levels from no-infection to high infection.

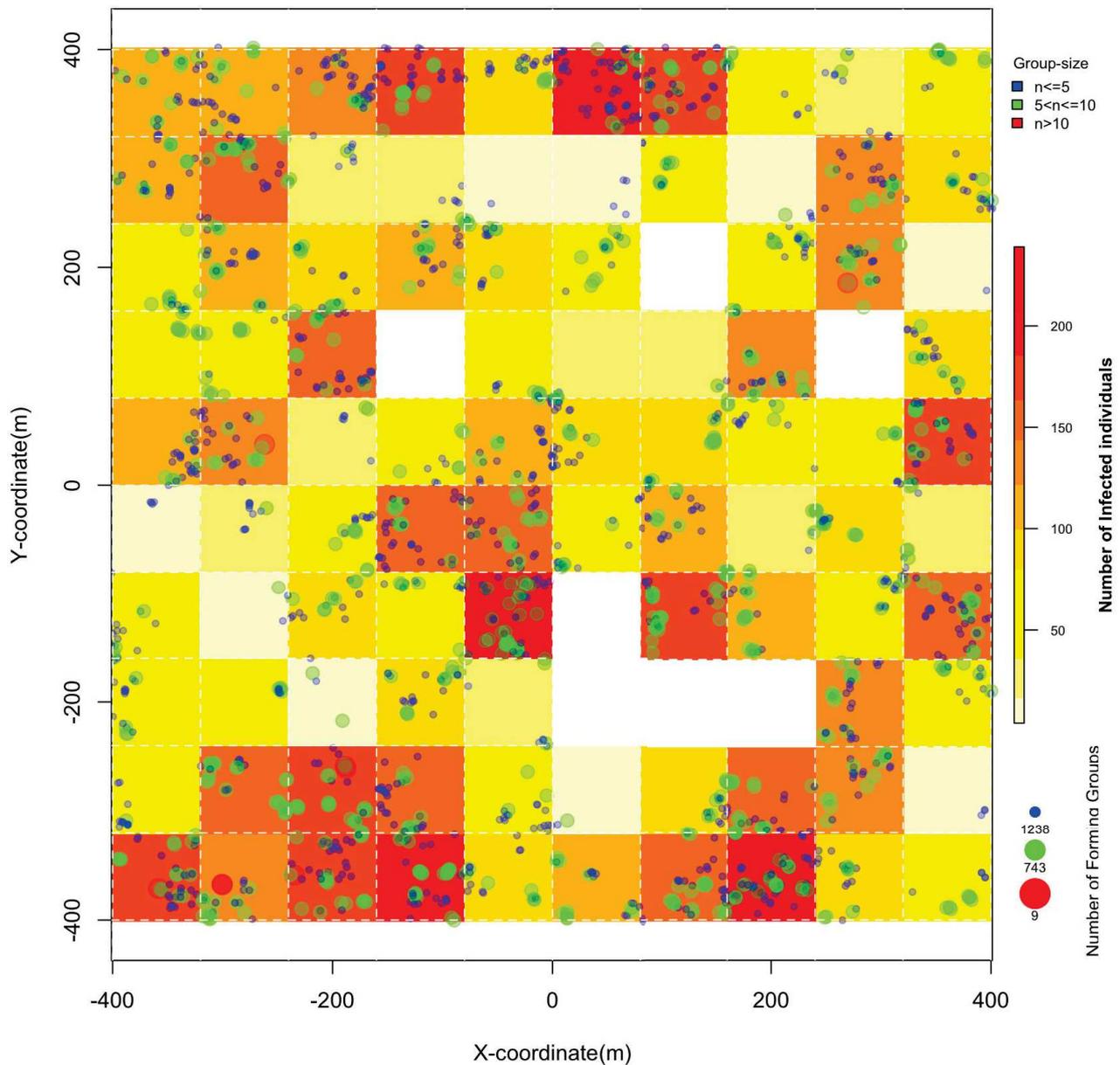
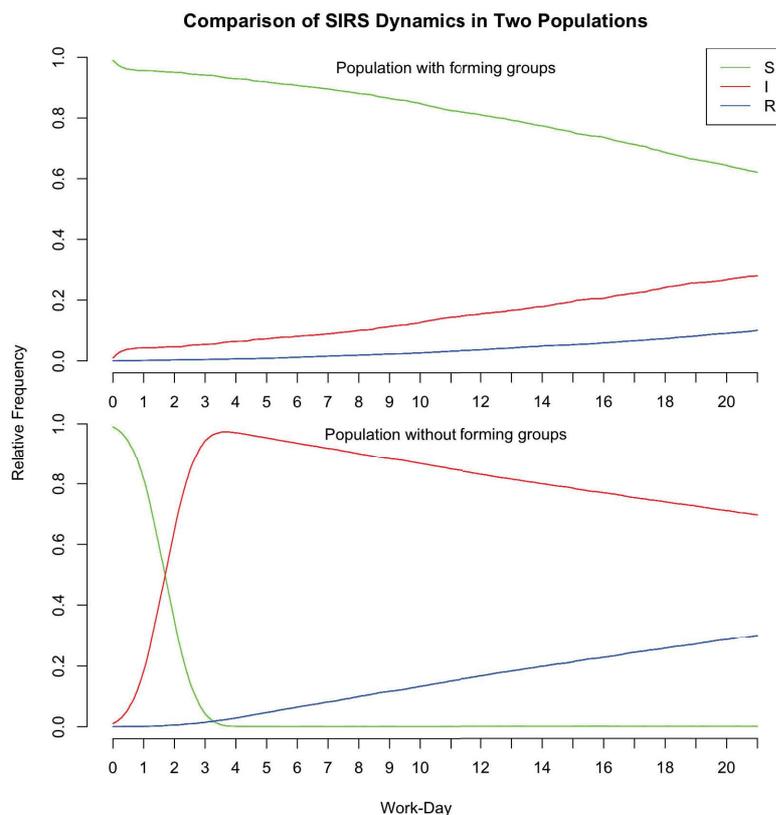


Figure 3.7 **The Group Formation & Number of Outbreaks in MAS-G Model:** Outbreaks are the total number of infected individuals per square meter for each area block (dotted area) of  $80 \times 80$  m, at the end of the simulation. Only the active groups are plotted in figure 3.6

When comparing the distribution of infection outbreaks in figure 3.7 with the densities in figure 3.6, confirms that a dynamic behaviour similar to that of the SSA-AC model is obtained: the blocks with a high grouped population density but few active groups also have few infected areas since infections occurrences are limited to active groups. Inversely, the areas containing higher numbers of active groups are more likely to have considerable epidemic outbreaks. The prevalence of the SIR epidemic outbreaks corresponds to the spatial location of groups.

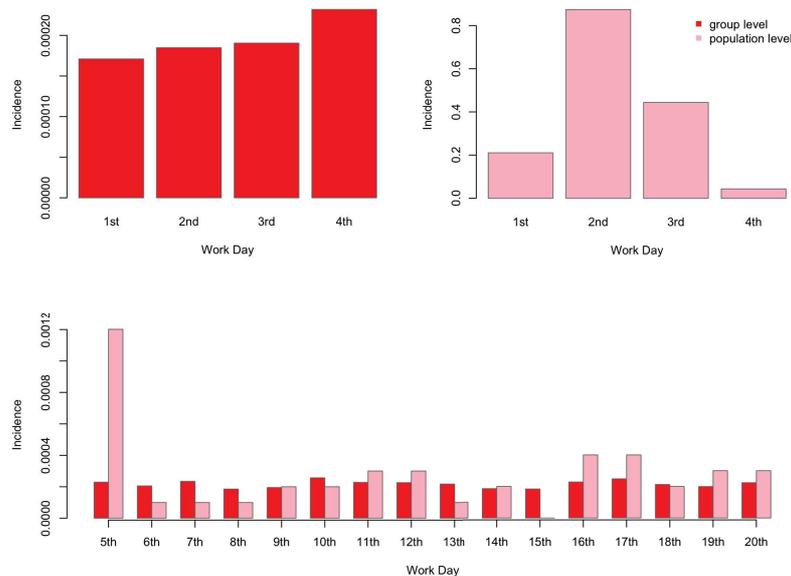
The question raised here is : *what would change if a given population was considered without any group structures ?* or said differently, *Would the SIR epidemic dynamics be remarkably different?*. To test this, we used the same SIR epidemic model with the same parameters ( $\beta = 0.003 \text{ km}^2 \cdot \text{indiv}^{-1} \cdot \text{day}^{-1}$ ,  $\gamma = 0.08 \text{ day}^{-1}$  and  $\lambda = 0.04 \text{ day}^{-1}$ ) on the simulated population without forming groups. Interactions between agents were only temporary (1 simulation step) and limited to the close neighbourhood surrounding the agents (within the interaction radius). The results of these two simulations are shown in figure 3.8.



**Figure 3.8 SIR Epidemic model with & without Groups Formation:** Two different SIR dynamics are obtained by simulating them in a population with group formation and in another population without any group formation. The simulated population density, in both cases, is 15,625 inhabitants per square kilometre. The common SIR model parameters are  $\beta = 0.003 \text{ km}^2 \cdot \text{indiv}^{-1} \cdot \text{day}^{-1}$ ,  $\gamma = 0.08 \text{ day}^{-1}$  and  $\lambda = 0.04 \text{ day}^{-1}$ . Both SIR dynamics are shown at the population level. When group formation is allowed (Top), the SIR process shows very slow dynamics, notably due to the very small proportion of active groups (in which there are infected individuals). On the contrary, when group formation is not allowed (Bottom) the infection peak is reached after only 2 days, due to a faster mixing of individuals, in all likelihood; free agents move indeed permanently and mix fastly compared to individuals belonging to groups. This however illustrates the strong effect of the presence of a group structure or not.

The MAS-G model also provides fine-grain information in the form of a microscopical resolution of time, space and a description of populations at the level of groups and single individuals. The possibility to observe epidemic events at both macroscopic (population) and microscopic (individual) levels gives access to information that is important to understand

epidemic dynamics. An example of such information is the incidence at the group level. Incidence is defined as the probability to develop a particular disease during a given period of time ( $\#$  of new cases during  $[t_1, t_2]$  / population at risk during  $[t_1, t_2]$ ). Incidences in populations with or without groups are shown in figure 3.9. This shows again the impact of groups on the SIR dynamics. The incidence at the population level (with no group formation) is – almost – always higher than at group level during all the 20 simulated days and this is more remarkable during the first four days. With the parameters used here, the infection does not develop in our group-based MAS simulation, but different parameters could be used that would produce results similar to those obtained in SSA-G simulations.



**Figure 3.9 Daily Incidences at Group & Population Levels:** The number of daily new cases are compared for two simulations, with (and without) group formation in red (in pink respectively).

Comparing the results of daily incidence between these two simulations confirms the importance of considering the notion of groups in the population. The study of an epidemic

process at the population level (without group formation) might cause an over-estimation of epidemic intensity. Groups play an important role in disease's spread in two ways: first, at the microscopic level within groups (e.g. the spread of a flu epidemic in a family of 5 member is not the same as in a company that consists of 10 to 30 employees); second, at the macroscopic level, diseases diffuse from groups to the global population, transforming micro-events (within groups) into big, catastrophic macro-events, when many groups are present in a population.

Data obtained from MAS-G -based simulations could also be analysed at other levels, notably at the individual level by tracking, individual after individual, the paths that the disease follows. One could provide a connection network map to understand who infects who or even investigate who is more susceptible to the infection, particularly to know whether they are individuals with high group tendencies or those who stay for a long time in a group. Here, we aimed to show the important role of group-structured populations in epidemiological studies but further developments are needed to study the evoked epidemic pathways, or even to inject more realistic social behaviours to our agents (like "intelligent" bots in video games). These sorts of analyses could be done in the context of participatory simulations (serious games) in which part or all the individuals are played by real humans.

# Chapter 4

## Discussion and Perspective

### 4.1 General discussion

By using different mathematical models (modified SIR dynamics with or without resistance) and different numerical implementations (ODE, homogeneous SSA, group-based SSA, group-based MAS), we improved the role of  $R$  individuals, that are as important as the  $I$  individuals, to produce and understand an epidemic process. Resistant individuals do not only introduce delays in the SIR dynamics; their presence, as obstacles inserted within groups with other types of individuals ( $S$  and  $I$ ), strongly influences the SIR dynamics. In communicable diseases the infection process (spreading) depends on the "pool of  $S$  individuals" susceptible to be in contact with  $I$  individuals. When more  $S$  individuals are present and close to  $I$  individuals, the latter are more likely to meet  $S$  people during their infectious period. Then,  $R$  individuals influence the infection process in two different ways: first, they introduce a latency (see [111]) by forming a pool of immune people delaying the reconstitution of the

$S$  reservoir; second, they dilute the population and block a disease's spreading at the same time. This last point is not taken into account in SIR dynamics computed by ODE or PDE. As we introduced this notion of group-complexes in SIR kinetic models, it has allowed us to study how the  $R$  individuals can play an inhibitory role against the spread of disease. In this way, we integrated local spatial structures in homogenous populations, whose dynamics are simulated by deterministic ODE or by the corresponding stochastic simulation algorithms.

We introduced social and local spatial structures, like groups, in SSA-G and MAS-G numerical models, but also in ODE simulations with modified SIR kinetics that include the notion of group-complexes, and in SSA-CA in which the effect of local grouping and resistance are intrinsic to the implementation (local neighbourhood on a grid). Although these implementations are very different, the dynamics obtained for a given epidemic process look very similar, as shown in figure 4.1 below. However, the understanding of an epidemic process, from local and individual events to global observations, depends on the simulation scale, and on the related observation scale, that give access to the different mechanisms involved at different levels (*e.g.* individual, groups or population, but also relational networks).

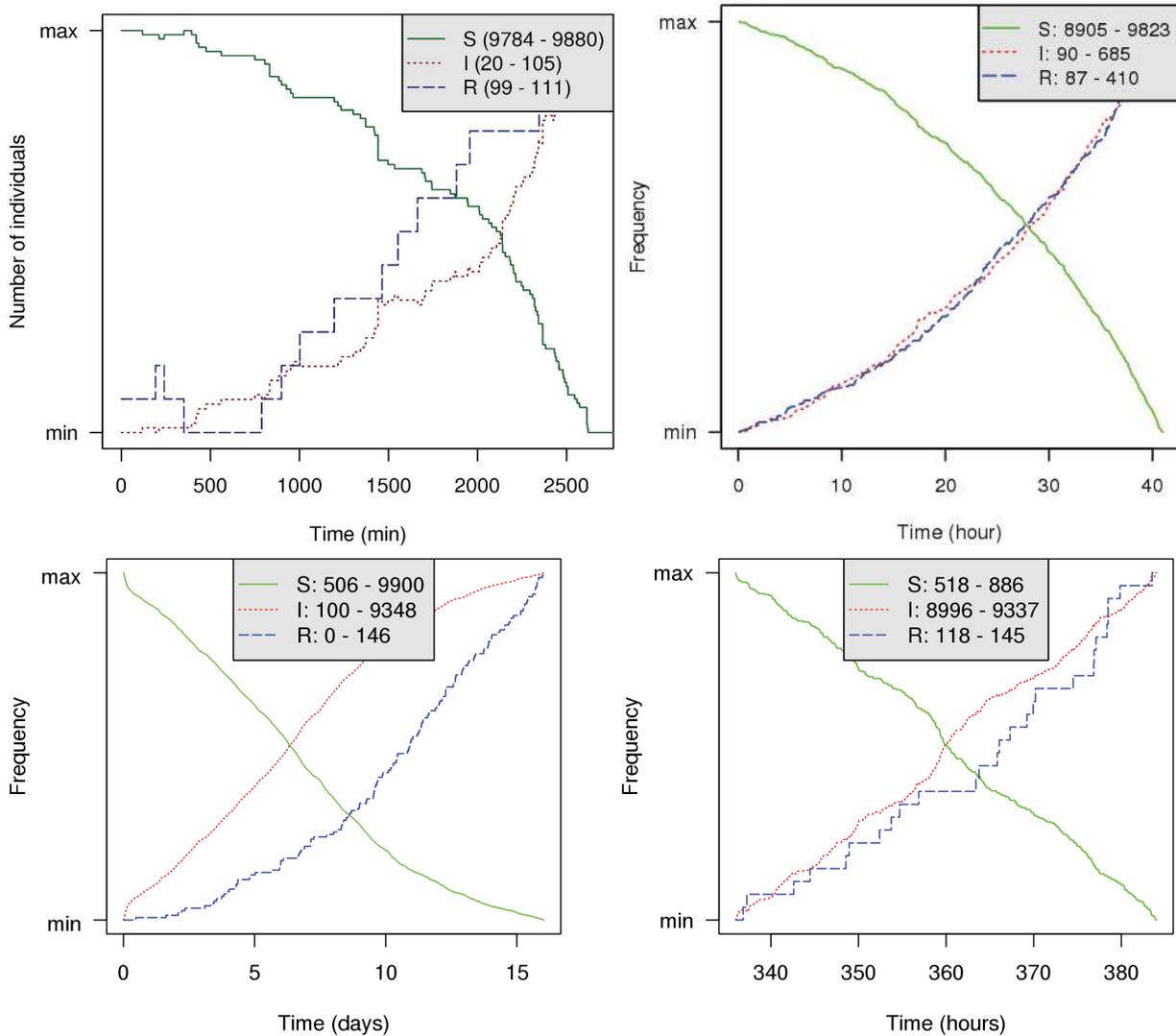


Figure 4.1 **The same epidemic seen by three different models** : In all these simulations, the parameters are the same as those in figure 3.3 given for a population of 10,000 inhabitants with a density of  $8700 \text{ km}^{-2}$ ). The results obtained by a stochastic simulation with group-formation (SSA-G: Top Left), a stochastic cellular automaton (SSA-CA: Top Right), and a multi-agent based simulation with group-formation (at two different observation scales: 15 days and 2 days) (MAS-G: Bottom), show very similar behaviours. However, they do not provide the same amount of information.

We also showed that the local structure in space and the composition of a group of individuals influence the infection dynamics due to individual and local parameters that can be taken into account in all our models, like individuals' speeds or local population

density (related to the effective surface). For example, simulating an epidemic by using a MAS-G models depends on which context the population dynamics are considered: low density (people in the street), more densely populated (colleagues in an office or a couple of friends in a cafeteria), or extremely densely populated (a crowd in a concert in which any movement is not easy, or several people in a lift). Then to take correctly into account the different characteristics of epidemics, to simulate all such situations at the same time, different appropriate models working at different levels (and having different observation scales) must be used. For this reason, we propose the use of hybrid models as appropriate combinations of different modelling methods. We showed that the different models presented in this thesis can use the same sets of parameters and that in these conditions, they can show the same behaviours. In this way, our models can be connected together to simulate different regions of space, or different moments (*e.g.* nights *vs* days), with more or less precision. For example, one could simulate the evolution of a disease (*e.g.* gastroenteritis) in France by a PDE using our modified SIR dynamics on a square grid, to simulate the same dynamics by a SSA-G within a small part of this grid corresponding to Grenoble city (*e.g.* which could be equivalent to only one cell of the PDE grid) so as to increase the precision of SIR dynamics in an area surrounding the region of interest, and finally, to simulate within this city another very limited region, the region of interest, *e.g.* the TIMC-IMAG laboratory, by a MAS-G numerical model.

Of course, we can not expect to "save the world" by such "toy models" as is often suggested by mathematicians looking for research funds. We did not investigate on whether our models are better adapted to explain an infection process or another, or to use more population classes

than the *S*, *I* and *R* ones (*e.g.* like children, pregnant women, healthy carriers, *etc.*), or even to take into account geographic and architectural features like reliefs, transports, city and building structures, or finally to include realistic individual behaviours (social interactions, moves, rhythms, *etc.*). However, we think that our models constitute a good basis to develop prevention tools and that they could be used in public health policies in the way they could guide (i) the structuration in time and space of the populations and their environments in order to reduce the epidemiological risks, (ii) the promotion of health behavioural patterns and (iii) to improve attitudes towards hygiene.

In an attempt to provide models with more realistic aspects, we also suggest applying complementary approaches like participatory simulations, the use of smart devices, social networks and the big-data, to enrich our hybrid models with features that are not easy to simulate.

## 4.2 Complementary Approaches

Epidemiological studies based on modelling should benefit from all existing and new tools or technologies to improve health care and social knowledge at both individual and public levels. Complementary methods can enrich epidemiological models to understand and simulate the complex behaviour of epidemic processes strongly linked to socio-cultural factors. Participatory simulation approaches, knowledge on the structure of and data from

social networks and big-data may help to put together information that seemed inaccessible until recently. This collection of information may lead to more reliable models.

#### 4.2.1 Participatory Simulations and video games

According to *Vygotsky* and related theoreticians, knowledge is social, constructed from cooperative efforts to learn, understand, and solve problems. The behavioural learning theory perspective focuses on the impact of group reinforcement and rewards on learning. Epidemiological modelling relies on methods to learn and understand the propagation process of diseases. They are expected to be used to prevent and even stop the epidemic outbreaks by promoting prevention strategies. However when we deal with socio-cultural impacts on epidemic dynamics, individual behaviour has a significant role. This role can be taken into account and understood by using multi-agent systems in which individuals' decisions are made by probabilistic based procedures or even in a more effective way, by using cooperative methods in simulations. One can effectively use participatory simulations (serious games) to model an epidemic process by considering in the same simulation the behaviours (interactions, moves, *etc.*) driven by real human agents, and others obtained by simulation procedures. Participatory simulations are then new tools to bridge between – human – behavioural sciences and computer simulations.

Another manner to include better social behaviours is to study how players behave in massive multiplayer online role playing games (MMORPG) like *Second Life*, Blizzard's *World of Warcraft*, *etc.*, in which socially evolved structures developed. Particular simulations are a simplified version of MMORPG.

### 4.2.2 Smart Device Applications, Social Network and Big-Data

Nowadays, smart devices (*e.g.* tablets, smartphones, smartwatches) are changing human life styles. Smartphones are the most common devices that are used by more than 2.23 billion people worldwide, estimated as 96.8% of the population while 47.2% of these also use mobile internet <sup>1</sup> in 2015. These statistics allow us to get a glimpse of how smartphones could play an important role in modelling, collecting and analysing health data: they allow geolocation, one can know a part of the spatio-temporal interaction network of an individual, but they also bring different aspects of individual healthcare domains, notably via health and fitness tracking software.

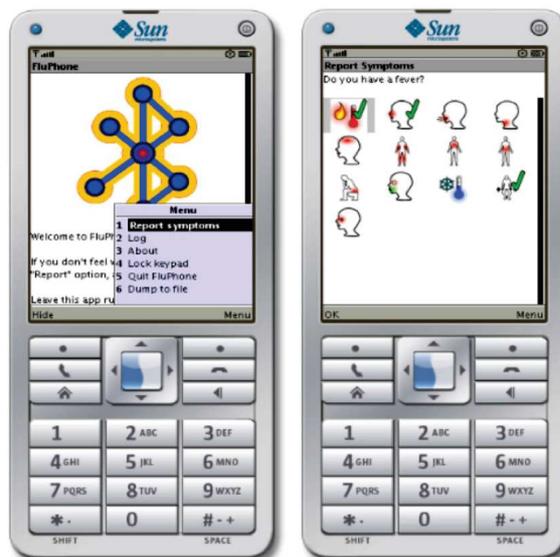


Figure 4.2 **FluPhone Project**. In this smartphone app, individuals can declare their state of health by selecting predefined pictograms showing different states of health (headache, cold, cough, *etc.*). The app yields this data while geolocating the individual. Data recorded from many users allows the tracking of diseases and constructs infection maps.

<sup>1</sup>International Telecommunication Union Statistics

As an example (see figure 4.7), in an epidemiological research project (FluPhone Project of the Cambridge university), mobile phones were used to understand behavioural responses to Flu outbreaks (also refer to Peruni & Tabourier, Plos One 2011). Volunteers were asked to download a dedicated application on their smartphone (Fluphone). The application anonymously recorded the encounters between people and also gave the users the possibility to declare their state of health through a system of pictograms (showing a healthy person, headache, abdominal pain, *etc*). Then, diseases could be tracked and infection maps understood. Such information could be at least compared with the results of models, but also potentially used as predefined behavioural patterns in epidemics models.

There are also other examples based on smartphone applications for mapping and tracking disease's spread and outbreaks. The use of social networks in virtual environments is increasing, particularly among the younger generation. Social networks, such as *Facebook*, are becoming valuable source of information about users' social behaviours.

<b>Social Network statistics</b>	<b>Data</b>
Do you ever use / have a profile on any social network	58% Yes
Total number of Facebook users worldwide	1.4 billion
Total percentage of 18-24 year olds who already use social media	98%
Average amount of tweets per day	190 million
Percent of teenagers who log on to Facebook over 10 times per day	22%
Percent of 18-34 year olds who check Facebook when they wake up	48%
Total pieces of content shared on Facebook each month	70 billion
Total amount of people who access Facebook with phone	250 million

Figure 4.3 **Social Network Data** Source: Browser Media, Socialnomics, MacWorld

The study of social network connections and all other related information obtained in *The Bigdata*, can lead to an understanding of the individual behavioural aspects in a given

population facing an epidemic outbreak. Such data provides information about people's lifestyles, their contacts, their frequented places and their trips. This information is now fairly easy to access by data-mining approaches from mobile smartphones, social networks, online newsletters, online consumer preferences (including that about travel, food, shops,*etc.*). Such data can be used to develop pre-topological maps (maps of consumers, workers,*etc.*). They provide a set of information about groups structures, relationship patterns, connections, interdependency, *etc.*. Such information might be used as complementary informative indicators to detect early stage disease outbreaks in real time and by actual geographical positions. For example, starting with information collected from *Facebook* data (see figure 4.4 below), one can obtain information about different groups of friends, their connections and where they live. This can go further with more detail to analyse the social network's statuses that are mentioned by users about specific subjects (*e.g.* common cold, influenza) as was done in [14] with *Twitter* accounts. The personal information published voluntarily by users on the internet could be viewed as a useful source of information to the benefit of public health systems. For example, in *Facebook*, the most popular users (with more followers) could be vectors to promote prevention strategies such as vaccination campaigns.

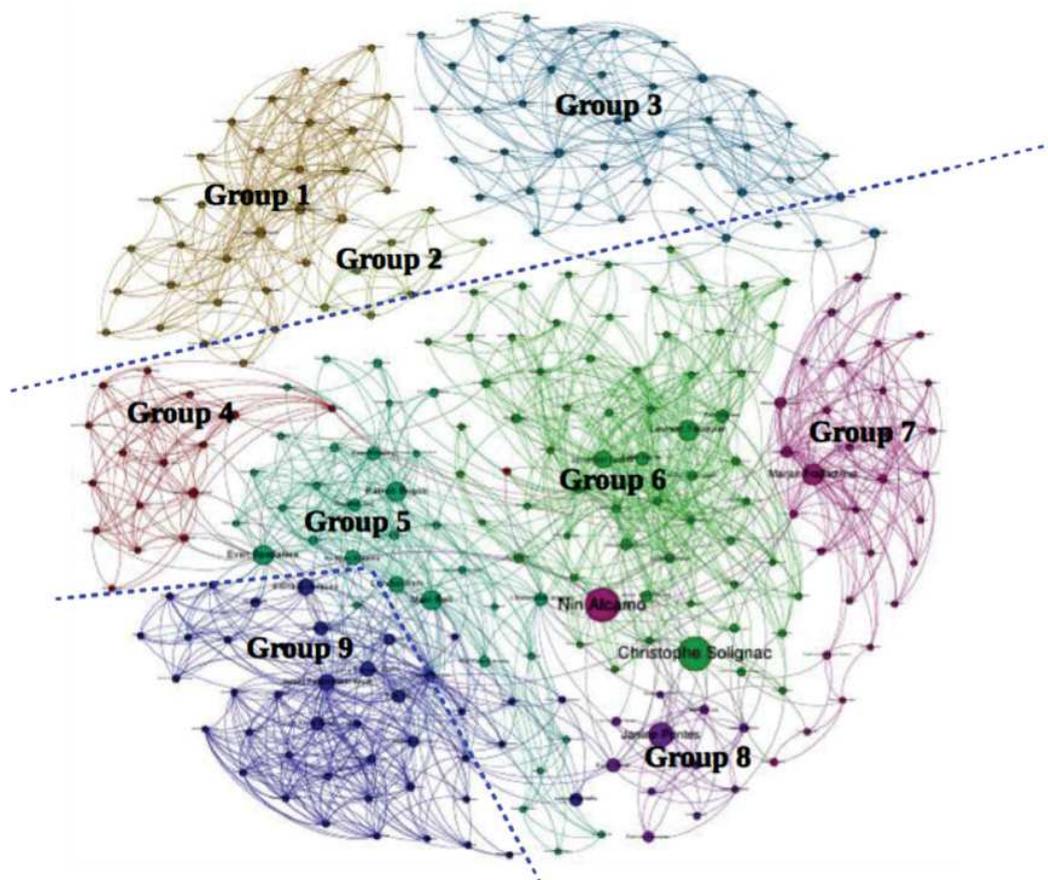


Figure 4.4 **FaceBook Data:** The author's Facebook data is analysed to detect different groups of friends, their connections and the current geographical positions (*i.e.*, all the places they visit). This information is declared on Facebook by personal users. On this graphic, three locations are specified as Iran, France and the United State of America, separated by dotted lines. Different groups of friends (9 groups) are detected (*e.g.* group #5 is known as a group of Brazilian friends who live in France). Each user is displayed by a single dot whose size depends on the number of the user's friends on Facebook. Connections between friends (users) are shown by arc lines.

Moreover, there already exists specific health applications for smartphones to follow disease outbreak. They are based on the collecting of health-related information from media sources on the internet (*e.g.* Twitter accounts, newsletters or publicly available sources by national and international public health organisations <sup>2</sup>).

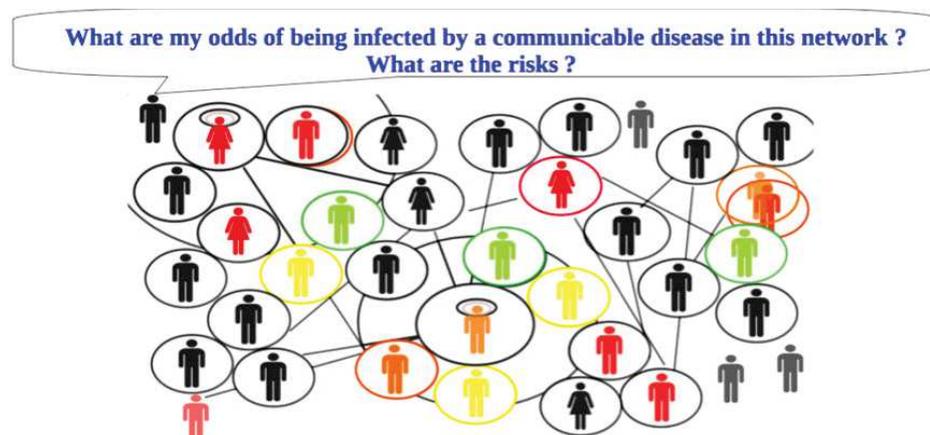
<sup>2</sup>Outbreaks Near Me: a mobile application from the HealthMap website founded by a team of researchers, epidemiologists and software developers at Boston Children's Hospital

In fact only having access to this level of details can not lead to a perfect disease spread monitoring system, but a combination of this complementary information with numerical simulations may produce more accurate hybrid epidemiological models. Future work should be based on these kinds of hybrid data-driven and predictive numerical models.

### 4.3 Possible Future Works

Monitoring the health dynamic of a given population is an essential way to identify the occurrence of epidemic events at an early stage and to control the spread of disease. The mathematical models used in epidemiology are not the same as their analogues in meteorology, whose analytical results are used frequently at the individual level by everyone as well as at the population level by national or international organisations. Although meteorological and epidemiological fields are two different worlds, they aim to study the dynamic behaviour of a dynamical process (the weather or an epidemic) in real time and in exact geographical locations. The weather forecasting models, beside their mathematical nature, are based on observations that come from different numerous information sources such as weather satellites, radiosondes, *etc.* These complementary observations help to enrich numerical predictive results and make the meteorological models the most accurate human-made models. In epidemiological studies, since we don't have much complementary information at the individual level on related factors in disease spread such as sociocultural patterns, individual connection networks and *etc.*, and since human (animal) behaviours are very complex com-

pared to atmospheric flows, one can not practically predict the chances of an individual (or a sub-population) being infected in a given population by a given communicable disease.



At the individual level, in order to predict the significant risks and estimate the probabilities of being infected in real time in actual locations, one can not limit the modelling to models at the population level (*e.g.* ODE and even PDE, and their corresponding homogeneous SSA-based models). Therefore, we propose epidemiological modelling approaches that are compatible with the individual level and that can also respond to other constraints existing at the population level. This might be possible by using multi-modal data approaches that consist of bottom-up data processing from the individual level to populations and the reciprocal (top-down).

An epidemiological hybrid model might consider a combination of different systems that collect information at individual, groups and population levels. The information at each level is used to improve the modelling methods at the other levels. For example, data collected from personal mobile data can be used, such as pre-topological information about spatial lo-

cations of individuals, their replacement positions and their connection networks to introduce spatiotemporal structures into models at the population level.

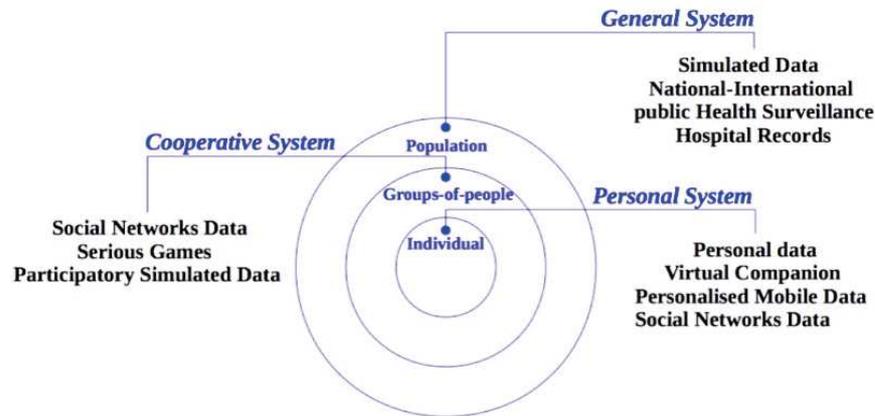


Figure 4.5 Multi-Modal Data Approaches

Analysing the information about epidemic outbreaks from several sources (based on different epidemic models and data-collecting systems) leads to an enrichment in public health surveillance at different levels. It is important then to provide models that can furnish measurements of epidemic processes from the microscopic (individual) level to the macroscopic (population) level. Models offer an anonymous data bank of outbreak tracks and records from single individuals to groups of people in a population, a way to determine who is infected by whom, when and where, as it's possible in spatial multi-agent based models (*e.g.* MAS-G) or spatial or structured stochastic based numerical models (*e.g.* SSA-CA and SSA-G).

It is time now for epidemiological models to serve as tools for providing fine-grain health awareness (at both microscopic and macroscopic levels) to inform and predict the potential epidemic risks in real-time and in actual positions. Models should be tools that help public

health policies to be accompanied by ongoing epidemic control strategies. This would be possible by means of hybrid models enriched by multidimensional information from an individual to groups of people or the global population. Such hybrid models could be used according two main axes: first to organise human environments (and rhythms) in order to reduce the economic impact of epidemic diseases, and second to personalise disease monitoring in health-care systems.

- **Hybrid Models to Organise Human Environment**

By considering a hybrid models, one can assess how a human environment arrangement can affect an epidemic process. Very simply speaking, there are three different human environments: living, working and learning environments. The dynamics of an epidemic would behave differently in each of those environments: for example, the dynamics of gastroenteritis in an elementary school is not the same as in a high school, notably because children are asked to clean their hands frequently at the elementary school. Then we must use the disease surveillance systems depending on different contexts.

Hybrid models proposed here, include different modelling approaches that are adapted to different contexts (*e.g.* educational institutions such as schools, administrations, universities) at different levels, for example a sophisticated multi-agent models (MAS-G) to model working or learning environments with accuracy by considering pre-topological information sources like agents' characteristics (*e.g.* working hours, meeting durations, lunch break, *etc.*) and a cellular automaton based modelling (SSA-AC) for dense environments in which individual's movements are reduced (*e.g.* the population of a crowded metro, school canteens,

*etc.*), or on the contrary by using more global modelling methods (like the SSA-G numerical model) to model large populations. These models can be developed by using complementary information (*e.g.*, achieved by big-data).

We propose here to use such hybrid models to simulate the impacts on epidemic dynamics caused by architectural adjustments in human environments, like the arrangement of offices, corridors or toilets, the number of washbasins, *etc.*. They can also be used to assess the impact of the health awareness campaigns for epidemic prevention like for the models used in socio-political studies to understand the social impacts on political opinions, see [34–39]. These models can be considered as advice tools for institutions (*e.g.*, schools, administrations, universities), public organisations or private companies to propose useful changes in human rhythms (including behavioural patterns) and human environments in order to reduce the impact of epidemic outbreaks. For example, to reorganise school or work environments with new sanitary and hygiene arrangements or to assess the prevention strategies to a specific targeted population for improving hygiene practices or changing hygiene behaviour. The figure below illustrates how such different models can be combined to model different contexts.

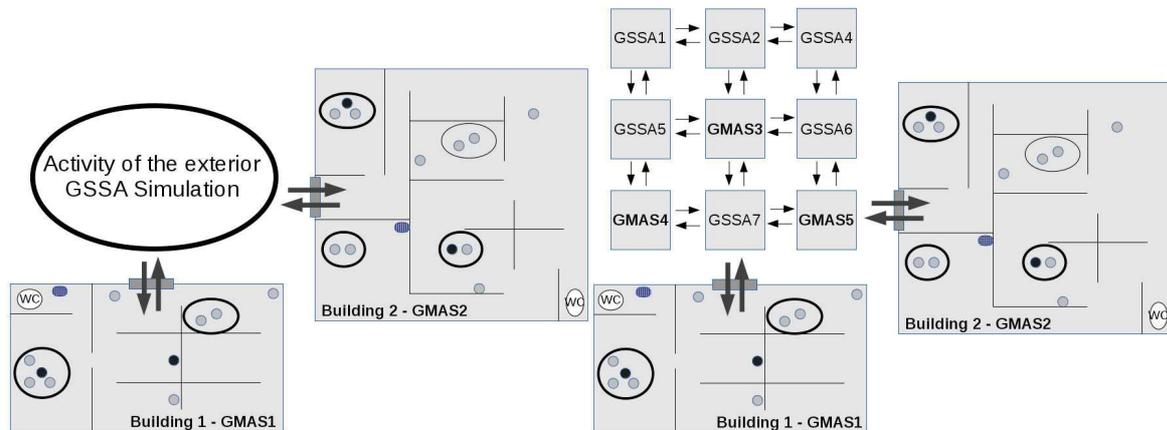


Figure 4.6 Hybrid Models to Organise the Human Environment

• **Personal Disease Surveillance Assistant**

The second potential application of our models is to elaborate a personalised disease surveillance that provides a new health-care surveillance method, meaning disease surveillance by everyone.



Figure 4.7 Disease Surveillance System

As mentioned before about the importance of epidemiological modelling approaches at different levels, we also believe that disease surveillance systems (figure 4.7) must be considered at different levels, especially at the individual level, notably as an improvement in individual responsibility in population health. It has to be noted that the idea here is not to predict the occurrence of an epidemic disease for a specific individual, but to bring all information related to a given individual, such as the record of recent and actual locations, claimed health status, social networks status and information, *etc.*, to personalise surveillance in health-care systems. This could lead to a win-win public healthcare system as everyone could benefit from its advantages. Of course, this raises the question about private data and ethics.

At the population level, as said before, one could provide more accurate epidemic models to study the population health dynamic (for governmental purposes, *e.g.* to target populations concerned by new vaccination policies), at a smaller scale (enterprises or educational institutions levels), one could reduce (optimise) the costs or lack of individual productivity due to a communicable disease at work (office) or in a learning environments (school). Moreover, at the individual level, each person could be aware of the local possible epidemic outbreaks in specific (personalised) geographic locations in real time, of the potential risks to that anyone is exposed. By receiving preventative and health awareness alerts, such a health surveillance systems would become more personalised and specific. Future work implying strongly modern social science studies have now to be coupled to epidemiology to combine these two fields efficiently. During this Ph.D. thesis, we contributed by preparing the ground

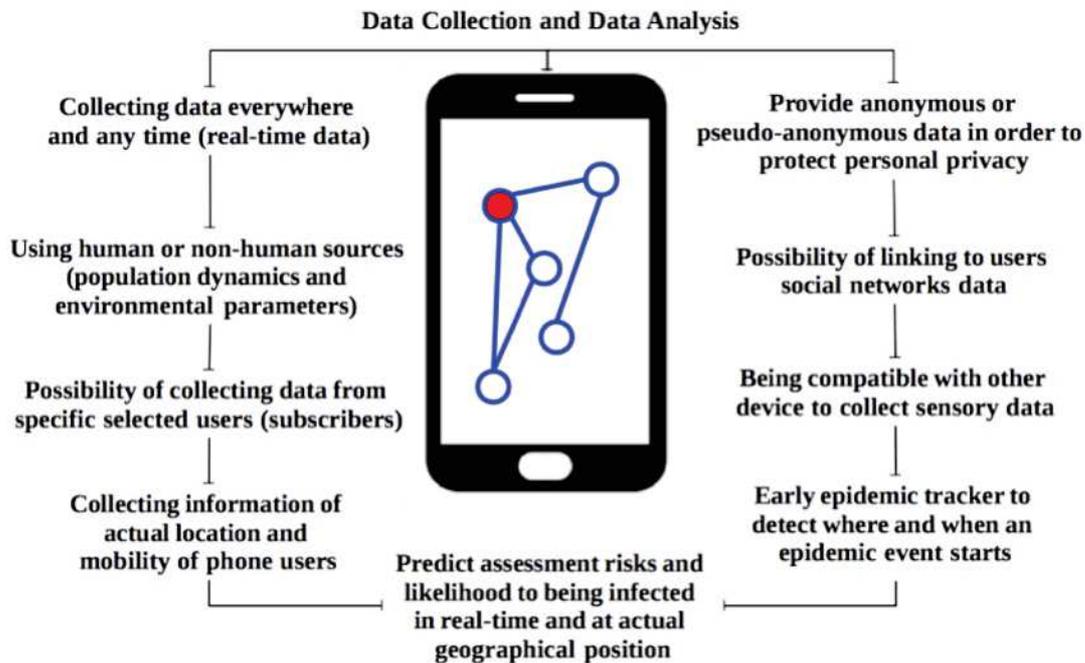


Figure 4.8 Personal Diseases Surveillance Assistant Applications such as Virtual Companion for coupling epidemiological and population dynamics, but a strong effort with many others has to follow so as to obtain the efficient tools we described above.

# References

- [1] An introduction to stochastic epidemic models, L. Allen. [ *Mathematical Epidemiology Springer, (2008)*,p. 81-130]
  
- [2] Etude comparée de quelques méthodes de Simulation de Reactions Biochimiques, P. Amar, [ *Le vivant discret et continu Editions Materiologiques, Chap. 10 (2013)* p. 267-300]
  
- [3] The role of mathematical models in the study of HIV transmission and the epidemiology of AIDS, R.M. Anderson [ *AIDS 1(1988)*, p. 241-246]
  
- [4] Mathematical Models of Transmission and Control of Sexually Transmitted Diseases, Roy M. Anderson, Geoffery P. Garnett [ *Mathematical Models Vol. 27 (2000) No. 10.* ]
  
- [5] Scaling of sexual activity, R.M. Anderson and R.M. May [ *Nature 333(1989)* pp. 514-519]
  
- [6] Quarantine in a multi-species epidemic model with spatial dynamics, Julien Arino, Richard Jordan, P. van den Driessche [ *Mathematical Biosciences 206 (2007)* p. 46-60]

- 
- [7] Emergence of Scaling in Random Networks, A.L. Barabasi and R. Albert [ *Science*, Vol. 286 no. 5439 (1999 ) pp. 509-512.]
- [8] When individual behaviour matters: homogeneous and network models in epidemiology, S. Bansal, B. T. Grenfell, L. A. Meyers [ *The Royal Society Interface* (2007)]
- [9] How Nature works The science of self-organized criticality, Per Bak [ *Springer-Verlag* (1996), New York.]
- [10] The mathematical theory of infectious diseases and its applications, N.T.J. Bailey [ *Griffin* (1975), London]
- [11] A clarification of transmission terms in host-micro parasite models: numbers, densities and areas, M. Begon, M. Bennett, R.G. Bowers, N.French, S.Hazel [ *Epidemiology and Infection*, Vol. 129 (2002) p. 147-153]
- [12] Essai d'une nouvelle analyse de la mortalité causée par la petite vérole, D. Bernouilly [ *Mém. Math. Phys. Acad. Roy. Sci. Paris* (1766)]
- [13] Economic epidemiology of avian influenza on smallholder-poultry farms, M.F. Boni, A.P. Galvani, A.L. Wickelgren, A. Malani [ *Theoretical Population Biology* 90 (2013).p.135-144]
- [14] National and Local Influenza Surveillance through Twitter: An Analysis of the 2012-2013 Influenza Epidemic, D.A. Broniatowski, M.J. Paul, M. Dredze [ *PLOS Currents: Outbreaks*, 8(12) (2013).]

- 
- [15] Scaling in animal group-size distributions, E. Bonneau, L. Dagorn, P. Freon [*Proc. Natl. Acad. Sci. Ecology, Physics, Vol. 96 (1999)*, p. 4472-4477]
- [16] Modelli deterministici e stocastici in epidemiologia, Bollettino del centro per la ricerca operativa, F. Brambilla [*Serie Metodologica 4 (1960)*, p. 3]
- [17] Age-of-infection and the final size relation, F. Brauer [*Mathematical biosciences and engineering Vol.5 n.4 (2008)*, p.681-690]
- [18] Stochastic epidemic models: a survey, T. Britton [*Mathematical Biosciences Vol. 225, Is.1, (2010)*, p. 24-35]
- [19] Community Surveillance of Respiratory Viruses Among Families in the Utah Better Identification of Germs-Longitudinal Viral Epidemiology (BIG-LoVE) Study, C. L. Byington, C. Stockmann, F.R. Adler [*Clinical Infectious Diseases N 61(9) (2015)*]
- [20] Internets Dirty Secret: Assessing the Impact of Online Intermediaries on HIV Transmission, J. Chan, A. Ghose [*social science research network (ssrn) MIS Quarterly, 38(4) (2013)*, p. 955-976]
- [21] Mixing patterns between age groups in social networks, S.Y. Del Valle, J.M. Hymanb, H.W. Hethcote, S.G. Eubank [*Social Networks 29 (2007)* pp.539-554]
- [22] Daniel Bernoulli's epidemiological model revisited, Klaus Dietz and J.A.P. Heesterbeek [*Mathematical Biosciences 180(2002)*, p.1-21]
- [23] Epidemics and Rumours, Klaus Dietz [*Journal of the Royal Statistical Society 130(1967)*, p.505-528]

- [24] Bernoulli was ahead of modern epidemiology, Klaus Dietz and J.A.P. Heesterbeek  
[*Nature* 408(2000),p. 513]
- [25] Global asymptotic properties of an SEIRS model with multiple infectious stages,  
Dessalegn Y. Melesse and Abba B. Gumel [*Journal of Mathematical Analysis and  
Applications* 366 (2010) pp. 202-217]
- [26] An Agent-Based Spatially Explicit Epidemiological Model in MASON, J.B. Dunham  
[*Journal of Artificial Societies and Social Simulation*, Vol. 9 (2005)]
- [27] Dynamical resonance can account for seasonality of influenza epidemics, J. Dushoff,  
J.B. Plotkin [*PNAS*, Vol. 101 no. 48 (2004)]
- [28] The landscape epidemiology of foot-and-mouth disease in South Africa: A spatially  
explicit multi-agent simulation, E. Dion, L. VanSchalkwyk, E.F. Lambin [*Ecological  
Modelling*, Vol. 222, Issue 13 (2011)p. 2059- 2072]
- [29] Epidemiology of Transmissible Diseases after Elimination, G. De Serres, N. Gay,C.  
Paddy Farrington [*American Journal of Epidemiology* V.151,N.11 (2000)]
- [30] Mesoscopic reaction-diffusion in intracellular signaling, Johan Elf, Andreas Doncic,  
Mans Ehrenberg [*Fluctuations and Noise in Biological, Biophysical, and Biomedical  
Systems Proceedings of SPIE* 5110 (2003)]
- [31] The Meta-population Dynamics of an Infectious Disease: Tuberculosis in Possums , G.  
R. Fulford, M. G. Roberts [*Theoretical Population Biology* 61 (2002) p. 15-29]

- 
- [32] Minority opinion spreading in random geometry, S. Galam [*The European Physical Journal B* 25 (2002), p. 403-406]
- [33] Terrorisme et percolation, S. Galam [*La Science N* 306 (2003)]
- [34] Sociophysics: A new approach of sociological collective behaviour,behaviour description of a strike, S. Galam, Y. Gefen, Y. Shapir [*The Journal of Mathematical Sociology* 9 (1982), p . 1-13]
- [35] Heterogeneous beliefs, segregation, and extremism in the making of public opinions, S. Galam [*Phys. Rev. E* (2005) 71, 046123]
- [36] Sociophysics : A review of Galam Models, S. Galam [*International Journal of Modern PhysicsC, V.19* (2008), p. 409-440]
- [37] Collective beliefs versus individual inflexibility: The unavoidable biases of a public debate, S. Galam [*Physica A* 390 (2011), p. 3036-3054]
- [38] Building up of individual inflexibility in opinion dynamics, A. C. R. Martins, S. Galam [*Phys. Rev. E* (2013) 87, 042807]
- [39] From Public Outrage to the Burst of Public Violence: An Epidemic-Like Model, S. Nizamani, N. Memon , S. Galam [(2013) *arXiv:1310.0731*]
- [40] On the Critical Behavior of the General Epidemic Process and Dynamical percolation, P. Grassberger [*Mathematical Biosciences* 63 : 157-172 (1983)]

- [41] An Analysis of the Belousov-Zhabotinskii Reaction, Casey Gray, Calhoun HS, Port Lavaca [*undergraduate Math Journal Vol. 3 Issue 1(2002)*]
- [42] A General Method for Numerically simulating the Stochastic Time Evolution of Coupled Chemical Reactions, Daniel T. Gillespie [*Journal of computational Physics Vol. 22 (1976)*, p. 403-434]
- [43] Infection, reinfection, and vaccination under suboptimal immune protection : epidemiological perspectives, M. Gabriela M. Gomes, Lisa J. White, Graham F. Medley [*Journal of Theoretical Biology (2004)*, p. 1-11]
- [44] The reinfection threshold promotes variability in tuberculosis epidemiology and vaccine efficacy, M. Gabriela M. Gomes, Ana O. Franco, Manuel C. Gomes, Graham F. Medley [*The Royal Society 271 (2004)*, p. 617-623]
- [45] Epidemic disease in England The Evidence of Variability and of Persistency of Type, W.H. Hamer [*Lancet, 1 (1906)*, p. 733-739]
- [46] Modeling HIV Transmission and AIDS in the United States, Herbert W. Hethcote and James W. Van Ark [*Springer-Verlag (1992)*, New York]
- [47] The Mathematics of Infectious Diseases, Herbert W. Hethcote [*Society for Industrial and Applied Mathematics Vol. 42, No. 4 (2000)*, p. 599-653]
- [48] Qualitative Analyses of Communicable Disease Models, Herbert W. Hethcote [*Mathematical Biosciences 28 (1976)* p. 335-356]

- [49] Three Basic Epidemiological Models, Herbert W. Hethcote [*Applied Mathematical Ecology* L. Gross, T. G. Hallam and S. A. Levin, *Biomathematics Vol. 18 (1989)* p. 119-144]
- [50] Random modelling of contagious diseases, J. Demongeot, O. Hansen, H. Hessami, A.S. Jannot, J. Mintsu, M. Rachdi, C. Taramasco [*Acta Biotheor. (2013)*]
- [51] The Theoretical Population-Level Impact of a Prophylactic Human Papilloma Virus Vaccine, James P. Hughes, Geoff P. Garnett, Laura Koutsky [*EPIDEMIOLOGY Vol. 13 No. 6 (2002)*]
- [52] The application of simulation models and systems analysis in epidemiology, H.Scott Hurd and John B.Kaneene [*Preventive Veterinary Medicine, 15 (1993)* p. 81-99]
- [53] Timothy Perry, Modeling and Analyzing HIV Transmission: The Effect of Contact patterns, John A. Jacquez, Carl P. Simon, James Koopman, Lisa Sattenspiel [*Mathematical Biosciences 92 (1988)* p. 119-199]
- [54] Reproduction Numbers and Thresholds in Stochastic Epidemic Models, John A. Jacquez and Philp O'neill [*Mathematical Biosciences 107 (1991)* p. 161-186]
- [55] Measles Outbreaks in a Population with Declining Vaccine Uptake, V.Jansen, N. Stollenwerk, H. Jensen, M.Ramsay, W. Edmunds, C. Rhodes [*Science V.301 (2003)*]
- [56] Spatio-Temporal self-organized in a model of disease spreading, A. Johansen [*Physica D 78 (1994)*]

- [57] Modeling Infectious Diseases in Humans and Animals, M. Keeling, P. Rohani  
[*Princeton University Press. (2007)*]
- [58] Stochastic Dynamics and a Power Law for Measles Variability, M. Keeling, B. Grenfell  
[*The Royal Society (1999)*]
- [59] Simple Stochastic Models and Their Power-Law Type Behaviour, M. Keeling  
[*Theoretical Population Biology, Vol. 58 (2000) p. 21-31*]
- [60] The effects of local spatial structure on epidemiological invasions, M. Keeling [The  
*Royal Society 266 (1999) p. 859-867*]
- [61] A Contribution to Mathematical Theory of Epidemics, W. O. Kermack and A. G. McKendrick [The  
*Royal Society of London, Series A Containing Papers of a Mathematical and Physical Character, Vol. 115, No. 772 (1927) p. 700-721*]
- [62] Invariances d'échelle, Michel Lagues Annick Lesne [Belin (2003)]
- [63] Early Stage Influenza Detection from Twitter, J. Li, C. Cardie [arXiv:1309.7340 (2013)]
- [64] Applications of Mathematics to Medical Problems, LieutCol and A. G. McKendrick  
[*From the Laboratory of the Royal College of Physicians, Edinburgh (1926) p. 98-130*]
- [65] A multi-agent simulation to assess the risk of malaria re-emergence in southern France,  
C. Linarda, N. Poncon, D. Fontenilles, F. Lambina [ecological modelling, Vol. 220  
(2009) p. 160- 174]

- [66] *Transmission Dynamics and Control of Severe Acute Respiratory Syndrome*, Marc Lipsitch, Ted Cohen, Ben Cooper, James M. Robins, Stefan Ma, Lyn James, Gowri Gopalakrishna, Suok Kai Chew, Chorh Chuan Tan, Matthew H. Samore, David Fisman, Megan Murray [Science vol.300 (2003)]
- [67] *The dynamics of spreading and immune strategies of sexually transmitted diseases on scale-free network*, Jie Lou, Tommaso Ruggeri [J. Math. Anal. Appl. 365 (2010) p. 210-219]
- [68] *Dynamics of Tropical Diseases*, G.A. Macdonald [Oxford Medicine Publications (1973)]
- [69] *Susceptible-Infectious-Recovered Models Revisited: From the Individual Level to the Population Level*, Pierre Magal and Shigui Ruan [Mathematical Biosciences, vol. 250, (2014) p. 26-40]
- [70] *What is a group*, Smith Mark K [the encyclopaedia of informal education (2008)]
- [71] *Infection dynamics on scale-free networks*, Robert M. May, Alun L. Lloyd [Physical Review E Vol. 64 (2001)]
- [72] *Stochastic Approach to Chemical Kinetics*, Donald A. McQuarrie [Journal of Applied Probability , Vol. 4, No. 3 (1967) p. 413-478]
- [73] *Inference in Epidemic Models without Likelihoods*, T. McKinley, A.R. Cook, R. Deardon [The International Journal of Biostatistics. Vol. 5, Is. 1(2009) p. 1557-4679]

- [74] *Global asymptotic properties of an SEIRS model with multiple infectious stages*, Dessalegn Y. Melesse, Abba B. Gumel [Journal of Mathematical Analysis and Applications 366 (2010) p. 202-217]
- [75] *Introduction to Epidemiology*, R.M. Merrill [Jones and Bartlett Publishers (2013) p. 68-71]
- [76] *An agent-based model of sleeping sickness: simulation trials of a forest focus in southern Cameroon*, G. Muller, P. Grebaut, J.P. Gouteux [C.R. Biologies, Vol. 327 (2004) p. 1-11]
- [77] *Mathematical Biology*, J. Murray [Springer-Verlag, Berlin (1993).]
- [78] *Stochastic models of some endemic infections*, I. Nasell [Mathematical Biosciences 179 (2002) p. 1-19]
- [79] *Power-law versus exponential distributions of animal group sizes*, HS Niwa [Journal of Theoretical Biology (2003)]
- [80] *Teaching Epidemiology: A guide for teachers in epidemiology, public health and clinical medicine*, J. Olsen, R. Saracii, D. Trichopoulos [Oxford University Press (2010)]
- [81] *The Impact of 2-Dose Routine Measles, Mumps, Rubella and Varicella Vaccination in France on the Epidemiology of Varicella and Zoster Using a Dynamic Model With an Empirical Contact Matrix*, Mario J.N.M. Ouwens, Kavi J. Littlewood, Christophe

*Sauboin, Bertrand Tehard, Francois Denis, Pierre Yves Boelle, Sophie Alain* [Clinical Therapeutics Vol. 37, N 4 (2015)]

[82] *Modelling the effect of changes in vaccine effectiveness and transmission contact rates on pertussis epidemiology, P. Pescoa, P. Bergeroa, G. Fabriciusa, D. Hozborba* [Epidemics 7 (2014), p.13-21]

[83] *Implementing the Stochastic Simulation Algorithm in R, M. Pineda-Krch* [Journal of Statistical Software , Vol. 25 Issue 12 (2008)]

[84] *Modeling dynamics of HIV infected cells using stochastic cellular automaton, M. Precharattana, W. Triampo* [Physica, A 407 (2014) p. 303-311]

[85] *On the critical behaviour of simple epidemics, C.J. Rhodes, H.J. Jensen, R.M. Anderson* [The Royal Society (1997)]

[86] *Power laws governing epidemics in populations, C.J. Rhodes, R.M. Anderson* [Letters to nature (1996)]

[87] *An Application of the Theory of Probabilities to the Study of a priori Pathometry part I, Ronald Ross* [The Royal Society of London, Series A Containing Papers of a Mathematical and Physical Character, Vol. 92, No. 638 (1916) p. 204-230]

[88] *A Priori Pathometric Equations, Ronald Ross* [The British Medical Journal (1915) p. 546-547]

- [89] *Multi-agent systems in epidemiology: a first step for computational biology in the study of vector-borne disease transmission*, B. Roche, J.F. Guegan, F. Bousquet [BMC Bioinformatics , Vol. 9 (2008) p. 435]
- [90] *Modélisation du paludisme*, C. Rogier, G.Sallet [Medecine Tropicale, Vol. 64 (2004) p. 89- 97]
- [91] *Host-parasite population dynamics under combined frequency and density dependent transmission*, J. Ryder, M. R. Miller, A. White, R. J. Knell and M. Boots [Oikos, Vol. 116 (2007) p. 2017-2026]
- [92] *Transmission Dynamics of the Etiological Agent of SARS in Hong Kong: Impact of Public Health Interventions*, Steven Riley, Christophe Fraser, Christl A. Donnelly, Azra C. Ghani, Laith J. Abu-Raddad, Anthony J. Hedley, Gabriel M. Leung, Lai-Ming Ho, Tai-Hing Lam, Thuan Q. Thach, Patsy Chau, King-Pan Chan, Su-Vui Lo, Pak-Yin Leung, Thomas Tsang, William Ho, Koon-Hung Lee, Edith M. C. Lau, Neil M. Ferguson, Roy M. Anderson [Science vol.300 (2003) p. 1961-1965]
- [93] *Stochastic models for the spread of HIV in a mobile heterosexual population*, A. Sani, D.P. Kroesea and P.K. Pollett [Elsevier Mathematical Biosciences (2006)]
- [94] *Introducing the history of epidemiology*, R. Saracci [Teaching Epidemiology: A guide for teachers in epidemiology, public health and clinical medicine, Oxford Scholarship (2010)]

- [95] *Modeling the Effects of Varicella Vaccination Programs on the Incidence of Chickenpox and Shingles*, M.C. Schuette, H.W. Hethcote [Bulletin of Mathematical Biology 61 (1999) p. 1031-1064]
- [96] *Analytic Comparison of Some Epidemic Models with Vaccination*, M. D. Sen, S. Alonso Quesada, A. Ibeas [Physics Procedia 22 (2011) p. 20-39]
- [97] *D. Bernoulli's work on probability*, O.B. Sheynin [Studies in the History of Statistics and Probability vol. II (1960) p. 105]
- [98] *Evolution towards criticality in an epidemiological model for meningococcal disease*, N. Stollenwerk, V. Jansen [Physics Letters , A 317 (2003)]
- [99] *Criticality in epidemiology*, N. Stollenwerk, V. Jansen [World scientific (2007)]
- [100] *Influence of the motion of individuals on the evolution of a SIRS epidemic*, G.J. Sibona, F. Peruani, G.R.Terranova [Proceedings of BIOMAT 2011 International Symp. on Mathematical and Comp. Biology. Ed. R. Mondaini, World Scientific co (2011)]
- [101] *Introduction To Multi-Agent Simulation*, P.Siebers, U.Aiekelin [University of Nottingham online documentary]
- [102] *A cellular automaton model for the effects of population movement and vaccination on epidemic propagation*, G.Ch Sirakoulis, I. Karafyllidis A. Thanailakis [Ecological Modelling 133 (2000) p. 209-223]
- [103] *Stochastic epidemic models for emerging diseases*, S. Spencer [University of Nottingham PhD thesis]

- [104] *Global analysis of an SEIR model with varying population size and vaccination*, Chengjun Sun, Ying Hen Hsieh [Journal of Mathematical Analysis and Applications 34 (2010) p. 2685-2697]
- [105] *Global analysis of an SEIR model with varying population size and vaccination*, C. Sun, Y. Hsieh [Applied Mathematical Modelling 34 (2010) p. 2685-2697]
- [106] *Estimation of HIV infection and incubation via state space models*, Wai-Yuan Tan, Zhengzheng Ye [Mathematical Biosciences 167 (2000) p. 31-50]
- [107] *Stochastic modeling of AIDS epidemiology and the HIV pathogenesis*, Wai-Yuan Tan [World Scientific publishing Book (2003)]
- [108] *A Stochastic Model for the HIV Epidemic in Homosexual Populations Involving Age and Race*, Wai-Yuan Tan, Z. Xiang [Elsevier Science (Math. Comput. Modelling ) vol.24 No.12 (1996) p. 67-105]
- [109] *Effects of HIV Counseling and Testing on Sexual Risk Behavior*, Lance S. Weinhardt Michael, P. Carey Balir, T. Johnson Nicole, L. Bickham [American Journal of Public Health Vol.89 (1999) N0.9 p. 1397-1405]
- [110] *Modeling epidemics using cellular automata*, S. White, A Martin Del Ray, G. Rodriguez Sanchez [Applied Mathematics and Computation 186 (2007) p. 193-202]
- [111] *Threshold dynamics in an SEIRS model with latency and temporary immunity*, Y. Yuan, J. Belair [Journal of Mathematical Biology Vol. 69 (2014) p. 875-904]

- 
- [112] *Mixing patterns between age groups in social networks*, Y. Del Valle, J. Hymanb, H. Hethcote, S. Eubank [Social Networks, Vol. 29 (2007)]
- [113] *Pulse vaccination delayed SEIRS epidemic model with saturation incidence*, Tailei Zhang, Zhidong Teng [Applied Mathematical Modelling 32 (2008) p. 1403-1416]



# Appendix A

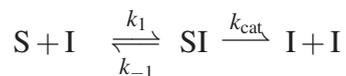
## *SI and SIR Dynamics*

The population density of *free* individuals  $S$ ,  $I$  and  $R$  (those who are not in contact with the others, *i.e.* that do not belong to any group-complex), are expressed by  $[S]_f = [S(t)]_{free}$ ,  $[I]_f = [I(t)]_{free}$  and  $[R]_f = [R(t)]_{free}$  respectively.

We denote also  $[S] = [S]_{tot}$ ,  $[I] = [I]_{tot}$  and  $[R] = [R]_{tot}$  the total quantities of  $S$ ,  $I$  or  $R$  individuals, *i.e.* including free individuals and individuals in groups (in  $SR$ ,  $SI$ ,  $RI$  or  $SIR$  group-complexes). These densities  $[S]$ ,  $[I]$  and  $[R]$  are those that are measurable in the simulations, while the different group-complexes cannot be measured.

## A.1 Modified SI kinetics without resistance (and with latency)

The equations of the reaction mechanism of the SI system are the following:



as we can see in the equation,  $R$  doesn't appear. Then the measurable infection flux can be derived by :

$$\frac{d[I]}{dt} = -\frac{d[SI]}{dt} = k_{cat}[SI]$$

Eq. A.1.1

Assuming the amount of  $[S]$  is large enough for any time  $t$  at which one can measure this quantity, one can formalise one of the following enzyme kinetic hypotheses: the *quasi-equilibrium approximation* of *Henri-Michaelis-Menten* or the *quasi-steady state approach* of *Briggs-Haldane*. In both cases, they specify that the formation of  $SI$  group-complexes is perfectly compensated by their disappearing due to pure dissociation (quasi-equilibrium) or by dissociation coupled to catalysis (steady state). Then one can formulate this assumption that  $\frac{d[SI]}{dt} = 0$  as follows:

$$\frac{d[SI]}{dt} = 0 \Leftrightarrow k_1[I]_f[S]_f = (k_{-1} + k_{cat})[SI] \text{ in the case of stationary equilibrium}$$

and

$$\frac{d[SI]}{dt} = 0 \Leftrightarrow k_1[I]_f[S]_f = k_{-1}[SI] \text{ in the case of quasi-equilibrium where } k_{cat} \ll k_{-1}.$$

The effect of considering one of these above hypotheses as well as the approximation for which  $[S]_f = [S]_{tot} = [S] \forall t$ , are discussed in the appendix A.3. Anyhow, in all the cases we will focus on the steady state that allows us to study the case in which the  $SI$  catalyst is comparable to its dissociation, so with following equation :

$$k_1[I]_f[S]_{tot} = (k_{-1} + k_{cat})[SI]$$

Eq. A.1.2

Then we can formulate the mass conservation equations. In the enzymatic case this is quite simple:  $[S]_{tot}$ , the total amount of substrate  $S$  over time, consists of  $[S]_f$  (*free* substrate),  $[ES]$  (the substrate linked to the enzyme  $E$ ), and  $[P]_f$  the product  $P$  that is the transformed substrate. In the epidemiological case,  $I$  species (individuals) are the same nature as  $S$  and  $R$  species (also individuals): they design persons (or more generally, living species). This means that the mass conservation of  $S$  includes the other classes  $I$  and  $R$  and *vice versa* : one must conserve the amount of individuals. This causes a significant problem for describing the kinetics equations in an epidemiological context. In the appendix A.3, we explain how to deal with this problem by taking into account the relativity of the rates of infection, immunisation and loss of immunity. Then for the modified SI kinetics system (without resistance) the mass conservation equations are:

$$[S]_{tot} = [S]_f + [SI]$$

$$[I]_{tot} = [I]_f + [SI]$$

Injecting  $[I]_f = [I]_{tot} - [SI]$  in  $[SI] = [I]_f[S]_{tot}/K_M$  (Eq. A.1.2) leads to :

$$[SI] = \frac{[I]_{tot}[S]_{tot} - [SI][S]_{tot}}{K_M} = \frac{[I]_{tot}[S]_{tot}}{K_M + [S]_{tot}}$$

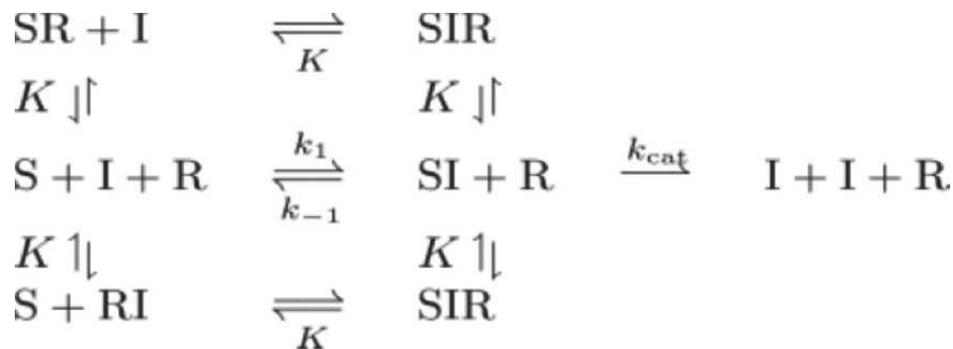
$$\text{then } \frac{d[SI]}{dt} = -\frac{k_{cat}[S]_{tot}[I]_{tot}}{K_M + [S]_{tot}} = -\frac{k_{cat}[S][I]}{K_M + [S]}.$$

$$\text{so } \frac{d[I]}{dt} = \beta_{app}[S][I] \text{ with } \beta_{app} = \frac{k_{cat}}{K_M + [S]}.$$

## A.2 modified SIR kinetics with resistance

By integrating the  $R$  individuals as non-competitive inhibitors in a kinetic system, the general

$SIR$  reaction scheme is described as follows :



**NB.** We distinguish here (compared to the SIR reaction scheme described in section 2.2.1.2) two dissociation constants for the formation of group-complexes  $SI$ ,  $RI$  and  $SIR$  (dissociation constant  $K_B$ ) on the one hand, and for the formation of group-complexes  $SR$  (dissociation constant  $K_A$ ) on the other hand, to maintain the possibility of expressing the fact that the formation of group-complexes including infected individuals may be less

important than the formation of group-complexes that do not contain  $I$  individuals (*i.e.* the group-complexes  $SR$ ). Then  $K_A \leq K_B$ .

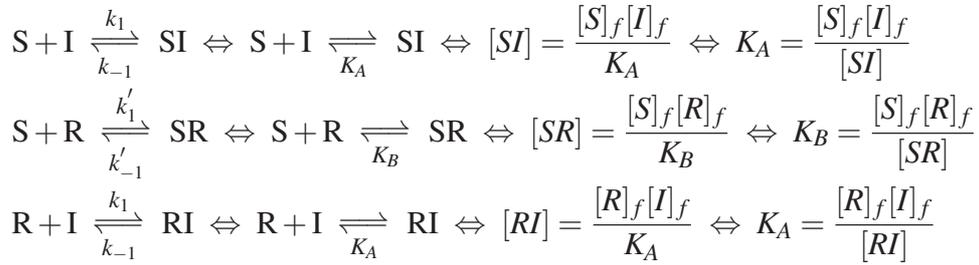
The Mass conservation equations are :

$$[S]_{tot} = [S]_f + [SI] + [SR] + [SIR]$$

$$[I]_{tot} = [I]_f + [SI] + [RI] + [SIR]$$

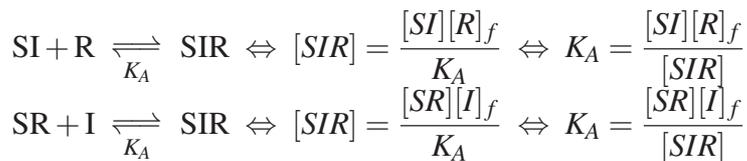
Eq. A.2.2

Several group-complexes can form. Then by writing the dissociation equilibria for bimolecular associations having two  $S$ ,  $I$  or  $R$  components, we have :



so  $K_A = \frac{[S]_f[I]_f}{[SI]} = \frac{[R]_f[I]_f}{[RI]}$  and  $K_B = \frac{[S]_f[R]_f}{[SR]}$  with  $K_A \leq K_B$ .

Similarly, for bimolecular associations with three  $S$ ,  $I$  and  $R$  components, we have:



$$[I]_{tot} = [I]_f + [SI] + [RI] + [SIR]$$

$$RI + S \xrightleftharpoons{K_A} SIR \Leftrightarrow [SIR] = \frac{[RI][S]_f}{K_A} \Leftrightarrow K_A = \frac{[RI][S]_f}{[SIR]}$$

$$\text{so } [SIR] = \frac{1}{3K_A} ([SI][R]_f + [SR][I]_f + [RI][S]_f)$$

Dissociation equilibria are now injected in the equation A.2.2 :

$$[I]_{tot} = [I]_f + [SI] + \frac{[R]_f[I]_f}{K_A} + \frac{1}{3K_A} ([SI][R]_f + [RI][S]_f + [SR][I]_f)$$

$$[I]_{tot} = [I]_f + [SI] + \frac{[R]_f[I]_f}{K_A} + \frac{1}{3K_A} \left( [SI][R]_f + \frac{[R]_f[I]_f}{K_A}[S]_f + \frac{[S]_f[R]_f}{K_B}[I]_f \right)$$

$$[I]_{tot} = [I]_f + [SI] \left( 1 + \frac{[R]_f}{3K_A} \right) + \frac{[R]_f[I]_f}{K_A} \left( 1 + \frac{[S]_f}{3K_A} + \frac{[S]_f}{3K_B} \right)$$

$$[I]_{tot} = [I]_f \left( 1 + \frac{[R]_f}{K_A} \left( 1 + \frac{[S]_f}{3K_A} + \frac{[S]_f}{3K_B} \right) \right) + [SI] \left( 1 + \frac{[R]_f}{3K_A} \right)$$

then:

$$[I]_f = \frac{[I]_{tot} - [SI] \left( 1 + \frac{[R]_f}{3K_A} \right)}{1 + \frac{[R]_f}{K_A} \left( 1 + \frac{[S]_f}{3K_A} + \frac{[S]_f}{3K_B} \right)}$$

Eq. A.2.3

As seen in Appendix A.1, the equation A.2.3 is used in the stationary equilibrium (equation

A.1.2):

$$[IS] = \frac{[I]_f[S]_{tot}}{K_M} = \frac{[I]_{tot}[S]_{tot} - [SI][S]_{tot} \left( 1 + \frac{[R]_f}{3K_A} \right)}{K_M \left( 1 + \frac{[R]_f}{K_A} \left( 1 + \frac{[S]_f}{3K_A} + \frac{[S]_f}{3K_B} \right) \right)}$$

$$[IS] \left( 1 + \frac{[S]_{tot} \left( 1 + \frac{[R]_f}{3K_A} \right)}{K_M \left( 1 + \frac{[R]_f}{K_A} \left( 1 + \frac{[S]_f}{3K_A} + \frac{[S]_f}{3K_B} \right) \right)} \right) = \frac{[I]_{tot} [S]_{tot}}{K_M \left( 1 + \frac{[R]_f}{K_A} \left( 1 + \frac{[S]_f}{3K_A} + \frac{[S]_f}{3K_B} \right) \right)}$$

then we have :

$$[IS] = \frac{[I]_{tot} [S]_{tot}}{K_M \left( 1 + \frac{[R]_f}{K_A} \left( 1 + \frac{[S]_f}{3K_A} + \frac{[S]_f}{3K_B} \right) \right) + [S]_{tot} \left( 1 + \frac{[R]_f}{3K_A} \right)}$$

and

$$\frac{d[I]}{dt} = \frac{k_{cat} [I]_{tot} [S]_{tot}}{K_M \left( 1 + \frac{[R]_f}{K_A} \left( 1 + \frac{[S]_f}{3K_A} + \frac{[S]_f}{3K_B} \right) \right) + [S]_{tot} \left( 1 + \frac{[R]_f}{3K_A} \right)}$$

and as we can reasonably assume that  $[S]_f = [S]_{tot} = [S] \quad \forall t$ ,  $[I]_f = [I]_{tot} = [I] \quad \forall t$  and  $[R]_f = [R]_{tot} = [R] \quad \forall t$ , we have :

$$\frac{d[I]}{dt} = \frac{k_{cat} [I] [S]}{K_M \left( 1 + \frac{[R]}{K_A} \left( 1 + \frac{[S]}{3K_A} + \frac{[S]}{3K_B} \right) \right) + [S] \left( 1 + \frac{[R]}{3K_A} \right)}$$

This gives the following apparent constant rate:

$$\beta_{app} = \frac{k_1 k_{cat}}{(k_{-1} + k_{cat}) \left( 1 + \frac{[R]}{K_A} \left( 1 + \frac{[S]}{3K_A} + \frac{[S]}{3K_B} \right) \right) + k_1 [S] \left( 1 + \frac{[R]}{3K_A} \right)}$$

It also has to be noted that :

$$\frac{d[I]}{dt} = \frac{k_{cat} [I]}{\left( 1 + \frac{[R]}{3K_A} \right)} \times \frac{[S]}{\frac{K_M}{\left( 1 + \frac{[R]}{3K_A} \right)} \left( 1 + \frac{[R]}{K_A} \left( 1 + \frac{[S]}{3K_A} + \frac{[S]}{3K_B} \right) \right) + [S]}$$

Then both the maximal infection rate  $V_{max\ app}$  and the Michaelis constant  $K_{M\ app}$  are affected by the presence of  $R$  individuals in a population.

$$V_{max\ app} = \frac{k_{cat}[I]}{\left(1 + \frac{[R]}{3K_A}\right)} = \frac{V_{max}}{\left(1 + \frac{[R]}{3K_A}\right)}$$

$$K_{M\ app} = \left(\frac{k_{-1} + k_{cat}}{k_1}\right) \frac{\left(1 + \frac{[R]}{K_A} \left(1 + \frac{[S]}{3K_A} + \frac{[S]}{3K_B}\right)\right)}{\left(1 + \frac{[R]}{3K_A}\right)} = K_M \frac{\left(1 + \frac{[R]}{K_A} \left(1 + \frac{[S]}{3K_A} + \frac{[S]}{3K_B}\right)\right)}{\left(1 + \frac{[R]}{3K_A}\right)}$$

### A.3 About Mass Conservation Equations of $[S(t)]_{tot}$

Unlike an enzyme  $E$  in the equations from enzymology, which is completely different from its substrate  $S$ , we have to consider that the nature of  $I$  (infected) individuals is not fundamentally different from those of  $S$  (susceptible) or  $R$  (resistant) individuals. In enzymology, the conservation of the total amount of substrate includes free substrate  $S$ , the substrate combined to the enzyme  $ES$  and the product  $P$ . Then, in the epidemiological context, one should formulate the total mass conservation of species  $S$ ,  $I$  or  $R$  as follows :

$$[S]_{tot} = [R]_{tot} = [I]_{tot} = [S]_f + [I]_f + [R]_f + [SR] + [SI] + [RI] + [SIR] = [N]_{tot}$$

which is obviously not applicable.

Let us see how one could provide a reasonable approximation of the amount  $[S]_{tot}$ . The

solution is related to relativity between the different reaction rates.

In this simplified conservation equation, the term  $R$  is the most problematic. First, the dynamics of formation of  $R$  individuals is much slower than the formation of new infected individuals in a population. In our simulations (see 3) the immunisation reaction is 5 to 50 times slower than the infection (*e.g.*, for a density of 5000 inhabitants per  $km^2$ , the apparent infection constant is  $\beta_{app} = 4 \cdot 10^{-4} km^2 \cdot indiv^{-1} \cdot day^{-1}$  and the immunisation constant rate  $\gamma = 0.08 day^{-1}$ , then the ratio of infection and immunisation is equal to 25). In this context, the resistance process is slow and it can potentially be neglected ( $R$  can be neglected in the conservation equation, at least in early stages of the infection). The second and probably the more important point is that *R individuals are not involved in the infection process itself* since they act as non-competitive inhibitors (in enzymology, this means that the inhibitor does not bind the same site as the ligand to the enzyme). Then it is reasonable to exclude them, as *free* individuals, from the conservation equation of  $S$  and  $I$ , those that are reactive species.

Regarding the presence of group-complexes  $SI$ ,  $RS$ ,  $RI$  and  $SIR$  in the conservation equation, the problem is easy to solve: the dissociation constants  $K_A$  and  $K_B$  are quite large, denoting a large turn-over of groups, which means that if many groups of individuals form, many groups split out at the same time, and this causes most individuals to be considered as *free* individuals. Then the number of group complexes  $SI$ ,  $SR$ ,  $IR$  or  $SIR$  is largely negligible compared to the number of *free*  $S$  or  $I$  individuals.

Finally, based on the same principle, while the dissociation constants are large and by assuming that the number of formed groups is very important ( $k_1$  large), this means that the

first order equilibrium constant  $k_{-1}$  is also very large, much larger than the catalytic constant  $k_{cat}$ , *i.e.* the group-complex dynamics are largely faster than the infection dynamics. In our simulations (with reasonably realistic parameters), we have  $k_{-1}/k_{cat} \simeq 100$  to 1000. Thus, the occurrence of new infected individuals  $I$  is very slow compared to the quasi-equilibrium  $S + I \rightleftharpoons SI$ , which strongly leads individuals to remain in the *free*  $S$  and  $I$  population. Then, if the number of  $I$  individuals is quite low at the beginning, it is reasonable to affirm that  $[S]_{tot} \simeq [S]_f$ .

It has to be added that in actual observed epidemic processes, the number of infected cases is effectively around 100 – 500 persons per 100000 in the population (*e.g.* during the occurrence of major outbreaks of viral gastroenteritis during the week 50 in December 2014 in France, the observed maximum number of infected people was 300 per 100000 persons). Therefore the number of  $I$  is in practice very small and negligible compared to the number of  $S$  individuals, and then it does not count in the conservation equation.

# Appendix B

## Effective Surface

What is the actual area in which individuals move and have interactions with others? The answer is not simple, but the effective surface should probably be a very small percentage of the actual surface. It might be comprised between 0.1% and 1% of the surface (according to an empirical estimation made in the laboratory and from maps of Grenoble).

This surface depends firstly on how crowded the occupied places are (see Figure B.1): desks, chairs, plants, heaters, baskets, fridges,... but also walls, doors. All these objects take up a large surface, and the individuals' movements are limited by the available space to move. Usually, one walks in a room or a hall at least at 20 to 30 cm from the objects and the walls. For example, our desk is about  $12\text{m}^2$  in surface, but we two are occupying an effective surface of about only  $2\text{m}^2$ .



Figure B.1 The plan and a picture of an open space in a building illustrates how space is already taken up by objects, and how individuals can only move in the free zone. Here, only the grey area can be effectively used by individuals. In practice, their movement is more limited due to the additional presence of various objects.

Above all, outdoor spaces consist of some areas (the middle of a road or spaces where cars are parked) that we rarely expect to see individuals: it is generally preferred to walk along the sides of a parked car and walk on the sidewalks. No walks or presence are expected (in general) in the green spaces of cities or in the middle of rivers. Therefore, all these limitations for occupying the spaces around us reduce the effective surface (Figure B.2) and this consequently increases the effective overall population density. For example, for a population density of  $8.7 \cdot 10^3$  inhabitants per  $km^2$  for Grenoble city and its suburbs, we can probably say that an effective density (modified in accordance with the available space to occupy) can be comprised between  $8.7 \cdot 10^5$  and  $8.7 \cdot 10^6$  inhabitants per  $km^2$ . By doing a quick calculation we find an adjusted density of  $1.4 \cdot 10^6$  to  $5.7 \cdot 10^6$   $hab.km^{-2}$  in a classic lift which can hold up to 8 people and have a surface of  $1.4 \times 1.0 m^2$ .



Figure B.2 Where are the people? We distinguish between the area that can be occupied by individuals as effective surface to meet and interact with others (in the context of disease transmission) and the areas that are not suitable are not more frequented for this aim. For example the real surface of Grenoble consists of space took up by the river that pass through the city (the Isère river), two graveyards, the mountains (here, the Mount Jala), parks, large boulevards ... ; this excludes a lot of space from the effective surface !

Then it is important to take into account the effective surface to calculate the adjusted population density in a given population. The adjusted density by effective surface as a more realistic value that may play a very important role in the control of epidemics. Density is a crucial factor in the spread of many communicable diseases, then reorganising the space and facilitating the movement of people should play an important role in the control and prevention of diseases.

