The 5 hallmarks of stochastic SIR and meta-populations

Introduction to Computational Science

Jorrim Prins
Universiteit van Amsterdam
11038934

Isabelle Brakenhoff
Universiteit van Amsterdam
11283912

1 INTRODUCTION

While the current COVID-19 pandemic has captured global attention, the importance of mathematically modelling infectious disease and epidemics sincerely exceeds the scope of this sole problem. Diseases like influenza, Ebola and measles still cause epidemics on a regular basis and understanding the dynamics of the spread of infectious disease could possibly help avoid an epidemic outbreak or decrease the burden on caretakers. Modelling of infectious disease has been done for almost a century, since the early work by Kermack and McKendrick [7], and helps identify spread mechanisms, predict future disease developments and control epidemics. The early work by Kermack and McKendrick [7] provides a framework for compartment models presently known as SIR, SIS and SEIR, these models have been adjusted and extended to fit real-world situations well. Current research is still heavily reliant on these models as can be seen in COVID-19 related work by Acemoglu et al. [1], Chen et al. [3], Cooper et al. [4].

We examine some additions and modifications on the standard deterministic SIR model and the effects of implementation thereof. Stochasticity is included into the dynamics by following an event-driven approach of modelling SIR, observing the rate of specified events instead of state variables. Defining the model in a stochastic way leads to various changes in the dynamics that can be explored by focusing on five key features that distinguish them from deterministic counterparts [6], we call these the five hallmarks. Furthermore, the assumption of homogeneous mixing of the population is relaxed by implementing a meta-population model, investigating the effect of the number of populations and the strength of interaction on disease dynamics.

2 THEORETICAL BACKGROUND

As the name of the model already implies, the SIR model consists of three compartments defined as Susceptible, Infectious and Recovered. These compartments represent the fraction of total population in each compartment, we will redefine the compartments as X, Y and Z, equal to the total number of individuals in the Susceptible, Infectious and Recovered compartment respectively.

2.1 Event-driven SIR

Whereas deterministic SIR models use differential equations of the movement between three states, event-driven SIR defines a number of possible events and a rate at which they happen. The event-driven approach induces *demographic stochasticity*, defined as fluctuations in population processes caused by the random nature of individual events [6]. We will start with a simple SIR model including demographics and imports that has seven possible events: birth, transmission, recovery, three times death (one separate event for death in every compartment) and imports. Most of these events has an intuitive rate of occurrence compared to the standard SIR (coming from a Poisson process) and a specified list of effects that is causes, moving individuals between compartments. The rate of import ι is set to be $0.0625\mu(R_0 - 1)\sqrt{N}$, as Bartlett [2] show that this is a generalized rate tot use for human populations. At the end of this section we will elaborate on import a little more, defining its importance.

The rates and movements are presented in Table 1.

Table 1: Seven possible events with their rates and resulting movements

Event	Rate	X	Y	Z
Birth	μN	X+1		
Transmission	$\beta\frac{XY}{N}$	X-1	Y+1	
Recovery	γY		Y-1	Z+1
Death from X	μX	X-1		
Death from Y	μY		Y-1	
Death from Z	μZ			Z-1
Import	ι	(X-1)	Y+1	(Z-1)

With μ the nativity and mortality rate, β the transmission rate, γ the recovery rate and ι the rate of import.

We implement this event-driven approach by Gillespie's Direct Algorithm, a stochastic simulation approach he presents in Gillespie [5]. The basic idea is determining the first arrival time for any event by randomly drawing from a Poisson distribution with a mean rate equal to the sum of rates specified in Table 1. Another random drawing multiplied by the sum of rates then determines which of the possible events this is and the current time is updated by adding the inter arrival time of the event. The complete process is described in Algorithm 1.

Algorithm 1: Gillespie's Direct Algorithm for 7-event SIR

Initialize vectors for X, Y and Z, $t \leftarrow 0$
Label the events $E_1, \ldots, E_7$
While t < T:
  N = X + Y + Z
  Determine $R_1, \ldots, R_7$ and $R_{tot} = \sum_{m=1}^{7} R_m$
  $\delta_t = \frac{-1}{R_{tot}} \ln(u1)$ and $P = u2 \cdot R_{tot}$
  with $u1$ and $u2$ two random uniform numbers.
  Event p occurs if $\sum_{m=1}^{p-1} R_m < P \leq \sum_{m=1}^{p} R_m$
  Perform p and $t \leftarrow t + \delta_t$

2.2 The five hallmarks

The event-driven approach will introduce stochasticity into SIR models and can and most likely will provoke behavior different from

the deterministic counterparts. Keeling et al. [6] list five key features that distinguish stochastic from deterministic models, some more obvious than others and with varying impact on total dynamics: variation between simulations, negative covariance between X and Y, increased transients, stochastic resonance and extinctions.

As the event-driven approach relies on simulation, drawing random numbers from an uniform distribution (see Algorithm 1), outcomes will be different for every instance. Whereas future predictions in deterministic models are certain, statistical quantities are necessary for evaluating and reporting stochastic models. The variation in the number of susceptible and infected individuals caused by the variability of the process brings us to the second feature, the generally negative covariance between the susceptible and infected group. This negative covariance is caused by the interaction between the underlying deterministic dynamics and the stochastic nature of the process, it usually has an effect on mean population levels and can lead to divergence from deterministic levels.

The third feature that results from stochastics in the model comprises the transient-like behavior that pulls disruptions back to the equilibrium. The stochastic nature can cause disruptions from the endemic equilibrium when it is reached, but the convergent nature of the underlying deterministic dynamics will force the model to return to the endemic level [6]. Transient-like returns towards the deterministic attractor become stronger when further disrupted from the equilibrium and these restorative forces cause the stochastic model to behave a lot like the underlying deterministic specifications.

SIR models usually reach the endemic state after several epidemics with decreasing amplitude, depending on the parameter settings. The increased transient behavior that is discussed above will also inflict something known as *stochastic resonance*, which can also be observed in e.g. neuroscience and physics [8] and is our fourth feature. Stochastic resonance implies the presence of oscillations around the endemic equilibrium that occur with a frequency that is close to the natural frequency of the initital peaks in infections.

The fifth and final feature of stochastic models is the appearence of stochasticity-induced extinctions. Persistence of a disease in deterministic models depends on the specified parameter settings that make up R_0, if this value is higher than 1, the disease will persist. In stochastic, closed-population models, stochastic fluctuations will eventually always cause extinction, independent of the R_0 value or the population size. Even short-term extinctions might occur during the early stages of infection. Long-term persistence of a disease can be obtained by importing the disease from an external individual into the population, these imports result in $Y \leftarrow Y + 1$ (with $X \leftarrow X - 1$ or $Z \leftarrow Z - 1$ when population size is assumed constant). As discussed in 2.1, the rate of imports has to be scaled with the population size.

The stochasticity-induced extinctions might be the most interesting feature we've discussed, as it is influenced by multiple variables and measures of persistence are all highly debatable. For example, extinction dynamics might change due to the population size, transmission rate and seasonality of the transmission rate. We will quantify some dynamics concerning extinction by presenting the number of extinctions per year and the fraction of the disease being extinct for varying population size.

2.3 Meta-population model

In the simple SIR model homogeneity of the population is assumed. For a small population this might be a valid assumption, but this assumption does not hold if we look at a country for example. We may wish to subdivide a country by town or state[9]. It is intuitive that disease transmission is most likely between individuals who are at the same location. Additionally, individuals switching locations facilitate the geographical spread of infectious diseases. The simplest spatial model (and most applicable to most human diseases) is the stochastic meta-population model. We start with dividing the population in n distinct subpopulations. Each subpopulation has independent dynamics and there is limited interaction between the subpopulations. The total individuals in subpopulation i is given by $N_i = X_i + Y_i + Z_i$, where X_i, Y_i and Z_i are respectively the number of susceptible, infected and recovered individuals. The total population is given by $\sum_{i=1}^{n} N_i$. We focus on the spatial spreading of the disease, so we ignore demography and just consider two events: infection and recovery. To determine the infection and recovery rate we set the meta-population model in the SIR framework. The rate of infection varies and depends on the the transmission rate β_i, the number of susceptibles in the subpopulation X_i, the number of infected in all subpopulations Y_j and the strength of interaction between the subpopulations. The rate of interaction in the population is captured by the matrix ρ. ρ_{ij} is the strength of infection for an individual of subpopulation j to subpopulation i. So the infection rate in the meta-population model can be mathematically formulated as: $X_i\lambda_i = X_i\beta_i \sum_{j=1}^{n} \rho_{ij}\frac{Y_j}{N_i}$. The introduced interaction term does not affect the recovery, so the recovery rate in a subpopulation is given by γY_i. We implement the stochastic meta-population model by using Gillespie's First Reaction Algorithm, a variation of Algorithm 1. The complete progress is described by Algorithm 2.

Algorithm 2: Gillespie's First Reaction Algorithm for meta-population model without demography

Initialize vectors for X, Y, Z and N, $t \leftarrow 0$
Label the events E_1, E_2
While t < T:
  Determine R_1, R_2
  $\delta_{t_i} = \frac{-1}{R_i} \ln(u1)$
  with $u1$ a random uniform number.
  Event p occurs if $\delta_{t_i} < \delta_{t_j} \forall j$.
  Perform p and $t \leftarrow t + \delta_{t_i}$.

3 RESULTS

We present results of several experiments regarding the implementation of event-driven SIR, the stochasticity that results from it and the five key features of stochastic SIR models. Thereafter, we present results from meta-population models that have a varying number of subpopulations and rate of interaction.

3.1 The effect of stochastics

First and foremost, stochasticity will obviously cause variability in the outcomes and equilibria of simulations and we observe the dynamics for varying transmission rates and population sizes. Figure

1 shows the number of infected individuals over time for 5 simulations of stochastic SIR with various values of β and N, compared to the deterministic model with equal parameter settings. The black line shows deterministic SIR dynamics and it clearly shows that the colored lines, representing the stochastic SIR simulations, exhibit a lot of variation around the deterministic dynamics. The variation relative to the initial peak in the first plot is significantly higher than in the second plot, as the second plot has a higher value for R_0 and a larger first peak. The third plot shows the dynamics for a model with equal settings to the first plot, only now simulating the model for a population that is 10 times larger. The deterministic dynamics obviously do not change, apart from a scaling factor 10. The stochastic dynamics in the third plot however show relatively little variation compared to the model with smaller population size.

As discussed in section 2.2, the variation shown above can lead to negative covariance between X and Y, which can cause equilibrium values in the stochastic model differing from the deterministic equilibrium. Table 2 presents results of 50 simulations for varying β and N, showing the average covariance between X and Y and the

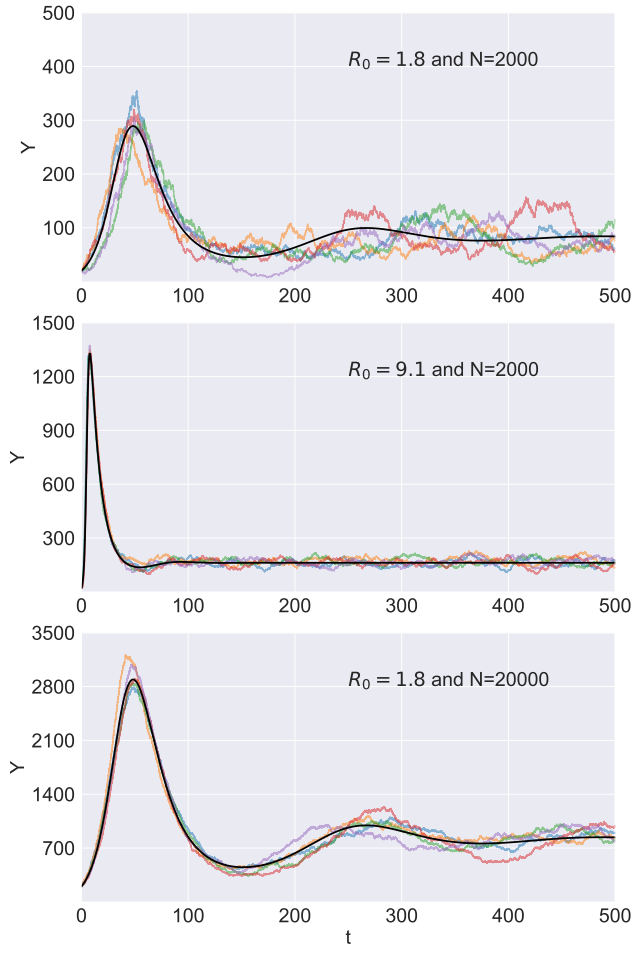

Figure 1: Variability between stochastic SIR simulations
Varying N and β, $\gamma = 0.1$, $\mu = 0.01$

Table 2: Covariances between susceptible and infected population and the equilibrium number of infected

N	β	$cov(X, Y)$	$\bar{Y}$	Y
200	0.2	-186	3.9	8.4
	0.5	-68	11.7	14.2
	1.0	66	14.3	16.2
2000	0.2	-1847	80.9	83.9
	0.5	3365	145.1	141.8
	1.0	9127	162.0	161.8

$\gamma = 0.1$, $\mu = 0.01$

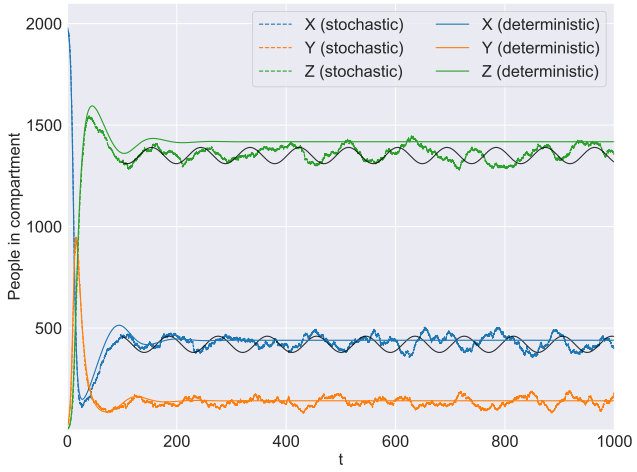

Figure 2: SIR dynamics plotted against periodic functions
$N = 2000$, $\beta = 0.5$, $\gamma = 0.1$, $\mu = 0.01$

average endemic equilibrium values for stochastic and deterministic Y. The covariance between X and Y is negative for $\beta = 0.2$ for both population sizes and for $\beta = 0.5$ in the small population. The equilibrium value of Y is clearly higher than $\bar{Y}$ in these three cases, the difference between deterministic and stochastic equilibria is smaller for the other parameter settings. As this negative covariance is caused by the interaction between deterministic model dynamics and the stochastic nature, it will logically be more frequently observed in high-variability settings. As seen in Figure 1, lower values for the rate of transmission and the population size lead to a relatively higher level of variability, thus cause a negative covariance more often.

Having solely inspected the dynamics of compartment Y, we will now focus on the behavior of all three compartments and more specifically on the transient-like returns and stochastic resonance that can occur in stochastic models. Figure 2 presents stochastic and deterministic SIR models for a population size of 2000 and an R_0 of 5. Perturbations from the deterministic equilibrium, caused by the stochastic nature of the model, can be observed for stochastic SIR. The transient-like returns to the deterministic equilibrium as discussed in section 2.2 are also clearly visible, especially looking at Figure 7 (Appendix) which does not contain the periodic functions. These functions are two sines that help show the occurrence of

stochastic resonance, the oscillations around the deterministic equilibrium that are supposed to have a frequency close to the natural one in the initial peaks. The sines in the plot have a frequency of $\frac{2\pi}{100}$ (peaks appear to be periodic every 100 days) and an amplitude of 40, 7 of the 10 initial peaks are matched by these sines and we can conclude that stochastic resonance is clearly present.

The graphs we have shown above have not presented extinctions in the time-scale that we present, so the import of infected individuals has not been visible yet. Nevertheless, extinctions are inevitable in stochastic models and the effect of several parameters, including the presence of imports, are interesting to examine.

Figure 3 provides dynamics for models that do or do not go extinct within the first 2000 days, depending on the underlying parameters regarding population size, transmission rates and imports. Plot A has the same specifications as 2, only without using imports in the model. Both plots show very similar behavior as extinction apparently does not occur within 2000 days and imports do not have a significant effect on the dynamics. Plot B decreases β to 0.15 (resulting in a low R_0 of 1.5) and immediately shows the phenomenon we are looking for. Theory states that increasing population size would prevent extinction from happening, so plot C multiplies population size by 10. Extinction does not occur and stable dynamics are again visible. Alternatively, plot D shows dynamics for low β (= 1.5) and low N (= 2000) but now uses imports to shut out complete extinction. The disease dies out after approximately 600 days, but due to the imports it prevails again after day 1200.

The dynamics from plot D in Figure 3 are again used for Figure 8 (Appendix), now plotting the number of extinctions and percentage of extinct time against population size averaged over 50 simulations. As expected, predominantly decreasing slopes can be observed for both graphs as extinctions are expected to be less frequent with increasing population size. The number of extinctions per year sometimes has an unexpected peak for rather high population size, whereas the percentage of time that the disease is extinct is steadily decreasing towards zero.

3.2 Spatial dynamics of a spreading infection

In our meta-population analysis of the a SIR framework, we ignore demography and assume that all subpopulations are of the same size ($N_i = 1000$). Furthermore, we set an equal transmission and recovery rate for every subpopulation ($\beta_i = \beta = 0.2$ and $\gamma_i = \gamma = 0.1$). We assume that the outbreak starts in subpopulation 1 with 50 infected individuals and that everyone else is susceptible, the strength of interaction within subpopulations is set to 1 ($\rho_{ii} = 1$) and varies between values lower than 1 for interactions between subpopulations ($\rho_{ij} < 1$).

Figure 4 presents the dynamics of a model with 2 subpopulations and different interaction terms ρ_{21}. From top to bottom and left to right, increasing values for ρ_{21} are shown. Whereas the peak of the disease in subpopulation 2 appears significantly later for low values of the interaction strength, the peaks quickly move closer together when it increases and are almost equal for $\rho_{21} = 0.25$. Increasing the interaction strength even further results in situations where the infection peak is reached in subpopulation 2 before it is reached in subpopulation 1. It also shows that the peak increases in amplitude

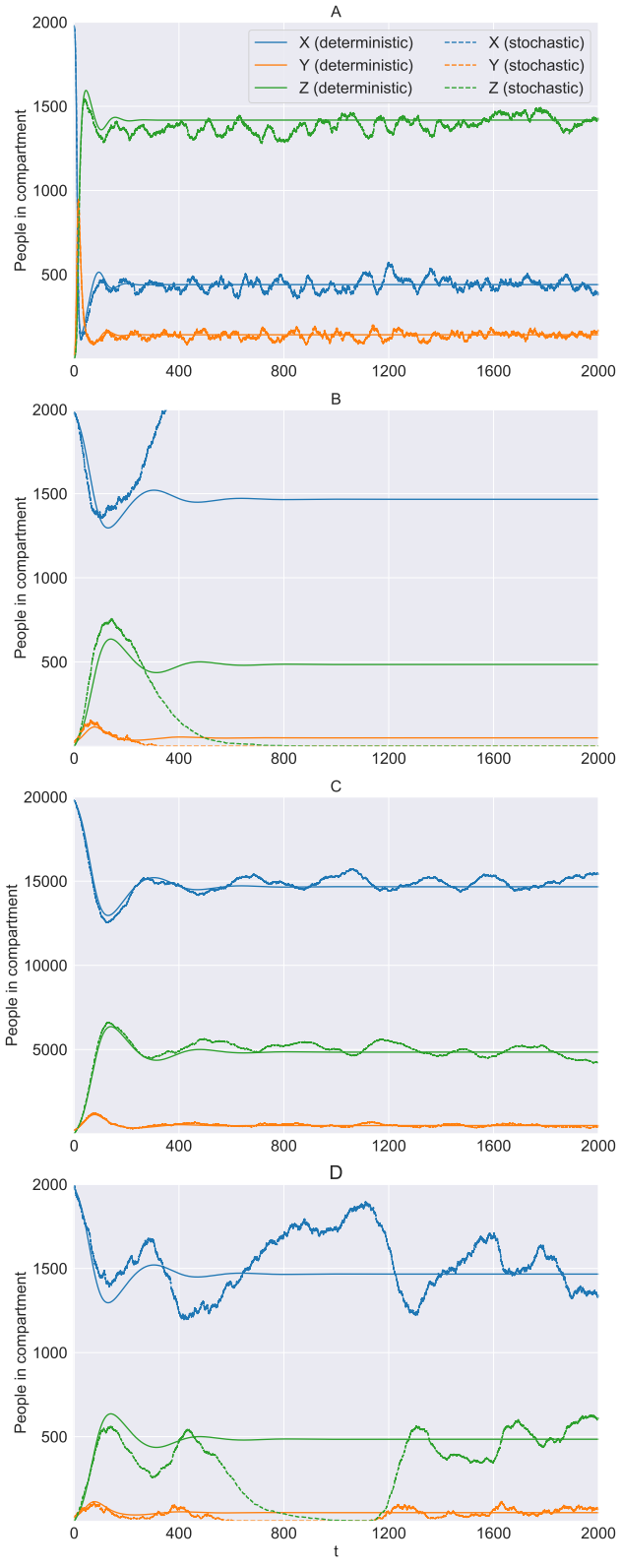

Figure 3: SIR dynamics with or without extinctions and imports

Varying N and β, γ = 0.1, μ = 0.01

as ρ_{21} increases, starting with equally leveled peaks for low interaction strength and eventually almost reaching double the amplitude. This effect can be explained by the fact that susceptible individuals in subpopulation 2 do not only interact with the infected population in their own subpopulation, but also with the infected individuals in subpopulation 1. As a consequence, a one-sided increase in infections is reached, as individuals in subpopulation 1 can not catch the disease from subpopulation 2. In Figure 5, we present the delay between the peak of the epidemic in subpopulation 1 and the peak of the epidemic in subpopulation 2, averaged over 50 simulations. We see that the time between the two peaks decreases for a higher ρ_{21} and also reaches a lot of instances below zero for $\rho_{21} > 0.5$. This clearly shows the effect we just discussed, if interactions between subpopulation occur often, the disease will spread almost immediately and may even reach the epidemic in subpopulation 2 first.

We retain all previously made assumptions and extend our meta-population model to contain six subpopulations, again we examine varying interaction patterns to understand the behavior of the model. Figure 6 presents dynamics of the meta-population model specified by three different interaction matrices. The matrices used for these models look (from top to bottom) as follows:

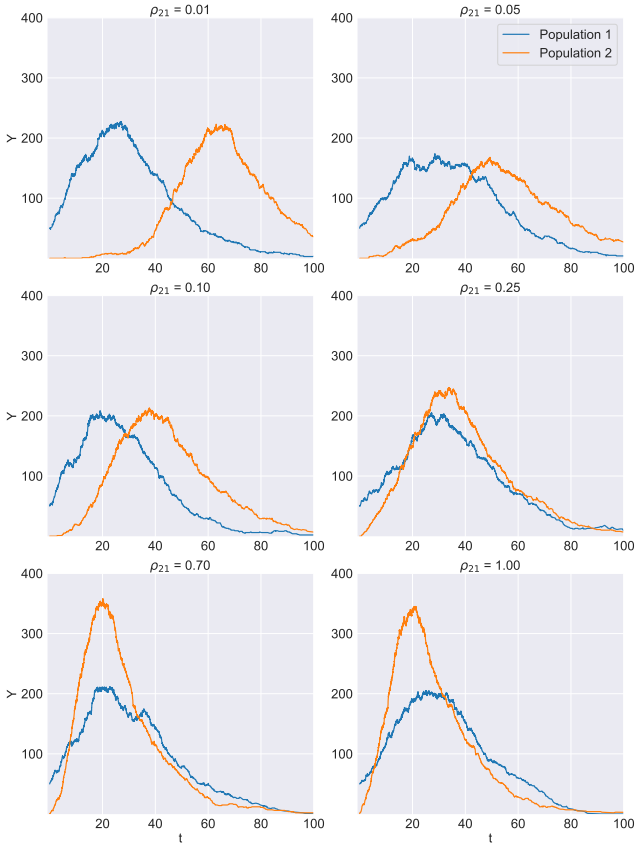

Figure 4: Meta-population without demographics and two subpopulations for different interaction terms

$\beta = 0.2$, $\gamma = 0.1$, $N_i = 1000$, $X_{0,1} = 50$

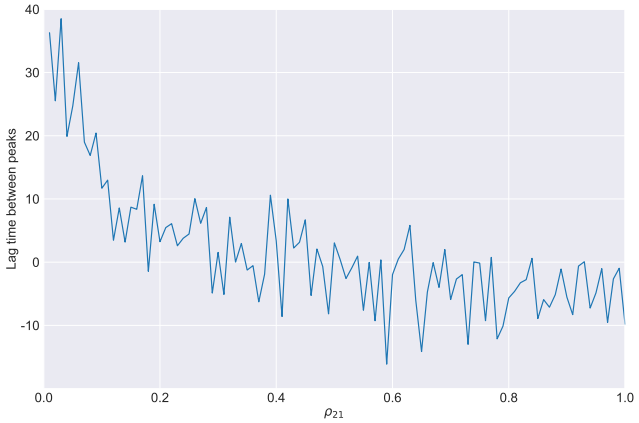

Figure 5: Delay in disease peaks of subpopulation 1 and 2

$$\rho = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0.01 & 1 & 0 & 0 & 0 & 0 \\ 0.05 & 0 & 1 & 0 & 0 & 0 \\ 0.1 & 0 & 0 & 1 & 0 & 0 \\ 0.25 & 0 & 0 & 0 & 1 & 0 \\ 0.7 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{1}$$

$$\rho = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0.1 & 1 & 0 & 0 & 0 & 0 \\ 0.1 & 0.1 & 1 & 0 & 0 & 0 \\ 0.1 & 0.1 & 0.1 & 1 & 0 & 0 \\ 0.1 & 0.1 & 0.1 & 0.1 & 1 & 0 \\ 0.1 & 0.1 & 0.1 & 0.1 & 0.1 & 1 \end{bmatrix} \tag{2}$$

$$\rho = \begin{bmatrix} 1 & 0.01 & 0.7 & 0.01 & 0.7 & 0.2 \\ 0.01 & 1 & 0 & 0 & 0 & 0 \\ 0.01 & 0 & 1 & 0 & 0 & 0 \\ 0.7 & 0 & 0 & 1 & 0 & 0 \\ 0.7 & 0 & 0 & 0 & 1 & 0 \\ 0.2 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \tag{3}$$

The first interaction matrix presents the case where individuals from subpopulation 1 interact with the other 5 subpopulations, with increasing intensity levels that are equal to the individual interaction levels from Figure 4. We observe similar behavior to the individual plots of Figure 4, with population 5 and 6 experiencing high and early peaks of infections as we would expect with the relatively high interaction level they experience.

Interaction matrix (2) contains interaction terms for all the subpopulations (in one direction) and results in dynamics that look more similar between the subpopulations. The lag between peaks decreases and the amplitude of the peaks is pulled towards each other as well. The collection of subpopulations starts behaving more like 1 population, which is an intuitive result.

The third plot presents dynamics of a variation where individuals from subpopulation 1 move to all other subpopulations and individuals from the other populations move to subpopulation 1, but not to any other subpopulations. The interaction happens at different levels as specified in interaction matrix (3). The first thing we notice

is that the peak of subpopulation 1 is in this case higher than in the other two plots. This results from the fact that (infected) individuals from all the other populations move to subpopulation 1, which was not the case in the other models. Secondly, we see that subpopulations 4 and 5 show quite similar development of the number of infections, with a slightly lower peak in subpopulation 5. Weak interactions are modelled from subpopulation 1 to subpopulations 2 and 3 and these subpopulations therefore have a low and broad infection peak. Subpopulation 3 eventually reaches a slightly higher level of infections as the interaction from subpopulation 3 to 1 is a 0.7 compared to 0.01 for subpopulation 2. A one-sided conclusion about the interaction rate from subpopulation i to subpopulation 1 can not be observed, as the comparison between subpopulations 2 and 3 and subpopulations 4 and 5 provides differing results.

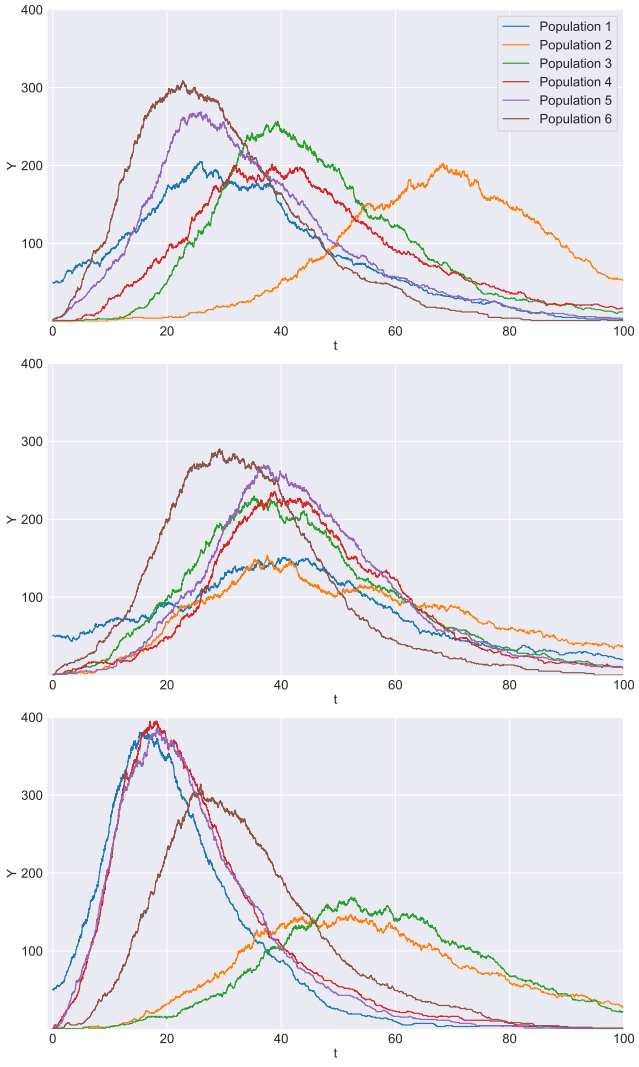

Figure 6: Meta-population without demographics and six subpopulations for different interaction matrices

$\beta = 0.2$, $\gamma = 0.1$, $N_i = 1000$, $X_{0,1} = 50$

4 CONCLUSION

We have examined two important modifications to the standard deterministic SIR model and the effects this has on model dynamics. First, stochasticity was included into the dynamics by modelling through an event-driven approach, implemented by Gillespie's Direct Algorithm. Five typical features of stochastic SIR models were discussed and we have clearly shown the appearance of variability, negative covariances, increased transients, stochastic resonance and extinctions. The behavior of models that suffer from extinctions and the implementation of imports as a solution has been investigated, as well as the direct effect of the transmission rate and population size. For these variables, we conclude that increasing either of them will help prevent extinction from happening.

Additionally, the assumption of homogeneous mixing of the population in the simple SIR model is relaxed by implementing a meta-population model, investigating the effect of the number of populations and the strength of interaction on disease dynamics. The spread of a disease among the specified subpopulations depends on how the interaction term ρ is defined. Whereas weak interaction terms present a significant epidemic delay between subpopulations, stronger interaction decreases this delay and can even increase the amplitude of the infection peak.

Interesting further research would focus on some of the assumptions that are not relaxed in this research. Transmission terms could be extended to include seasonal effects, infection-induced deaths could be specified and variations of SIR could replace it (SIS, SEIR, SI). The meta-population approach could be explored in more depth by relaxing the assumption of equally-sized subpopulations or by specifying more complex interaction matrices.

REFERENCES

[1] D. Acemoglu, V. Chernozhukov, I. Werning, and M. D. Whinston. A multi-risk sir model with optimally targeted lockdown. *NBER Working Paper*, 2020.
[2] M. S. Bartlett. Measles periodicity and community size. *Journal of the Royal Statistical Society. Series A (General)*, 120(1):48–70, 1957.
[3] Y.-C. Chen, P.-E. Lu, C.-S. Chang, and T.-H. Liu. A time-dependent sir model for covid-19 with undetectable infected persons. *IEEE Transactions on Network Science and Engineering*, 2020.
[4] I. Cooper, A. Mondal, and C. G. Antonopoulos. A sir model assumption for the spread of covid-19 in different communities. *Chaos, Solitons & Fractals*, 139:110057, 2020.
[5] D. T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *The journal of physical chemistry*, 81(25):2340–2361, 1977.
[6] M. Keeling, P. Rohani, and B. Pourbohloul. Modeling infectious diseases in humans and animals. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*, 47:864–865, 10 2008. doi: 10.1086/591197.
[7] W. O. Kermack and A. G. McKendrick. A contribution to the mathematical theory of epidemics. *Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character*, 115(772):700–721, 1927.
[8] F. Moss, L. M. Ward, and W. G. Sannita. Stochastic resonance and sensory information processing: a tutorial and review of application. *Clinical neurophysiology*, 115(2):267–281, 2004.
[9] C. Viboud, O. Bjørnstad, D. Smith, L. Simonsen, M. Miller, and B. Grenfell. Synchrony, waves, and spatial hierarchies in the spread of influenza. *Science*, 312 (5772):447–551, 2006. doi: 10.1126/science.1125237.

APPENDIX

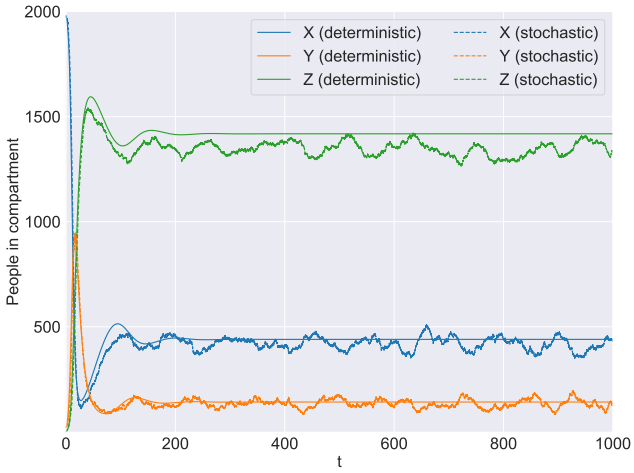

Figure 7: Stochastic and deterministic SIR dynamics
$N = 2000$, $\beta = 0.5$, $\gamma = 0.1$, $\mu = 0.01$

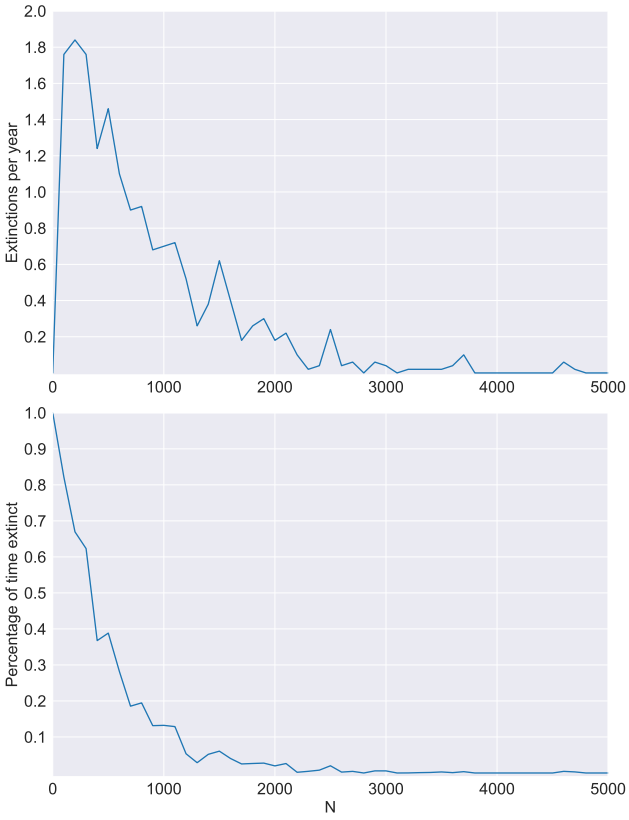

Figure 8: Extinctions for varying N
$\beta = 0.15$, $\gamma = 0.1$, $\mu = 0.01$