Quantifying the effect of synchrony on the persistence of infectious diseases in a metapopulation

Tran Thi Cam Giang, Marc Choisy, Jean-Daniel Zucker, Yann Chevaleyre

16/04/2015 Version 1.1

Abstract

Global persistence of infectious diseases is a big problem for epidemiologists. Studies have showed that there are a lot of reasons to answer why many communicable diseases still exist and have been developed in more dangerous form [4, 28, 35, 36, 42]. The asynchrony and the recolonization among subpopulations are two key reasons pointed out. However, why these are the asynchrony and the recolonozations in a metapopulation is still an open question. Here we study the combined effects of forcing phase heterogeneity in the seasonally forced contact rate on global persistence of disease. We carry out an exploitation of stochastic dynamics in a susceptible-exposed-infectious-recovered (SEIR) model of the spread of infectious diseases in a metapopulation of n subpopulations. Starting with continuous-time Markov description of the model of deterministic equation, the direct method of Gillespie(1977) [16] in the class of Monte-Carlo simulation methods allows us to simulate exactly the spread of disease with the SEIR model. Our finding of the exploitation of stochastic dynamics points out that the persistence of the disease in the meta-population is characterized as an exponential survival model on data simulated by the stochastic model. Using a parametric survival model for an exponential distribution (R package 'survival' [43]), we estimate the global extinction rate which represents the global persistence of disease in the meta-population. We find how bigger the forcing phase heterogeneity becomes, and how larger the global persistence gets.

Keywords: SEIR model, Markov chain, Monte-Carlo simulation methods, spatial synchronization, disease persistence, meta-population.

1 INTRODUCTION

News about infectious diseases has always been a subject of worry to parents as well as all. It has brought many problems to humain society. Recent works have shown that infectious diseases do spread in space [11, 17, 41, 44]. The exact form of disease movement depends on a number of local factors (demographic including population size [21], growth rate and death rate [9], sociological such as school period of children, work tendency from rural to urban, environmental and

climatic comprising seasonal variations in seasonality [10, 19], temperature and rainfall, immunological for diseases, etc...) as well as the connections between the different populations (i.e. spatial structure) [46] such as distance [21], coupling rate, number of individuals between populations, etc....Hence, we focus on spread of disease in space by using variations in the seasonal aspects of subpopulations and after examine how synchrony could affect the persistence of infectious diseases in a metapopulation.

In modeling of ecological system, presenting interactions between humans, subpopulations, geographic conditions and the metapopulation model is a good choice. Metapopulation is a set of subpopulations with mutual interaction [33] here a subpopulation can only go extinct locally and be recolonized by another after it is emptied by extinction [7, 22, 33]. It is the reason that the ecological theory of metapopualtion can point out that the persistence of a population depends on the dynamics of migrations between the different sub-populations of the metapopualtion [15, 21, 22, 33]. The "colonization" caused by migrations brings infection to an uninfected subpopulation and we call infected individuals colonizers.

In addition, the disease persistence capability in a metapopulation depends positively on level of asynchrony between subpopulations [21, 24]. Many studies showed that the asynchronization of epidemics among subpopulations causes the recolonization of diseases for locally extinct subpopulation [7, 21, 22, 24, 45]. So, the recolonization becomes the main reason for which disease persistence exists. The disease always appears in metapopulation if and only if there is at least one non-extinct subpopulation. In 1996, in order to explain why measles persists after lot of vaccination policies, Bolker [7] used the measles data before and after vaccination from 1964 to 1988 in England and Wales. Vaccination has broken high synchrony between the UK cities in prevaccination era, and at the same time, causes decorrelation and enhances global persistence of the infection, because of decorrelating factors of vaccination such as the starting moment of vaccination policies, number of susceptibles vaccinated and interaction between vaccination policies [7, 38]. A decrease in correlation between subpopulations may make a metapopulation more difficult to eradicate infectious diseases [7, 13]. In addition, the level of synchrony between the subpopulations is strongly governed by the migration rate and distances between them [24]. In our modern world, the distance problem isn't large anymore for individuals who want to travel. There are a lot of cities very remote, but very connected, and thus very synchronous as in the USA [8]. In contrast, migration among subpopulations has become a big problem. The disease synchrony speed within a metapopulation can strongly increase when migration rate there is strong [31]. The migration rates are directly proportional to the amount of variation in metapopulation size, but inversely to the amount of variation in subpopulation size, over time [12, 19]. So, migration is key to the recolonization of empty subpopulation and simultaneously increases the degree of synchrony between subpopulations in spatially structured metapopulation. However, the vast majority of infectious diseases control policies that are applied in the world are still based on rationales that do not consider the local extinction/recolonization dynamics. This is maybe a reason why measles persists around the world, despite highly local vaccine coverages [9]. For example, in the start of the 2014, the World Health Organization (WHO) had officially

stated the global measles epidemic outbreak. In the first three months of the year 2014, there were about 56,000 cases of measle infections in 75 countries, including countries in south-east Asia and most particularly, Vietnam. We discovered measles persistence in the world for many years without extinction, from one nation to another as well as from cities to other cities. Though the moment of disease outbreaks in each region differs. For neighbouring regions with disease persistence, there is a time-lag differs between disease outbreak. This is explained in sociology by difference in culture, in geographic condition and more particularly seasonality.

Seasonality has been one in rather robust ingredients influencing the disease persistence process. Seasonal changes can alter migration tendency between urban and rural areas [14], also residence time of hosts, vectors and pathogens. Seasonal variation can thus determine population size, migration and interaction capabilities and particularly infection rate at which susceptible individuals become infected [1, 26]. Hence, infectious disease outbreak occurs due to this infection rate. However, finding a clear mechanism of seasonal forcing for modelling is a very difficult work because of unidentified formula for seasonal forcing [1, 14]. For indirectly transmission diseases such as water-born and vector-born, finding seasonality characteristics is less a problem, however, in direct contrast to transmission diseases such as measles. Seasonal forcing in a metapopulation is influenced by weather and climate in region, school schedule of children, and rural-urban migration in countries [5, 9, 14, 20]. In these factors, the seasonal aggregation of children in primary schools affects clearly the infection rate in metapopulation. The infection rate decreases due to children holidays but is inverse when the children come back to school [9]. So, exploring the influence of seasonality for the infection rate in simulation has been developed in many previous years. If the infection rates given are the same in all subpopulations, so this metapopualtion model is a rather simple model [39, 40] and the symmetry of the fixed points among subpopulations will not be broken. In contrast, if the infection rate is different in all subpopulations, so we have a more complex oscilation metapopulation model, but close to the oscilations in reality. Thus, realizing the oscilations of infectious diseases in life within metapopulation simulation models to estimate global disease persistence time has become a large problem and the infectious disease eradication has become our aim [13].

Here we propose a simulation study to quantify the effect of synchrony on the persistence of infectious diseases. We use stochastic simulations for infectious diseases in a metapopualtion, then we consider different spatial structures from the simplest to complex. These forcings can reflect local demographic, sociological, environmental, or climatic factors. The level of synchrony is computed from the phases of forcing in the different subpopulations and persistence is quantified by using statistical tools from survival analysis. Here, we are concerned about measles and simultaneously use the parameter values from articles and measles reports. As the persistence of measles has been largely studied in the literature and its still unexplained while global persistence is a growing concern for WHO and public health authorities around the world [10].

To do this, we first build the SEIR deterministic model for a metapopualtion. Then, we introduce a new interpretation of the infection force for a metapopulation, that is very different from all previous interpretations. We describe

the spatial structure and the stochastic version of the SEIR metapopulation model. We propose a small improvement in the original stochastic method of Gillespie [16] for a metapopulation. Finally, we show a new method to estimate global extinction in a metapopulation by using a parametric survival model for an exponential distribution.

2 MATERIAL AND METHODS

2.1 MATERIAL

2.1.1 Deterministic model for many subpopulations

The standard SEIR model (susceptible-exposed-infective-recovered) has been strongly developed for the dynamics of directly infectious disease [5]. For disease-based metapopulation models, we give here a suitable new version of the SEIR equation that would be as follows:

Consider a metapopulation of n sub-populations. In a subpopulation i of size N_i, disease dynamics can be deterministically described by the following set of differential equations [2]:

$$\frac{dS_i}{dt} = \mu N_i - \lambda_i S_i - \mu S_i \quad (1)$$

$$\frac{dE_i}{dt} = \lambda_i S_i - \mu E_i - \sigma E_i \quad (2)$$

$$\frac{dI_i}{dt} = \sigma E_i - \mu I_i - \gamma I_i \quad (3)$$

$$\frac{dR_i}{dt} = \gamma I_i - \mu R_i \quad (4)$$

where S_i, E_i, I_i et R_i are the numbers of susceptible, exposed, infectious and recovered in this sub-population i respectively. Individuals are born susceptible, die at a rate μ, become infected with the force of infection λ_i, infectious after a latency period of an average duration of $1/\sigma$ and recover at the rate γ. In a case the infectious contact rate is constant, the equilibrium values of the variables S, E, I and R can be expressed analytically (see appendix). The force of infection λ_i (the amount of time spent in the infectious class) depends not only on the total population size N_i and the number of infected I_i in subpopulation i, but also in other subpopulations [31]. In addition, many studies have also pointed out that the infection force depends upon three main factors of the disease transmission from S to E : the prevalence of infecteds, the underlying population contact structure, and the probability of transmission given contact. For a directly transmitted disease, the key factor is the contact between susceptible and infected individuals. Hence, in this article, we interpret the "infectious period" in more detailed, and realistic contexts. We show a new interpretation of the infection force. We study the case where individual x native from subpopulation i visits subpopulation j. In a metapopulation of subpopulations, during a small interval of time δt, each native individual of the subpopulation

i visits one single subpopulation j (with probability ρ_{ij}) and will see on average K_j individuals who come from all subpopulations. Therefore, we look at the probability that a susceptible $x \sim \mathcal{X}_i$ visiting j gets infected or not after δt time steps. Let $\mathcal{Y}$ be the uniform distribution over $V_{j,t}$. The correct mathematical approach for this would be to assume that for each subpopulation k, the number of people native from k that we meet during δt follows a Poisson process. So both the number of people we meet and the number of infected people we meet during δt should be random variables. However, in the approach described in [31], the authors did not do this. They assumed that both the number of people we meet and the number of infected people we meet *are fixed* (otherwise the maths they write would have been different). We call this interpretation proposed by "Keeling & Rohani" the old interpretation of the infection force (for short, OIIF) that we can see more detailed in appendix. This assumption simplifies the relations between individuals and between subpopulations. It's the reason for that the formula of the infection force did not present clearly the complex connections between individuals and between subpopulations in a metapopulation. Hence, here we introduce a new absolutely approximation of the probabilistic derivation of multipopulation epidemic model that has the name "New Interpretation of Infection Force" (for short, NIIF). We assume that the number κ of people we meet during δt is *fixed*, but each of these people has *some probability* to be infected. This is an *in-between interpretation*, easier than the Poisson process maths, but better than the OIIF (see more detailed in appendix). We have the NIIF as follows:

$$\lambda_i = \sum_j \rho_{ij}\kappa_j \log\left[1 - \sum_{k=1}^{M}\left(\frac{|I_{k,t}|}{N_k} \times c_{ik} \times \xi_{jk}\right)\right] \tag{5}$$

Where $c_{i,k}$ $(0 \leqslant c_{ij} \leqslant 1)$ is the probability that a susceptible individual native from i being in contact with another infected individual native from k gets infected. ξ_{jk} $(0 \leqslant \xi_{ij} \leqslant 1)$ refers to the probability that an individual y meeting x in C_j comes from C_k. κ_j is the average number of contacts per unit of time a susceptible will have when visiting subpopulation j. $\rho_{i,j}$ $(0 \leqslant \rho_{ij} \leqslant 1)$ is denoted as the probability that an individual from subpopulation i visits subpopulation j, of course, $\sum_{j=1}^{M} \rho_{ij} = 1$. See appendix for detail on the construction of this equation. We can verify that in the limit case on one single subpopulation in the metapopulation ($i = j$ and $n = 1$) we have

$$\lambda_i = -\kappa_i \log(1 - \frac{I_i}{N_i} \times c_{ii}) \tag{6}$$

Consider that the average number of contacts per unit of time κ_i is seasonally forced [1] and seasonality is an annually periodic function of time [18]. As a result, for the subpopulation i :

$$\kappa_i(t) = \kappa_{i0}\left[1 + \kappa_{i1}\cos\left(\frac{2\pi t}{T} + \varphi_i\right)\right] \tag{7}$$

where t is the time, κ_{i0} and κ_{i1} are the mean value and amplitude of the average contact rate κ_i at which a susceptible

will have when visiting subpopulation i per unit of time, T and φ_i are the period and the phase of the forcing. With the annual sinusoidal form of the average contact rate, we really have the sinusoidally forced SEIR metapopulation model.

2.1.2 Stochastic model for many subpopulations

In order to study the persistence of the disease, we must consider a stochastic version of the model [30, 34, 37]. We use for that a population-based time-to-next-event model based on Gillespie's algorithm [16]. Table 1 lists all the events of the model, occurring in subpopulation i.

Table 1 – Events of the stochastic version of the model of equations 1-4, occuring in subpopulation i.

Events	Rates	Transitions
birth	μN_i	$S_i \leftarrow S_i + 1$ and $N_i \leftarrow N_i + 1$
death of a susceptible	μS_i	$S_i \leftarrow S_i - 1$
death of an exposed	μE_i	$E_i \leftarrow E_i - 1$
death of an infected	μI_i	$I_i \leftarrow I_i - 1$
death of an immune	μR_i	$I_i \leftarrow I_i - 1$
infection	$\lambda_i S_i$	$S_i \leftarrow S_i - 1$ and $E_i \leftarrow E_i + 1$
becoming infectious	σE_i	$E_i \leftarrow E_i - 1$ and $I_i \leftarrow I_i + 1$
recovery	γI_i	$I_i \leftarrow I_i - 1$ and $R_i \leftarrow R_i + 1$

To apply the stochastic method of Gillespie [16] in a metapopulation of n subpopulations, we know that at a moment t, there are only one single event fired in one single subpopulation. Hence, we improve this stochastic method of one single population for a metapopulation by adding a random number to calculate chance which subpopulation to fire in the metapopulation at the moment t. Starting from the initial states, the stochastic simulation algorithm simulates the trajectory in population processe by repeatedly answering the following three big questions and updating the states.

- When (time) will the next event occur?
- Which subpopulation where the event will occur next?
- Which event in the subpopulation will occur next?

It is a small improvement in the original stochastic method of Gillespie [16], however it turns back much more benefits for a metapopulation, it well presents the interactions between subpopulations in the metapopulation.

2.1.3 Spatial structures

A metapopulation is a population of populations (subpopulations). Such a structure implies an heterogeneity in the sense where the probability of contact (or contact rate) between individuals from a same subpopulation is higher than

the probability of contact between individuals of different subpopulations [15]. Such heterogeneity is actually the result of the interaction between two phenomena that are often difficult to disentangle in nature. The first one relates to the granularity of the metapopulation (as rendered by the number of and sizes of subpopulations) and the second one relates to the isolation between subpopulations (as can be rendered, among others, by physical distances separating each pair of subpopulations). Moreover, according to the findings of Benjamin Bolker (1995) [5], there is no coexistence between periodicity and disease persistence in non-spatial measles models, and spatial structure is an important factor to both enhance persistence and create new types of dynamic behaviour.

To identify clearly the causes of observed phenomena, these two aspects will be modeled distinctly. In this article, our null model (model 0) will be a metapopulation without any explicit spatial distance (all the subpopulations are at the same distance from each other) and where all the metapopulation have the same population size N. Like the original Levins's model [33], this model considers that all the subpopulations are at equal distance from each other:

$$\rho_{ij} = \rho, \qquad 0 \leqslant \rho \leqslant 1, \qquad \forall i, \forall j. \tag{8}$$

The structure of this metapopulation is thus characterized by 3 parameters: (i) the number n of sub-populations, (ii) the population size N ($N_i = N$, $\forall i$) of all these subpopulations and, (iii) the coupling (or distance) ρ_{ij} between two subpopulations i and j that refers to the the probability that an individual from subpopulation i visits subpopulation j.

2.2 METHOD

2.2.1 Stationary distribution in metapopulation

Here we show some assumptions for the stationary distribution model as follows :

- Assumption 1. For each subpopulation V_i, there exists a markov chain M_i describing where (i.e. in which subpopulation) individuals native from V_i travel at each time step.
- Assumption 2. Each M_i has a stationary distribution $\rho(M_i)$.
- Assumption 3. At time t=0, each individual is located in a subpopulation randomly drawn from $\rho(M_i)$.

Therefore, when we consider a simplified model in which the dynamics of the individuals is stationary: each individual native from V_i no more follows a markov chain, but is relocated at each time step on a subpopulation randomly drawn from $\rho(M_i)$.

Then, under assumptions 1, 2, 3, at any time t, when the total number of individuals grows to infinity, the size of the populations under the markovian dynamics converges towards the size of the populations under stationary dynamics.

Thus, any statistics computed on the densities of individuals from the same population in various cities will not distinguish the markovian from the stationary dynamics.

Based on this conclusion, we will deploy a stationary distribution in a metapopulation by building a transition matrix and then computing its convergence. This matrix converges towards a stationary distribution matrix. Finally, we will apply the stationary distribution matrix in the metapopulation of n subpopulations.

2.2.2 Global persistence in a metapopulation

In order to examine questions of interaction between disease transmissibility and phase of seasonal forcing, we start in this section by studing the stochastic SEIR model in a metapopualtion of n subpopulations. For this meta-population, we observe the disease extinction in time due to spatial synchrony/asynchrony that are influenced by phase difference in seasonal forcing. To create the phase difference, we change the value of the forcing phase for each subpopulation. In this experience, we use a parameter φ_{max} in radian that runs in the interval from zero to π. With each value of φ_{max}, based on n the number of subpopulations in the metapopulation, we divide the interval $[0, \varphi_{max}]$ into a set of $(n-1)$ equal samples, so the value of the forcing phase of the i^{th} subpopulation is correspondent to i^{th} value in the set. We call φ_{max} asynchrony parameter.

The persistence of disease in the metapopulation was characterized by fitting an exponential survival model [10, 32] on a data simulated by a stochastic model. To measure the persistence in ecology and epidemiology, so many methods we can use [10, 20, 30]. For example, Keeling et al.(2002) [30] gave two methods. One method was for an isolated metapopulation without migration by calculating the expected extinction time or the extinction rate during a given period. This was a theoretical measure as no real data exists to compare with model results. The other method for a population with migration was found by calculating the number or the total duration of extinctions. Then in 2010, "mean annual fade-out" and "fade-outs post epidemic" methods proposed by Conlan [10] were used to quantify persistence by basing on the proportion or on the frequency of zero reports in a given reporting interval. In this work, we propose a new rate that presents also the survival of disease in a metapopulation. This is the mass extinction rate. For our metapopulation of n subpopulations, to do so we run first m independent simulations of our stochastic model. We calculate then the average metapopulation size by summing subpopulations at each sample time and averaging across the entire time series for each metapopulation. Lastly, we record the dates t of global disease extinction in all these m metapopulations. These dates allow to draw Kaplan-Meier survival curves from which we estimate the global extinction rates χ:

$$M(t) = \exp(-\chi t) \tag{9}$$

where $M(t)$ $(0 \leqslant M(t) \leqslant m)$ is the number of metapopulations in which the disease is not extinct at time t.

We use the parametric survival model for the exponential distribution (R package '*survival*' [43]) to estimate the

mass extinction rate of a metapopulation during 100 years. Due to that, we can capture one of the most important features of stochastic systems in spatial structure : its global extinction characteristics of disease.

2.2.3 Characterization of synchrony

Call $\delta_{ij} = \delta_{ji}$ $(0 \leqslant \delta_{ij} < 2\pi)$ the phase difference between subpopulations i and j :

$$\delta_{ij} = |\varphi_i - \varphi_j| \bmod 2\pi \tag{10}$$

where φ_i and φ_j are the phases of the contact rates (equation 7) in subpopulations i et j. Populations i and j are perfectly in phase if $\delta_{ij} = \delta_{ji} = 0$ or 2π and in opposition of phase if $\delta_{ij} = \delta_{ji} = \pi$. We can thus define the degree of synchrony $\xi_{ij} = \xi_{ji}$ $(0 \leqslant \xi_{ij} \leqslant 1)$ between populations i and j as

$$\xi_{ij} = \left| 1 - \frac{\delta_{ij}}{\pi} \right|. \tag{11}$$

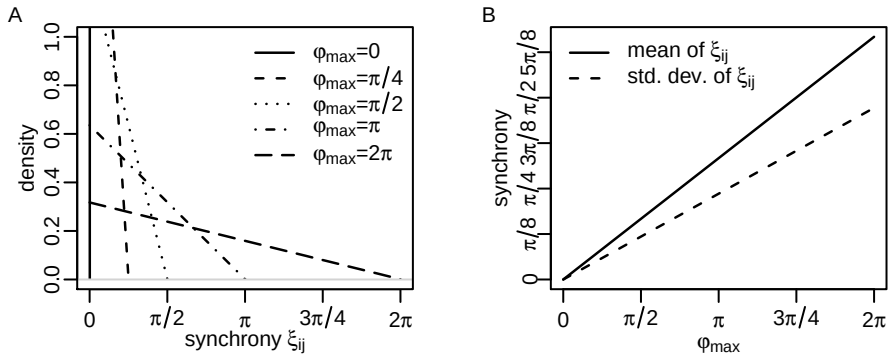

Figure 1 – Synchrony in the case of model 0. (A) distribution of synchrony ξ_{ij} for various values of φ_{max}. (B) mean and standard deviation of the distribution of ξ_{ij} as functions of $\varphi_{\max}$.

Consider that in the metapopulation the phases φ_i of the contact rates in the n subpopulations are evenly distributed between 0 and $\varphi_{\max}$ $(0 \leqslant \varphi_{\max} \leqslant \pi)$. We can express the mean of the pairwise phase differences $\delta_{ij} = \delta_{ji}$ as

$$<\delta_{ij}> = <\delta_{ji}> = 2\varphi_{\max} \sum_{k=1}^{n-1} \frac{(n-k)k}{(n-1)n^2} = \frac{n+1}{3n}\varphi_{\max} \tag{12}$$

and thus the mean of the synchronies $\xi_{ij} = \xi_{ji}$ as

$$<\xi_{ij}> = <\xi_{ji}> = 1 - \frac{n+1}{3n}\frac{\varphi_{\max}}{\pi} \tag{13}$$

and thus

$$\lim_{n\to\infty} <\xi_{ij}> = 1 - \frac{\varphi_{\max}}{3\pi} \tag{14}$$

As illustrated in Figure 1, for a high enough number n of subpopulations, the mean value of the ξ_{ij} does not depend on the number of subpopulation.

The values of φ_i are chosen so that they are uniformly distributed between $\varphi_{\min} = 0$ and $\varphi_{\max}$. The distribution of ξ_{ij} doesn't depend on n the number of subpopulation, but only depends $\varphi_{\max}$ and may be is characterized by one single parameter (we choose the average value of all ξ_{ij}), view figure 1.

2.3 Plan of experience

In this section, we will describe our plan of experience to quantifying the effect of synchrony on the persistence of infectious diseases in a metapopulation. We have four big concerns that we must verify.

2.3.1 Quantifying disease persistence in the simplest metapopulation

Firstly, we are interested in the correlation between global disease persistence time and recolonization in metapopulation when the forcing phase of the average contact rate κ in each subpopulation alters. In order to simplify this concern, we start with the metapopulation of two subpopulations. To have the two subpopulations in synchrony, we choose $\varphi_{max} = 0$. It means that the forcing phase of $subpopulation_1$ and $subpopulation_2$ are the same phase, $\varphi_1 = \varphi_2 = 0$. The fluctuations of the contact rates $\kappa(t)_1$ and $\kappa(t)_2$ are in the same annual sinusoidal form. In contrast, to have the two subpopulations in asynchrony, we create the phase difference or the phase lag between the subpopulations. It means that we have created a metapopulation in heterogeneity. Here, we choose $\varphi_{max}= \pi/2$ and $\varphi_{max}= \pi$. For $\varphi_{max}= \pi$, we have $\varphi = 0$ and $\varphi_2 = \pi$, then the fluctuations of $\kappa(t)_1$ is phase lag for $\kappa(t)_2$'s. Based on φ_{max}, we are successfully buiding a complex disease metapopulation mode.

2.3.2 Quantifying global extinction and asynchrony level φ_{max}

We will exploit more strongly the role of asynchrony φ_{max} for global disease extinction rate in metapopulation. φ_{max} is an important parameter that we use to break fixed points as well as first fixed points at begining moments between subpopulations. Our goal is to examine the mass extinction for each φ_{max} in the metapopulation. Simply, we apply survival regression model over global peristence curve at each different φ_{max} value to estimate its global extinction rate. We start also with the metapopualtion of n subpopulations given and φ_{max} from 0 to π. It means that we increase level of phase difference from the $subpopulation_1$ to $subpopulation_n$. We deploy m the number of different simulations. Then, we use survival analysis to quantify extinction rate at each value of φ_{max} with confidence interval to 95%.

2.3.3 Influence of other parameters on the mass extinction rate

The experiences will turn out how the metapopulation size, the coupling strength between subpopulations and the number of subpopulation affect the mass extinction rate as well as when the asynchrony parameter φ_{max} varies.

2.3.4 Stochastic metapopulation simulations

In order to run simulations, we use the same values of all parameters for all subpopulations. We use the Gillespie's direct algorithm [16] for metapopulation model as described in the previous part. With the SEIR metapopulation model, measles is modeled [3, 20]. Moreover, in this work, we use also the values of parameters for the measles to do experiences. We have a table of the convenient values for parameters of measles as follows :

Table 2 – Some Disease Parameter Values for Measles from the Literature [5, 8, 10, 29, 30, 31]

parameter	**description**
μ	birth and death rate per day
κ_0	mean value of the number of contacts κ per unit of time a susceptible will have when visiting one subpopulatio
κ_1	amplitude of the number of contacts κ per unit of time
γ	recovery rate per day
σ	average exposed duration per day
ρ	coupling rate (ρ_{ij} the probability that an individual from subpopulation i visits subpopulation j)
φ_{max}	synchrony parameter in radian
N	population size of subpopulation
n	number of subpopulation
t_{max}	simulation time

Following the table in detail about the convenient values of parameters, we will use them throughout all simulations. We start doing a simulation from an initial random number. Then, we aggregate the daily data (number of individuals in the suceptible, exposed, infected and recovered groups) into one-day intervals, and use this as the time step in the model.

3 RESULT

3.1 Quantifying disease persistence in the simplest metapopulation

We start quantifying global disease persistence in the metapopulation of two subpopulations by examining degree of synchronization between infected individuals in two connected subpopulations. This is the simplest metapopulation, for which we can view easily the influence of the level of synchrony on the disease persistence and recolonization of disease in the metapopulation over time due to the asynchrony parameter φ_{max}.

We chose here the metapopulation of two subpopulations, $N_1 = N_2 = 300,000$, the rate of coupling $\rho = 0.01$, the number of simulation repeats $m = 100$, and $\varphi_{max} = \{0, \pi/2, \pi\}$. Now, we have 100 metapopulations, we gather the

first time where the metapopualtion gets global extinction. We have three Kaplan-Meier survival curves for each value of φ_{max} as figure 2.

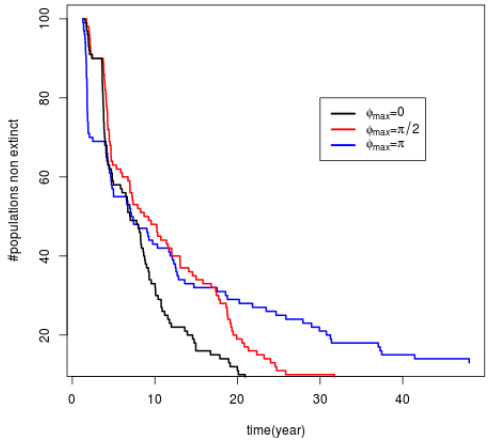

Figure 2 – Kaplan-Meier survival curves for disease persistence after 100 different simulations of $\varphi_{max} = 0$, $\varphi_{max} = \pi/2$ and $\varphi_{max} = \pi$. The disease persistence time of $\varphi_{max} = 0$ is the shortest and that of $\varphi_{max} = \pi$ is the longest.

The phase of forcing of the $subpopulation_1$ is always fixed at 0, but that of the $subpopulation_2$ increases from 0 to π. In the first experience $\varphi_{max} = 0$, the two subpopulations are in synchrony with all beginning conditions. In the simulation time, they are easy to find local extinctions together. The metapopulation goes globally extinct in a short time. The disease persistence time is the shortest. Moreover, the two subpopulations become asynchronous when $\varphi_{max} = \pi/2$ or π. The symmetry of fixed points is just broken at the starting moment. The level of synchrony of the metapopulation decreases. The level of asynchrony φ_{max} is directly scaled with the phase difference in the forcing rates, but inversly with the global persistence time. When $\varphi_{max} = \pi$, the two subpopulations are in antiphase. The balance points at the begining moment of $subpopulaiton_1$ are inverse to that of $subpopulation_2$, they accounts for why the two subpopulations take a long duration to obtain the balance state. This is the most difficult case to find global extinction, the persistence time is the longest. Moreover, theses obtained results are also explicated by the recolonization between the two subpopulations. When the disease no longer exists in one subpopulation at the moment t. If at this moment, all neighbouring subpopulations obtain also the extinction, so we have a global extinction in the metapopulation and the disease is entirely extinct. However, in a metapopulation, due to different factors for migration between subpopulations, we hardly see extinction at the same time in all subpopulations in first duration. The infected in other subpopulation migrates to the extinct subpopulation, so the disease is active. It is the reason why the disease exists for long term. In short, the level of synchrony between subpopulation is stronger, metapopualtion is easier to find global extinction. Make all subpopulations synchronize is the easiest way at which disease goes to extinct.

Additionally, we have exploited experiences on local fluctuations of subpopulations. Here, local fluctuations is called

local dynamics or local noises. This is an important factor that can increase local extinction numbers but make global extinction rate reduce as the following figure 3.

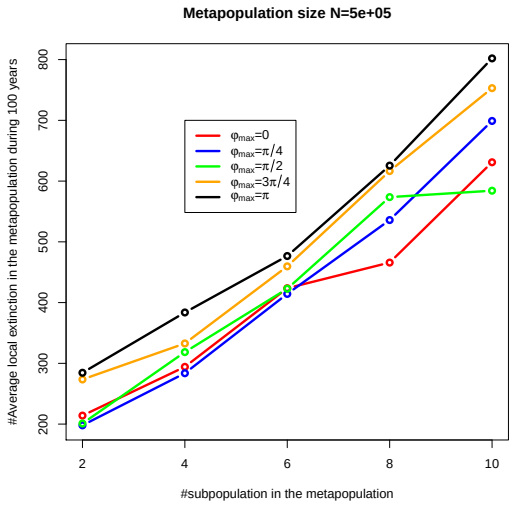

Figure 3 – Influence of the variation's asynchronic parameter on the relation between the average number of local extinctions and the number of subpopulations in a metapopulation. The number of iterations = 100 for each parameter; the probability of foreign visit=0.1; the metapopulation size = 5e+05.

As the result shown in the figure 3, the rate of asynchrony φ_{max} robustly affects the relation between the average local extinction number and the number of subpopulation in a metapopulation. We can find two main results here. First, the average local extinction number in any metapopulation has an clear increase when the rate of asynchrony φ_{max} augments. At the rate $\varphi_{max} = 0.0$, it means to set the entire metapopulation in the synchrony state, the dynamics of all subpopulations are synchronous and the local extinction positions of the subpopulation may be also synchronous. Therefore, this average number in the metapopulation is minimum. Inversely, when the rate of asynchrony φ_{max} starts tending to increase, the fluctuations are also changed and fallen into the difference phase. It is the reason why when a subpopulation goes locally extinct and after it is redominated in a short time by disease due to the recolonisation among subpopulations. At the rate $\varphi_{max} = \pi$, the difference phase is maximum. The average number of local extinction in a metapopulation is maximum. Second, in a metapopulation, if the subpopulation number increases, then the local extinction number augments also. Because when the subpopulation number is directly scaled with the metapopulation size and the time of disease persistence. In addition, the interaction among subpopulation increases with the subpopulation number. One subpopulation is easy to be dominated by the other subpopulations due to the recolonisation.

3.2 Quantifying global extinction rate and asynchrony level φ_{max}

As mentioned about, we use survival regression model for global peristence curve to estimate its global extinction rate for each value φ_{max} with confidence interval 95%. The result below is pointed for the metapopulation of 15 subpopulations, the metapopulation size $N = 10^6$ and the coupling rate $\rho = 0.1$ as following figure 4.

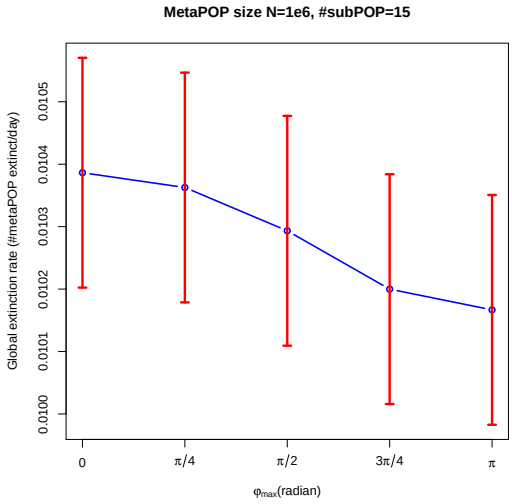

Figure 4 – Estimated th global etinction rates in the metapopulation of the 15 subpopulations after 100 different simulations $N = 10^6$, coupling rate $\rho = 0.1$. Here, with 95% confidence interval, the red lines and the red points are respectively the confidence intervals and the estimated rates for the global extinction rate of each value of φ_{max}.

This figure 4 shows to us that the amplitude of the confidence intervals for each value of φ_{max} are quite far to each other. Furthermore, it goes down robustly when φ_{max} runs from 0 to π. The phase difference strongly influences the global disease extinction rate. Figure 4 indicates the trend of the extinction rate with decreasing the level of asynchrony. The asynchrony between subpopulations is the main reason why the infectious disease goes extinct in the slow way.

3.3 Influence of other parameters on global extinction rate

3.3.1 Number of subpopulation in a metapopulation

In this part, we develop the relation between the local extinction rate and the number of subpopulations in a metapopulation. We perform simulations with metapopulations from two to 45 subpopulations, the size of metapopulation fixed is $N = 1e + 06$ and $\varphi_{max} = \pi$. Here, we use the stationaire distribution for metapopulations. In this case, all subpopulations in the metapopulation is the same initial values of variables. The result is as follows:

We find that, the local extinction rate of synchrony is larger than that of asyschronie. Because when the subpopulations are in asynchrony, the phase difference among subpopulations causes recolonization among the subpopulations.

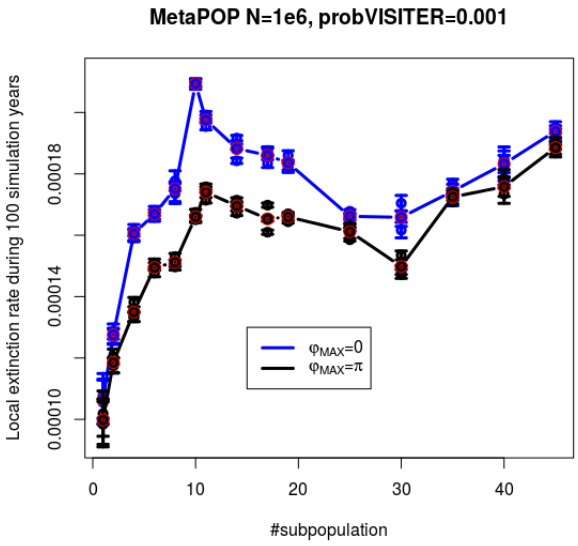

Figure 5 – Estimated local extinction rates in metapopulations that have the same metapopulation size but the number of subpopulation is different after 100 different simulations. The metapopulation size fixed $N = 1 \times 10^6$, coupling rate $\rho = 0.001$ and $\varphi_{max} = \pi$.

So the metapopulation is difficult to obtain local extinction. In addition, the distance between the rate of synchrony and asynchrony is more and more decreased when the number of subpopulation increases. Because with φ_{max} fixed, the phase difference between the subpopulation decreases due to the increase of the number of subpopulation, the subpopulations tend to fluctuate identically. Moreover, when the number of subpopulations in the metapopulation is small (from 1 to 10), the local extinction rate increases, because in this case the subpopulation size is large, the phase difference between subpopulation is great too. So, the time to maintain local extinction is great, and the total of the local extinction in the metapopulation increases due to the increase of the number of sub-populations. Then, when the number of subpopulations in the metapopulation is average (from 10 to 30). In this case, the subpopulation size begins decreasing. However, with the fixed value of φ_{max}, the phase difference between the subpopulations is decreased, the subpopulations tend to resemble. Therefore, the total of local extinction clearly increases, and the sum of the local extinction duration in the metapopulation is large due to the large number of subpopulations in the metapopulation. Hence, the slope of the profile "bell" in the histogram of the vector that contains the sum of the local extinction time for the metapopualtion, is very soft. This is the reason for that the profile estimated goes down. Finally, when the number of subpopulation is high (over 30), the metapopulation becomes very complicated. Luckily, in this case, for the fixed metapopulation size and the value of φ_{max} fixed, then the size of each subpopulation is very small and it seems that the subpopulations are in synchrony. Therefore, the total of the local extinction goes up due to the increase of the number of subpopulation, but the maximume value of the local extinction duration is small. Therefore, the local extinction rate increases.

3.3.2 Influence of the metapopulation size

In the metapopulation of 05 subpopulations, we implemented experiences with different population sizes of subpopulation. We performed 100 different simulations for the metapopulation of 05 subpopulations in which their sizes follow a stationary distribution of the metapopulation size N. We set N from 5×10^5 to 5×10^6 individuals. The result (Figure 6) affirms that the size N influences strongly the global persistence of an infectious disease in a metapopulation.

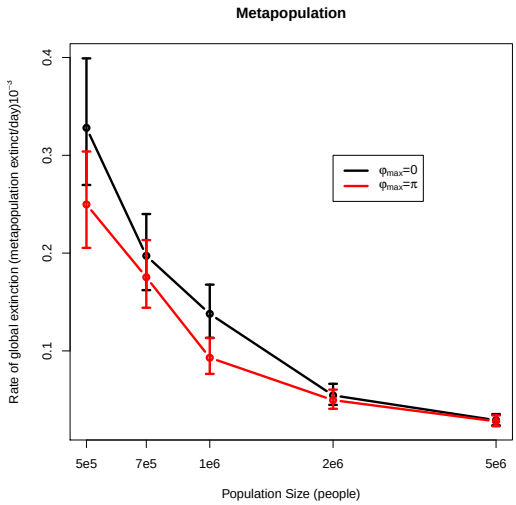

Figure 6 – The relation between the metapopulation size and the global extinction rate for the metapopulation of five subpopulations.

From the figure 6, we find that the mass extinction rate goes down when the metapopulation size augments, at the same time, these rates decreases when the asynchrony φ_{max} goes up. It is obvious that the metapopualtion size is big, then the time of disease persistence augments, then the mass extinction rates is declined..

3.3.3 Coupling rate

One more factor that was pointed is coupling strength between subpopulations. Here, the coupling rate or the dispersal rate ρ can be considered as migration strength. The disease transmission speed grows fast when coupling rate goes up in metapopulations, but the global extinction rate is inverse (Figure 7). In this part, we change coupling rate from weak to strong in a metapopualtion of five subpopulations with the metapopulation size $N = 10^6$. The dispersal rate ρ is divided into three intervals. These are low, intermediate and high coupling rate intervals. In each interval, we chose some coupling rates that highlight the coupling strength among subpopulations in a metapopulation. When the coupling rate is small from 0.0 to 0.005, the mass extinction rate decreases very slowly. Because, in this case, the subpopulations seem to be independent. They fluctuate independently. They are easy to go extinct. We are also easy to find the mass extinction in the metapopulation. However, this extinction rate is declined in a sudden way when the coupling rate is modified from 0.01 to 0.3. Lastly, the extinction rate increases when the coupling rate is so robust from

0.5 to 1.0. Based on this figure, the mass extinction rate in a metapopulation is one inverse bell for the coupling rate. The medium coupling rate (from 0.01 to 0.1) minimizes the mass extinction rate in metapopulation. As in the case of the small and average coupling rates, the coupling rate and the speed of migration among subpopulations are directly proportional. The dispersal speed increases, thereby the local recolonization speed rises, the duration of persistence grows. However, this trend of global extinction rate with decreasing coupling rate, is not right any more when the dispersal rate is strong. The duration of persistence falls, because the metapopulation has tendency to become one big population. In this case, the phase difference or the recolonization among subpopulations are no longer significant.

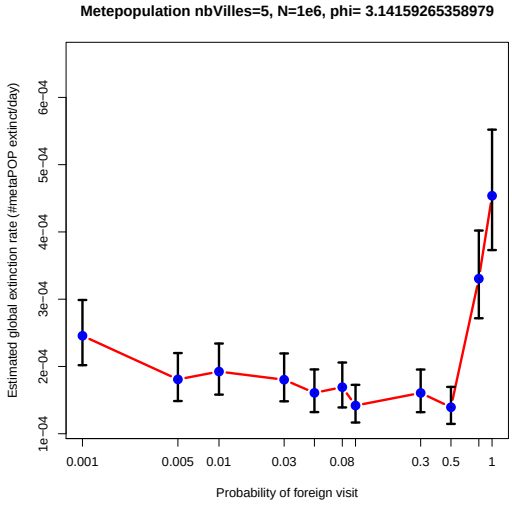

Figure 7 – Correlation between the coupling rate and the mass extinction rate in the metapopulation of five subpopulations. Here, the coupling rate ρ runs from 0 to 1. The level of asynchrony φ_{max} is π and the metapopulation size N=1e6.

4 Discussion and Conclusion

We successfully have built a version for the susceptible-infectied-recovered stochastic metapopulation model. The infection rate λ_i for $subpopulation_i$ has portrayed all effects inside as well as outside of the disease tranmission chain between individuals in the same subpopulation or in other subpopulations. Moreover, our metapopulation model became more detailed when we brought seasonality in metapopulation model to create periodic transmission in year that highlighted seasonal changes as well as school period of children [6, 5, 13, 30]. We have metapopulation model with different contact rates for each subpopulation. This is a more complex model than any used metapopulation model. We have sketched successfully in-phase and sometime out-of-phase ("antiphase") models across suburbs of He's 2003 [23]

This complex metapopulation model is also an expected result of Rozhnova(2012) [39]. It's a good result for scientists wanting to use the SEIR metapopulation model for simulating dynamics of infectious diseases. Our results roughly

support those of Rozhnova's 2012. The authors gave different values of the contact rates β of each subpopulation. However, the rates β here are fixed by constants and the number of subpopulations in experiences are maximum of three. Comparing this result with our's, in a coupling metapopulation, the degree of synchorny is maintained when the coupling rate between subpopulations is weak.

Moreover, our stochastic SEIR metapopulation model with subpopulations connected to each other, we have quantified disease persistence of seasonality as well as spatial synchrony. With our model, we can easily create level of seasonality in year and at the same time, phase difference in seasonality between subpopulations. It's the reason why we have model quite close to the metapopulation model in reality.

Due to the phase difference between the number of contact κ, we can change by an increase or a decrease in level of synchrony. We want to decrease level of synchrony, by simply increasing the phase difference between forcing phase coefficients in the formulas of the number of contact κ. Clearly, the level of synchrony between two subpopulations are the worst when the two fluctuations are in antiphase (as figure 2). When the phase difference between oscillations increases, the desynchronizing effect on population dynamics of the subpopulations augments. This enhances disease persistence time though the global extinction rate is reduced. Moreover, as the result above (figure 4), in the local shape, our result, along with those of Bolker (1995) [5] and Heino (1997) [24], stress the number of local extinctions being inversely proportionnel with the global extinction rate in a metapopulation. When the level of synchrony is at a reduction and the global persistence time gets an increase, the global extinction rate of metapopulation goes down and the number of local extinction goes up. Due to the result about local extinction, we also affirm that disease is always available in metapopulation if and only if at least one subpopulaiton is not extinct.

Our finding has specified the two main factors influencing the persistence ability of an infectious disease. One factor is transmission characteristics of the infectious disease and the other is interplay between mixing subpopulations in metapopulation. The interaction between the disease persistence and the spatial heterogeneity becomes a major key to unlock questions about infectious disease in epidemiology. This result takes a large part in epidemic disease persistence domain that has being exploited in scientific epidemically research works. We gained a robust understanding of how disease persistence is affected by local factors such as spatial heterogeneity, demographic asynchrony and seasonality, as well as mixing factors such as migration, disease transmission between hosts and pathogens. Lastly, we also highlighted recolonization effects. It is like rescue for disease. Because of connection between subpopulation, individuals can go everywhere. Subpopulation is quickly re-infected althought the disease has became extinct. Thus, the disease rescus makes local extintions difficult to extend into global extinctions.

As a matter of the fact of coupling strength among subpopulations in metapopulation, we proved that the curve of the global extinction rate is a convex function following the coupling rate. Inversely, the persistence of the disease in the metapopulation is an exponential survival function over time and is a humped function for the coupling rate. In addition, when the coupling rate between subpopulations is just medium where the mass extinction rate is minimum and

the global disease persistence in metapopulation is maximum . This finding is similar to those of Huffaker(1958) [27], Holyoak and Lawler(1996) [25], and Yaari et al. (2012) [45] when they exhaustively explored the disease persistence behavior of many different metapopulation models. And our result one more time affirms that the interaction in metapopulation models plays a big role when the interaction strength ρ is from 10^{-3} to 0.1 [31].

To summarize, we have built successfully a sinusoidally forced SEIR stochastic metapopulation model. This model is like a physical system of coupled oscillators. We have pointed out that spatial synchronization consistently and predictably makes extinction risk increase by using the model 0 where all the subpopulations have the same population size N and there is no explicit spatial distance. So, it's good for the future, we can continue this work with different population size of each subpopulation and different spatial distance between subpopulations and then, create synchronous metapopulations that optimize vaccination policies.

References

[1] S. Altizer, A. Dobson, P. Hosseini, P. Hudson, M. Pascual, and P. Rohani. Seasonality and the dynamics of infectious diseases. *Ecol Lett*, 9(4):467–484, Apr 2006. 1, 2.1.1

[2] R. M. Anderson and R. M. May. *Infectious Diseases of Humans: Dynamics and Control*. Oxford University Press, 1992. 2.1.1

[3] R. M. Anderson and R. M. May. *Infectious Diseases of Humans: Dynamics and Control*. Oxford University Press, 1992. 2.3.4

[4] Sue Binder, Alexandra M Levitt, Jeffrey J Sacks, and James M Hughes. Emerging infectious diseases: public health issues for the 21st century. *Science*, 284(5418):1311–1313, 1999. (document)

[5] B. Bolker and B. Grenfell. Space, persistence and dynamics of measles epidemics. *The Royal Society*, 348:309–320, 1995. 1, 2.1.1, 2.1.3, 2, 4

[6] B. M. Bolker and B. T. Grenfell. Chaos and biological complexity in measles dynamics. *Proc Biol Sci*, 251(1330):75–81, Jan 1993. 4

[7] B. M. Bolker and B. T. Grenfell. Impact of vaccination on the spatial correlation and persistence of measles dynamics. *Proc Natl Acad Sci U S A*, 93(22):12648–12653, Oct 1996. 1

[8] M. Choisy and P. Rohani. Changing spatial epidemiology of pertussis in continental usa. *Proc Biol Sci*, 279(1747):4574–4581, Nov 2012. 1, 2

[9] A. J. K. Conlan and B. T. Grenfell. Seasonality and the persistence and invasion of measles. *Proc Biol Sci*, 274(1614):1133–1141, May 2007. 1

[10] A. J. K. Conlan, P. Rohani, A. L. Lloyd, M. Keeling, and B. T. Grenfell. Resolving the impact of waiting time distributions on the persistence of measles. *J R Soc Interface*, 7(45):623–640, Apr 2010. 1, 2.2.2, 2

[11] G. S. Cumming. The impacts of low-head dams on fish species richness in wisconsin, usa. *Ecological Applications*, 14:1495–1506., 2004. 1

[12] S. Dey and A. Joshi. Stability via asynchrony in drosophila metapopulations with low migration rates. *Science*, 312(5772):434–436, Apr 2006. 1

[13] D. J. Earn, P. Rohani, and B. T. Grenfell. Persistence, chaos and synchrony in ecology and epidemiology. *Proc Biol Sci*, 265(1390):7–10, Jan 1998. 1, 4

[14] M. J. Ferrari, A. Djibo, R. F. Grais, N. Bharti, B. T. Grenfell, and O. N. Bjornstad. Rural-urban gradient in seasonal forcing of measles transmission in niger. *Proc Biol Sci*, 277(1695):2775–2782, Sep 2010. 1

[15] Ilkka Hanski; Oscar E Gaggiotti. *Ecology, Genetics and evolution of metapopulations*. 2004. 1, 2.1.3

[16] Daniel T Gillespie. Exact stochastic simulation of coupled chemical reactions. *The journal of physical chemistry*, 81(25):2340–2361, 1977. (document), 1, 2.1.2, 2.1.2, 2.3.4

[17] B. T. Grenfell, O. N. Bjørnstad, and J. Kappey. Travelling waves and spatial hierarchies in measles epidemics. *Nature*, 414(6865):716–723, Dec 2001. 1

[18] B.T. Grenfell, B. M. Bolker, and A. Klegzkowski. Seasonality and extinction in chaotic metapopulation. *The royal society*, 259:97–103, 1995. 2.1.1

[19] B. D. Griffen and J. M. Drake. Environment, but not migration rate, influences extinction risk in experimental metapopulations. *Proc Biol Sci*, 276(1677):4363–4371, Dec 2009. 1

[20] C. E. Gunning and H. J. Wearing. Probabilistic measures of persistence and extinction in measles (meta)populations. *Ecol Lett*, 16(8):985–994, Aug 2013. 1, 2.2.2, 2.3.4

[21] T. J. Hagenaars, C. A. Donnelly, and N. M. Ferguson. Spatial heterogeneity and the persistence of infectious diseases. *J Theor Biol*, 229(3):349–359, Aug 2004. 1

[22] I. Hanski. Metapopulation dynamics. *Nature*, 396, 1998. 1

[23] D. He and L. Stone. Spatio-temporal synchronization of recurrent epidemics. *Proc Biol Sci*, 270(1523):1519–1526, Jul 2003. 4

[24] M. Heino, V. Kaitala, E. Ranta, and J. Lindstrom. Synchronous dynamics and rates of extinction in spatially structured populations. *The Royal Society*, 264:481–486, 1997. 1, 4

[25] M. Holyoak and S. P. Lawler. Persistence of an extinction-prone predator-prey interaction through metapopulation dynamics. *Ecology*, pages 1867–1879, 1996. 4

[26] P. R. Hosseini, A. A. Dhondt, and A. Dobson. Seasonality and wildlife disease: how seasonal birth, aggregation and variation in immunity affect the dynamics of mycoplasma gallisepticum in house finches. *Proc Biol Sci*, 271(1557):2569–2577, Dec 2004. 1

[27] C. B. Huffaker. Experimental studies on predation: dispersion factors and predator-prey oscillations. *Hilgardia*, 27:343–383, 1958. 4

[28] Kate E Jones, Nikkita G Patel, Marc A Levy, Adam Storeygard, Deborah Balk, John L Gittleman, and Peter Daszak. Global trends in emerging infectious diseases. *Nature*, 451(7181):990–993, 2008. (document)

[29] M. J. Keeling and B. T. Grenfell. Disease extinction and community size: modeling the persistence of measles. *Science*, 275(5296):65–67, Jan 1997. 2

[30] M. J. Keeling and B. T. Grenfell. Understanding the persistence of measles: reconciling theory, simulation and observation. *Proc Biol Sci*, 269(1489):335–343, Feb 2002. 2.1.2, 2.2.2, 2, 4

[31] M. J. Keeling and P. Rohani. *Modeling Infectious Diseases in Humans and Animals*. Princeton University Press, 2011. 1, 2.1.1, 2, 4, 6.3, 6.3.2

[32] David G. Kleinbaum. *Survival analysis*. 2005. 2.2.2

[33] R. Levins. Some demographic and genetic consequences of environmental heterogeneity for biological control. *Bulletin of the Entomological Society of America*, 15:237–240, 1969. 1, 2.1.3

[34] A. L. Lloyd. Realistic distributions of infectious periods in epidemic models: changing patterns of persistence and dynamics. *Theor Popul Biol*, 60(1):59–71, Aug 2001. 2.1.2

[35] David M Morens, Gregory K Folkers, and Anthony S Fauci. The challenge of emerging and re-emerging infectious diseases. *Nature*, 430(6996):242–249, 2004. (document)

[36] Jonathan A Patz, Paul R Epstein, Thomas A Burke, and John M Balbus. Global climate change and emerging infectious diseases. *Jama*, 275(3):217–223, 1996. (document)

[37] E. Renshaw. *Modelling biological populations in space and time*, volume 11. Cambridge University Press, 1993. 2.1.2

[38] P. Rohani, D. J. Earn, and B. T. Grenfell. Opposite patterns of synchrony in sympatric disease metapopulations. *Science*, 286(5441):968–971, Oct 1999. 1

[39] G. Rozhnova, A. Nunes, and A. J. McKane. Phase lag in epidemics on a network of cities. *Phys Rev E Stat Nonlin Soft Matter Phys*, 85(5 Pt 1):051912, May 2012. 1, 4

[40] G. Rozhnova1, A. Nunes, and A. J. McKane. Impact of commuting on disease persistence in heterogeneous metapopulations. 2013. 1

[41] D. J. Smith, A. S. Lapedes, and J. C. de Jong. Mapping the antigenic and genetic evolution of influenza virus. *Science 305, 371 (2004);*, 2004. 1

[42] Fred C Tenover and James M Hughes. The challenges of emerging infectious diseases: development and spread of multiply-resistant bacterial pathogens. *Jama*, 275(4):300–304, 1996. (document)

[43] T. M. Therneau. *A Package for Survival Analysis in S*, 2014. R package version 2.37-7. (document), 2.2.2

[44] C. Viboud, O. N. Bjørnstad, D. L. Smith, L. Simonsen, M. A. Miller, and B. T. Grenfell. Synchrony, waves, and spatial hierarchies in the spread of influenza. *science*, 312(5772):447–451, 2006. 1

[45] G. Yaari, Y. Ben-Zion, N. M. Shnerb, and D. A. Vasseur. Consistent scaling of persistence time in metapopulations. *Ecology*, 93(5):1214–1227, May 2012. 1, 4

[46] G. Yan, Z. Q. Fu, J. Ren, and W. X. Wang. Collective synchronization induced by epidemic dynamics on complex networks with communities. *Phys Rev E Stat Nonlin Soft Matter Phys*, 75(1 Pt 2):016108, Jan 2007. 1

5 Appendix : equilibrium values of the system 1–4

We start with ordinary differential equations for a $subpopulation_i$ in a metapopulation as follows:

$$\frac{dS_i}{dt} = \mu N_i - \lambda_i S_i - \mu S_i \tag{15}$$
$$\frac{dE_i}{dt} = \lambda_i S_i - \mu E_i - \sigma E_i \tag{16}$$
$$\frac{dI_i}{dt} = \sigma E_i - \mu I_i - \gamma I_i \tag{17}$$
$$\frac{dR_i}{dt} = \gamma I_i - \mu R_i \tag{18}$$

In simulation, we know that the equilibrium state allow a disease to persist in a population for a long time. So, an infectious disease in the $subpopulation_i$ is available in long term this system is at equilibrium. It means that at which $\frac{dS_i}{dt} = \frac{dE_i}{dt} = \frac{dI_i}{dt} = \frac{dR_i}{dt} = 0$ (*). Thus, we let all equations (equations 15 - 18) in the system be equal to zero, then calculate the values of the variables (now denoted by S_i^*, E_i^*, I_i^* , and R_i^*) that satisfy this condition (*). We have these values as folllows:

$$S_i^* = N_i \frac{(\gamma+\mu)(\sigma+\mu)}{\beta\sigma} \tag{19}$$
$$E_i^* = N_i\mu\left(\frac{1}{\sigma+\mu} - \frac{\gamma+\mu}{\beta\sigma}\right) \tag{20}$$
$$I_i^* = N_i\mu\frac{\beta\sigma - (\sigma+\mu)(\gamma+\mu)}{\beta(\sigma+\mu)(\gamma+\mu)} \tag{21}$$
$$R_i^* = N_i - S_i^* - E_i^* - I_i^* \tag{22}$$

Here, if we set $R_0 = \frac{\beta\sigma}{(\gamma+\mu)(\sigma+\mu)}$, so we have

$$S_i^* = N_i \frac{1}{R_0} \tag{23}$$
$$E_i^* = N_i \frac{\mu\sigma}{R_0}(R_0 - 1) \tag{24}$$
$$I_i^* = N_i \frac{\mu}{\beta}(R_0 - 1) \tag{25}$$
$$R_i^* = N_i - S_i^* - E_i^* - I_i^* \tag{26}$$

One nomal conditions for all population availabes is that the equilibrium values cannot be negative. Therefore, an infectious disease is available in the $subpopulation_i$ if $R_0 > 1$. Now, the endemic equilibrium in the system is given by

$(S_i^*, E_i^*, I_i^*, R_i^*) = (N_i \frac{1}{R_0},\, N_i \frac{\mu\sigma}{R_0}\,(R_0 - 1),\, N_i \frac{\mu}{\beta}(R_0 - 1),\, N_i(1 - \frac{1}{R_0} - \frac{\mu\sigma}{R_0}\,(R_0 - 1) - \frac{\mu}{\beta}(R_0 - 1)).$

6 Appendix: derivation of the equation 5

Here, we will point out that the contact rate β is a function of the average contact number per unit of time and the probability of successful disease transmission following a contact.

Definition 1. During the small time interval δt, each individual native of the subpopulation i visits one single subpopulation j (with the probability ρ_{ij}) and will see in average κ_j individuals. These individuals come from all the cities.

6.1 Notation :

Here, we present list of sets and events describing the state of the system at time t :

- C_i is the set of all individuals born in subpopulation i.
- $V_{i,t}$ is the set of all individuals physically located in subpopulation i from time t to time $t + \delta t$. This includes foreigners traveling in subpopulation i at time t, and all natives from subpopulation i which are not traveling abroad at time t.
- S_t, E_t, I_t, R_t are the sets of all individuals respectively susceptible, exposed, infected and recovered at time t. Note that these set include individuals from all subpopulations.
- $S_{i,t}, E_{i,t}, I_{i,t}, R_{i,t}$ are the same sets, restricted to natives of subpopulation i. So formally, $S_{i,t} = S_t \cap C_i$, $E_{i,t} = E_t \cap C_i$, $I_{i,t} = I_t \cap C_i$, and $R_{i,t} = R_t \cap C_i$.
- $Transmit(y, x)$ is an event indicating that individual x gets infected by individual y which was already infected
- $c_{i,k}$ is the probability that a susceptible individual native from i being in contact with another infected individual native from k gets infected.
- κ_j is the average number of contacts per unit of time a susceptible will have when visiting subpopulation j.
- ξ_{jk} refers to the probability that an individual y meeting x in C_j comes from C_k.
- $\rho_{i,j}$, the probability that an individual from subpopulation i visits subpopulation j. Of course, $\sum_{j=1}^{M} \rho_{ij} = 1$.

Proposition 1. *The coefficient κ should also depend on i, because an individual native from subpopulation i meets more people in his own subpopulation than abroad ($\kappa_{i,i} > \kappa_{i,j}$).*

6.2 The background

One general question is always posed "how does the population of exposed individuals of subpopulation i evolve ?". For the sake of simplicity, in the process of transmission of the SEIR model, we focus on the incidence and we assume for now that the latent period and the recovery rate, repectively $\mu = \sigma = 0$. Thus, we write a probabilistic formulation of $\frac{dE_i}{dt}$. Assuming the time is discrete, we have $\frac{dE_i}{dt} \approx \mathbb{E}\left[E_{i,t+1} \setminus E_{i,t}\right]$. Then,

$$\begin{aligned}
\mathbb{E}\left[E_{i,t+1} \setminus E_{i,t}\right] &= \mathbb{E}\left[E_{i,t+1} \cap S_{i,t}\right] \\
&= \sum_{x \in C_i} Pr\left[x \in E_{t+1} \wedge x \in S_t\right] \\
&= \sum_{x \in C_i} Pr\left[x \in S_t\right] * Pr\left[x \in E_{t+1} \mid x \in S_t\right] \\
&= Pr_{x \sim \mathcal{X}_i}\left[x \in E_{t+1} \mid x \in S_t\right] * \sum_{x \in C_i} Pr\left[x \in S_t\right] \\
&= |S_{i,t}| \times Pr_{x \sim \mathcal{X}_i}\left[x \in E_{t+1} \mid x \in S_t\right]
\end{aligned}$$

Assume there are M cities. An individual x of the subpopulation i may be visiting another subpopulation, or staying in its own subpopulation. Applying the law of total probabilities, we get:

$$\begin{aligned}
Pr_{x \sim \mathcal{X}_i}\left[x \in E_{t+dt} \mid x \in S_t\right] &= \sum_{j=1}^{M} Pr_{x \sim \mathcal{X}_i}\left[x \in E_{t+dt} \wedge x \in V_{j,t} \mid x \in S_t\right] \\
&= \sum_{j=1}^{M} Pr_{x \sim \mathcal{X}_i}\left[x \in E_{t+dt} \mid x \in S_t \wedge x \in V_{j,t}\right] . Pr_{x \sim \mathcal{X}_i}\left[x \in V_{j,t}\right] \\
&\quad \sum_{j=1}^{M} Pr_{x \sim \mathcal{X}_i}\left[x \in E_{t+dt} \mid x \in S_t \wedge x \in V_{j,t}\right] \times \rho_{ij}
\end{aligned}$$

Where $\rho_{i,j} = Pr_{x \sim \mathcal{X}_i}\left[x \in V_{j,t}\right]$, the probability that an individual from subpopulation i visits subpopulation j. Of course, $\sum_{j=1}^{M} \rho_{ij} = 1$.

6.3 Study of case where agent x native from subpopulation i visits subpopulation j

Here, we look at the probability that a susceptible $x \sim \mathcal{X}_i$ visiting j gets infected or not after δt time steps. Let $\mathcal{Y}$ be the uniform distribution over $V_{j,t}$. The correct mathematical approach for this would be to assume that for each subpopulation k, the number of people native from k that we meet during δt follows a Poisson process. So both the number of people we meet and the number of infected people we meet during δt should be random variables.

In the approach described in [31], the authors did not do this. They assumed that both the number of people we

meet and the number of infected people we meet *are fixed* (otherwise the maths they write would have been different). We will call this the old interpretation of the infection force (for short, OIIF) interpretation that we will present it in the following parts.

We introduce an alternative approximation, where we assume that the number κ of people we meet during δt is *fixed*, but each of these people has *some probability* to be infected. This is an *in-between interpretation*, easier than the Poisson process maths, but better than the OIIF. We will call this the new interpretation of the infection force (for short NIIF).

6.3.1 The interpretation NIIF

Proposition 2. *Agent x meets* exactly κ_j *other individuals, and each of these individuals has a probability $\frac{|I_{k,t}|}{N_k}$ of being infected, where k is its native subpopulation. Let $y_1 \dots y_{\kappa_j}$ be the individuals that x meets. We get:*

$$\begin{aligned} & Pr_{x\sim\mathcal{X}_i}\left[x \in S_{t+\delta t} \mid x \in S_t \wedge x \in V_{j,t}\right] \\ = \; & Pr_{x\sim\mathcal{X}_i, y_1 \dots, y_{\kappa_j}\sim\mathcal{Y}}\left[\bigwedge_{p=1}^{\kappa_j} \neg\left(y_p \in I_t \wedge Transmit(y_p, x)\right) \mid x \in S_t \wedge x \in V_{j,t}\right] \end{aligned}$$

So we have:

$$\begin{aligned} & Pr_{x\sim\mathcal{X}_i}\left[x \in S_{t+\delta t} \mid x \in S_t \wedge x \in V_{j,t}\right] \\ = \; & Pr_{x\sim\mathcal{X}_i, y\sim\mathcal{Y}}\left[\neg\left(y \in I_t \wedge Transmit(y, x)\right) \mid x \in S_t \wedge x \in V_{j,t}\right]^{\kappa_j \delta t} \end{aligned}$$

Moreover, we have:

- the probability so that a susceptible individual x is infected by an infected individual y :

$$\begin{aligned} & Pr_{x\sim\mathcal{X}_i, y\sim\mathcal{Y}}\left[y \in I_t \wedge Transmit(y, x) \mid x \in S_t \wedge x \in V_{j,t}\right] \\ = \; & \sum_{k=1}^{M} Pr_{x\sim\mathcal{X}_i, y\sim\mathcal{Y}}\left[y \in I_t \wedge Transmit(y, x) \mid x \in S_t \wedge x \in V_{j,t} \wedge y \in C_k\right] . Pr_{y\sim\mathcal{Y}}\left(y \in C_k\right) \\ = \; & \sum_{k=1}^{M} \{Pr_{x\sim\mathcal{X}_i, y\sim\mathcal{X}_k}\left[y \in I_t \mid x \in S_t \wedge x \in V_{j,t}\right] \\ & \quad \times Pr_{x\sim\mathcal{X}_i, y\sim\mathcal{X}_k}\left[Transmit(y, x) \mid y \in I_t \wedge x \in S_t \wedge x \in V_{j,t} \wedge y \in C_k\right] \times Pr_{y\sim\mathcal{Y}}\left(y \in C_k\right)\} \\ = \; & \sum_{k=1}^{M} \left(\frac{|I_{k,t}|}{N_k} \times c_{ik} \times \xi_{jk}\right) \end{aligned}$$

$\xi_{jk} = \frac{N_k \rho_{kj}}{\sum_{v=1}^{M} N_v \rho_{vj}}$ refers to the probability that an individual y meeting x in C_j comes from C_k.

- hence, the probability so that a susceptible individual x is not infected by an infected individual y :

$$1 - \sum_{k=1}^{M} \left(\frac{|I_{k,t}|}{N_k} \times c_{ik} \times \xi_{jk} \right)$$

- thereby, the probability so that a susceptible individual x is not infected after κ_j contacts per unit time δt.

$$\left[1 - \sum_{k=1}^{M} \left(\frac{|I_{k,t}|}{N_k} \times c_{ik} \times \xi_{jk} \right) \right]^{\kappa_j \delta t}$$

- thus, the probability so that a susceptible individual x becomes infected after κ_j contacts per unit time δt.

$$Pr_{x \sim \mathcal{X}_i} \left[x \in E_{t+\delta t} \mid x \in S_t \wedge x \in V_{j,t} \right] = \left[1 - \sum_{k=1}^{M} \left(\frac{|I_{k,t}|}{N_k} \times c_{ik} \times \xi_{jk} \right) \right]^{\kappa_j \delta t}$$

We now apply the *log* approximation which consists in approximating $1 - (1-u)^v$ by $v \log(1-u)$:

$$Pr_{x \sim \mathcal{X}_i} \left[x \in E_{t+\delta t} \mid x \in S_t \wedge x \in V_{j,t} \right] = -\kappa_j \delta t \log \left[1 - \sum_{k=1}^{M} \left(\frac{|I_{k,t}|}{N_k} \times c_{ik} \times \xi_{jk} \right) \right]$$

So, the transmission rate per susceptible individual is as follows :

$$\frac{dPr_{x \sim \mathcal{X}_i} \left[x \in E_{t+dt} \mid x \in S_t \wedge x \in V_{j,t} \right]}{dt} \simeq -\kappa_j \log \left[1 - \sum_{k=1}^{M} \left(\frac{|I_{k,t}|}{N_k} \times c_{ik} \times \xi_{jk} \right) \right]$$

In fact, we use the parameter λ to present this quantity, and it is denoted as the "force of infection" :

$$\lambda_i = \sum_j \rho_{ij} \kappa_j \log \left[1 - \sum_{k=1}^{M} \left(\frac{|I_{k,t}|}{N_k} \times c_{ik} \times \xi_{jk} \right) \right]$$

If there is only one subpopulation i, then

$$\lambda_i = \kappa_j log(1 - \frac{|I_i|}{N_i} \times c_{ii})$$

6.3.2 The interpretation OIIF

Proposition 3. *Agent x meets* exactly $\kappa_j \delta t \xi_{jk} \frac{|I_{k,t}|}{N_k}$ *other infected individuals native from subpopulation k.*

Let $l_k = \kappa_j \delta t \xi_{jk} \frac{|I_{k,t}|}{N_k}$.

Let $y_1^k \dots y_{l_k}^k$ be the infected individuals native from k that our individual x meets between t and $t + \delta t$.

We have the probability so that a susceptible individual x is not infected after having seen l_k individuals between t and $t + \delta t$:

$$\begin{aligned}
& Pr_{x\sim\mathcal{X}_i}\left[x \in S_{t+\delta t} \mid x \in S_t \wedge x \in V_{j,t}\right] \\
= \; & Pr_{x\sim\mathcal{X}_i}\left[\bigwedge_{\substack{k=1\ldots M \\ p=1\ldots l_k}} \neg\left(Transmit(y_p^k, x)\right) \mid x \in S_t \wedge x \in V_{j,t}\right] \\
= \; & \prod_{k=1}^{M} Pr_{x\sim\mathcal{X}_i}\left[\bigwedge_{p=1\ldots l_k} \neg\left(Transmit(y_p^k, x)\right) \mid x \in S_t \wedge x \in V_{j,t}\right] \\
= \; & \prod_{k=1}^{M} (1 - c_{ik})^{\kappa_j \delta t \xi_{jk} \frac{|I_{k,t}|}{N_k}}
\end{aligned}$$

Then, we plug this back into the previous formula, and we get:

$$Pr_{x\sim\mathcal{X}_i}\left[x \in E_{t+\delta t} \mid x \in S_t \wedge x \in V_{j,t}\right] \quad = \quad 1 - \prod_{k=1}^{M} (1 - c_{ik})^{\kappa_j \xi_{jk} \frac{|I_{k,t}|}{N_k} \delta t}$$

The first order approximation of $1 - \prod_{k=1}^{M}(1 - c_{ik})^{v_k}$ is $\sum_{k=1}^{M} -v_k \log(1 - c_{ik})$. Applying this approximation here, we get:

$$Pr_{x\sim\mathcal{X}_i}\left[x \in E_{t+\delta t} \mid x \in S_t \wedge x \in V_{j,t}\right] \quad \simeq \quad \delta t \sum_{k=1}^{M} \left(-\kappa_j \xi_{jk} \frac{|I_{k,t}|}{N_k} \log\left(1 - c_{ik}\right)\right)$$

Define $\beta_{ijk} = -\kappa_j \log\left(1 - c_{ik}\right)$, let δt converge to zero, and we get:

$$\frac{dPr_{x\sim\mathcal{X}_i}\left[x \in E_{t+dt} \mid x \in S_t \wedge x \in V_{j,t}\right]}{dt} \simeq \sum_{k=1}^{M} \left(\xi_{jk} \frac{|I_{k,t}|}{N_k} \beta_{ijk}\right)$$

If there is only one subpopulation i, then we fall back to the formula of [31]. We have :

$$\beta_i = -\kappa_i \log\left(1 - c_i\right)$$

$$\frac{d}{dt}\mathbb{E}\left[|E_{i,t+dt} - E_{i,t}|\right] \simeq -|S_{i,t}|\left(\frac{|I_i|}{N_i}\beta_i\right)$$

and the force of infection as follows :

$$\lambda_i = \beta_i \frac{|I_i|}{N_i}$$

6.4 Final Formula

We simply have to plug in the probability ρ_{ij} that i visits j.

We get, for the interpretation NIIF :

$$\frac{d}{dt}\mathbb{E}\left[|E_{i,t+dt} - E_{i,t}|\right] \simeq -|S_{i,t}| \sum_j \rho_{ij}\kappa_j \log\left[1 - \sum_{k=1}^{M}\left(\frac{|I_{k,t}|}{N_k} \times c_{ik} \times \xi_{jk}\right)\right]$$

And for the interpretation OIIF:

$$\frac{d}{dt}\mathbb{E}\left[|E_{i,t+dt} - E_{i,t}|\right] \simeq -|S_{i,t}| \sum_j \rho_{ij} \sum_{k=1}^{M}\left(\xi_{jk}\frac{|I_{k,t}|}{N_k}\beta_{ijk}\right)$$