Hierarchical Spatial Structure in Compartmental Epidemiological models

James Bender

Sections

Copyright notice 3

Abstract 4

Declaration 5
Standard thesis . . . 5
Thesis including published works declaration . . . 6

Acknowledgements 7

1 Introduction 0

I Background 1

2 Compartmental Models 2
2.1 SIR Dynamics . . . 3

3 Compartmental models as stochastic processes 4
3.1 Stochastic processes . . . 4
3.2 Stochastic simulation . . . 6

4 Metapops 9
4.1 Multi Patch 'Metapopulation' Model . . . 9
4.2 Example: Origin-destination Spatial Metapopulation model . . . 11

II Current Study 19

5 Hierarchical metapopulation models 20
5.1 Hierarchical Structure of the SA classification . . . 20
5.2 Comparisons between models of different scales . . . 22
5.3 Mixing strategies in meta-population models . . . 22

6 Mixing Matrix Interventions 32
6.1 Motivation . . . 32
6.2 Formulation . . . 32

6.3 Implimentation 33
6.4 Simulation 33

Bibliography **35**

Appendices **36**

A Additional stuff **36**

Copyright notice

Abstract

Epidemiological modelling is essential to understand and predict the spread of infectious diseases and plan effective interventions. Disease spread is an inherently spatial process, particularly in diseases that spread through contact with body fluids, and spatially structured epidemiological models provide an opportunity to incorporate a wealth of geographical information into the modelling process. Here, we present a compartmental meta-population model encoding the hierarchical spatial structure of the Australian Statistical Geography Standard (ASGS). Relationships between spatial groupings are represented in a mixing matrix, which influences the spread of infection between patches (i.e. Patches in the same spatial grouping are more likely to spread disease to each other compared to patches outside this grouping). As an example, we model the Greater Melbourne Statistical Area (SA) with patches representing SA2, SA4, and SA4 level groupings, and stochastically simulate the course of an epidemic as a Continuous Time Markov Chain (CTMC). We compare the results of simulations with varying influence of the spatial relationships between patches (through the construction of the mixing matrix), different levels of spatial resolution (which level of the spatial hierarchy is represented by compartmental patches), and different infectious disease parameters (e.g. R_0), as well as with the results of a spatial compartmental model incorporating an alternative source of geographic information (i.e. empirically derived mobility data). We conclude by speculating on the potential use of such geographically informed epidemiological models in the management of infectious diseases.

Declaration

Use only one of the following declarations.

Standard thesis

This thesis is an original work of my research and contains no material which has been accepted for the award of any other degree or diploma at any university or equivalent institution and that, to the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

Student name:

Student signature:

Date:

Publications during enrolment

Remove this section if you do not have publications.

The material in sec-intro has been submitted to the journal *Journal of Impossible Results* for possible publication.

The contribution in sec-litreview of this thesis was presented in the International Symposium on Nonsense held in Dublin, Ireland, in July 2022.

Reproducibility statement

This thesis is written using Quarto with renv (Ushey, 2022) to create a reproducible environment. All materials (including the data sets and source files) required to reproduce this document can be found at the Github repository `github.com/SusanSu/thesis`.

Thesis including published works declaration

I hereby declare that this thesis contains no material which has been accepted for the award of any other degree or diploma at any university or equivalent institution and that, to the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

This thesis includes ?? original papers published in peer reviewed journals and ?? submitted publications. The ideas, development and writing up of all the papers in the thesis were the principal responsibility of myself, the student, working within the Department of Econometrics & Business Statistics under the supervision of ??

(The inclusion of co-authors reflects the fact that the work came from active collaboration between researchers and acknowledges input into team-based research.)

In the case of (??insert chapter numbers) my contribution to the work involved the following:

I have / have not renumbered sections of submitted or published papers in order to generate a consistent presentation within the thesis.

Student name:

Student signature:

Date:

I hereby certify that the above declaration correctly reflects the nature and extent of the student's and co-authors' contributions to this work. In instances where I am not the responsible author I have consulted with the responsible author to agree on the respective contributions of the authors.

Main Supervisor name:

Main Supervisor signature:

Date:

Acknowledgements

I would like to thank my pet goldfish for …

In accordance with Chapter 7.1.4 of the research degrees handbook, if you have engaged the services of a professional editor, you must provide their name and a brief description of the service rendered. If the professional editor's current or former area of academic specialisation is similar your own, this too should be stated as it may suggest to examiners that the editor's advice to the student has extended beyond guidance on English expression to affect the substance and structure of the thesis.

Free text section for you to record your acknowledgment and gratitude for the more general academic input and support such as financial support from grants and scholarships and the non-academic support you have received during the course of your enrolment. If you are a recipient of the "Australian Government Research Training Program Scholarship", you are required to include the following statement:

> "This research was supported by an Australian Government Research Training Program (RTP) Scholarship."

You may also wish to acknowledge significant and substantial contribution made by others to the research, work and writing represented and/or reported in the thesis. These could include significant contributions to: the conception and design of the project; non-routine technical work; analysis and interpretation of research data; drafting significant parts of the work or critically revising it so as to contribute to the interpretation.

1 Introduction

Part I

Background

2 Compartmental Models

"Compartmental models are the most frequently used type of epidemic model. In this class of models, individuals can be in a finite number of discrete states. Some of these states are simply labels that specify the various traits of individuals. Of these, some will be changing with time, such as age class, and others will be fixed, such as sex or species. Other states indicate the progress of an infection: for example, an individual can upon becoming infected, typically first enter a state of latency, then progress to a state of infectiousness, and then lose infected status to progress to a recovered/immune state. With each state one can associate the subpopulation of individuals who are in that particular state at the given time (e.g. a female in a latent state of infection). Often the same symbol is used as a label for a state and to denote the corresponding subpopulation size, either as a fraction or as a number (e.g. I or Y for individuals in an infectious state)" - (**diekmann2010?**)

2.0.1 The SIR compartmental model

The simplest comparmental model of infectious disease spread is the SIR Compartmental Model, with three compartments: S - for individuals *susceptible* to the disease; I - for *infected* individuals; and R - for previously infected individuals who have *recovered* (or been otherwise *removed* from that compartment). With the simplifying assumption of a constant population size, N, i.e.

$$N = S + I + R \tag{2.1}$$

Individuals move between compartments in a fixed set of ways, they may either become infected (moving from S -> I), or recover from infection (I -> R).

At each time unit an infected individual can come into contact with, on average, $k\frac{S}{N}$ Susceptible individuals. π is the probability of infecting somebody on coming in contact, so $\beta = k\pi$ is the average rate at which an infected indivudual will infect a susceptible. Infected individuals recover at the constant rate, γ, with $1/\gamma$ the mean recovery time.

Parameter	Interpretation
β	Transmission rate
γ	Recovery rate

Thus, an SIR model can be represented by the schematic

2.1 SIR Dynamics

2.1.1 ODE representation

Given the rate parameters defined above, the change in compartment composition over time can be described by the system of differential equations

$$\begin{aligned} \frac{dS}{dt} &= -\beta I \frac{S}{N} \\ \frac{dI}{dt} &= \beta I \frac{S}{N} - \gamma I \\ \frac{dR}{dt} &= \gamma I \end{aligned} \tag{2.2}$$

Importantly, if the total population size is known, given Equation 2.1, R = N - S + I and the entire system can be described by two of the equations in Equation 2.2.

EXAMPLE

Limitations of Deterministic ODE

While Equation 2.2 provides a neat solution for the expected behaviour of a epidemic, it fails to caputure the variability inherent in a complex process like disease spread.

3 Compartmental models as stochastic processes

3.1 Stochastic processes

A Stochastic Process is a collection of random variables, $X(t)$. In the case of a compartmental SIR model **?@sec-ref**, the state of the model at time t can be represented by the number of individuals in the Susceptible and Infected compartments, i.e.

$$X(t) = S(t) + I(t)$$

Where $S(t), I(t) \in \{0, 1, 2, \dots, N\}$, $S(t) + I(t) \leq N$, and $t \in [0, \infty)$

We can represent the probability of a transitioning one particular state $x = (s, i)$[1] to another $x = (s + k, i + j)$ (i.e. due to infection or recovery) after a time $\Delta t > 0$, as

$$P(X(t + \Delta t) = (s + k, i + j) \mid (S(t), I(t)) = (s, i))$$

We will make two assumptions about the probability of states, $P(X(t))$:

1. Time Homogeneity

   $$P\left(X(t + \Delta t) = j \mid X(t) = i\right) = P\left(X(t) = j \mid X(0) = i\right)$$

   Which states that the transition probabilities depend on the time between events, Δt, but not on the specific time t.

2. Memorylessness (Markov Property)

   $$P(X(t + \Delta t) = j \mid X(t) = i, X(u) = k, 0 \leq u < t) = P(X(t + \Delta t) = j \mid X(t) = i)$$

   Which states that the probability of transitioning from a state i at time t to a state j at time $(t + \Delta t)$ depends on the state at t and not on any previous state.

[1] lower case letters (e.g. s, i) represent specific values of the random variables ($S(t)$ and $I(t)$)

Representing our compartmental model as such a 'Time Homogeneous Markov Chain' means that we can succinctly describe the probabilities of possible state transitions ('events') because they only depend on the intervening time (Δt) and the current state $x_t = (s_t, i_t)$.

Table 3.1: Transition events in a SIR time homogeneous markov chain

Event	Change (ΔS , ΔI)	Probability
Infection	$(-1, +1)$	$\beta i \frac{s}{N} \Delta t + o(\Delta t)$
Recovery	$(0, -1)$	$\gamma i \Delta t + o(\Delta t)$

Time homogeneity and the markov property allow us to define the 'transition function' $p_{ij}(t)$, which gives the probabilty of transitioning from state i to state j after time t

$$p_{ij} = P(X(t) = P(X(t) = j \mid X(0) = i)$$

and represent the possible state transition probabilities for our SIR model as

$$p_{(s,i),(s+k,i+j)}(\Delta t) = \begin{cases} \beta i \frac{s}{N} \Delta t + o(\Delta t), & (k,j) = (-1,+1) \\ \gamma i \Delta t + o(\Delta t), & (k,j) = (0,-1) \\ 1 - (\beta i \frac{s}{N} + \gamma i)\, \Delta t & \\ \quad + o(\Delta t), & (k,j) = (0,0) \\ 0(\Delta t), & \text{otherwise.} \end{cases} \tag{3.1}$$

Moreover, if we consider the state space of $X(t)$, (e.g. containing all possible comninations of numbers of susceptible and infected individuals (s, i),

$$\mathbb{S} = \{s, i : 0 \leq s, i; s + i \leq N\}.$$

For a particular ordering of this space[2] (e.g. $(s, i) \in \{(N, 0), (N - 1, 1), \ldots, (0, 0)\}$), $P(t)$ can be represented as a matrix specifying the probability of every state transition at time t. Our bivariate SIR state space contains $(N + 1)(N + 2)/2$ elements, and therefore P(t) corresponds to a $(N + 1)(N + 2)/2 \times (N + 1)(N + 2)/2$ matrix.

We can now consider the derivative of the transition function, $p'_{i,j}(t)$, which has two forms known as the forward (Equation 3.2) and backward (Equation 3.3) Kolmogorov equations

$$\frac{dp_{i,j}(t)}{dt} = \sum_{k \neq i} p_{i,k}(t) q_{k,j} - q_{i,i} p_{i,j}(t) \tag{3.2}$$

[2]Note that while the ordering of state space is arbitrary, it is not without consequence Black & Ross (2015)

$$\frac{dp_{i,j}(t)}{dt} = \sum_{k \neq i} q_{i,k} p_{k,j}(t) - q_{i,i} p_{i,j}(t) \tag{3.3}$$

where $q_{k,j}$,$q_{i,i}$, and $q_{i,k}$ are the transition *rates* defined in Equation 3.1, i.e. for our SIR model, the forward Kolmogorov equations are

$$\frac{dp_{a,(s,i)}(t)}{dt} = p_{a,(s+1,i-1)}(t)\frac{\beta}{N}(s+1)(i-1) + p_{a,(s,i+1)}(t)\gamma(i+1) - p_{a,(s,i)}(t)\left[\frac{\beta}{N}si + \gamma i\right].$$

and the backward Kolmogorov equations are

$$\frac{dp_{(s,i),b}(t)}{dt} = \frac{\beta si}{N} p_{(s-1,i+1),b}(t) + \gamma i p_{(s,i-1),b}(t) - \left[\frac{\beta}{N}si + \gamma i\right] p_{(s,i),b}(t).$$

We can simplify (Equation 3.2) and (Equation 3.3), by considering the matrix Q specifying the rates of all transitions q_{ij} for $i, j \in \mathbb{S}$.Similarly to the transition probability matrix $P(t)$, Q has dimension $(N+1)(N+2)/2 \times (N+1)(N+2)/2$.

The system of equations Equation 3.2 and Equation 3.3, can now be written as $dP(t)/dt = P(t)Q$ and $dP(t)/dt = QP(t)$, respectively. Considering$p'(t) = qp(t)$, where p is a differentiable function and q is a constant, when $p(0) = 1$, there is a unique solution $p(t) = e^{tq}$, we can similarly solve the Kolmogorov Equation 3.2 and Equation 3.3 as as $P(t) = e^{Qt}$ and $P(t) = e^{tQ}$ (when $P(0) = I$), respectively.

Thus, the theoretically and empirically derived event rates presented in Equation 3.1, along with the assumptions of time homogeneity and memorylessness, allow us to compute the probability of any state transition at time t. This technique can be used, for example, to derive important statistics about an epidemic, such as its mean final size and the effects of vaccination programs Teo (2017), Teo et al. (2021). However, "computing the matrix exponential is often numerically challenging. Finding accurate and eficient algorithms is still a topic of current research" Dobrow (n.d.)

3.2 Stochastic simulation

Instead of analytically deriving the marginal probability of a particular series of events by the method outlined above, we can simulate instantiations of the continuous time markov chain ("sample paths") numerically. There are several stochastic simulation algorithms Simoni et al. (2019), but in this thesis, we will use only the original 'Direct-method' algorithm put forward by Gillespie (1976).

Originally a method for simulating chemical reaction networks, the Gillespie direct algorithm is applicable to many domains of stochastic simulation with discrete state spaces, continuous time, and known rate equations.

In general, the core of the Gillespie algorithm takes in the possible state changes in a system, and a 'propensity function' for each state change. Random number generation is used to select a particular state change, u and a time interval, τ, which can be used to update the state of the system $X(t) = i \rightarrow X(t + \tau) = i + u$ (Algorithm 1).

Algorithm 1 Gillespie Direct Method

Input: v, a

Output: μ, τ

set $a_{net} = 0$
for all j **do**
    compute a_j
    update $a_{net} = a_{net} + a_j$
end for
generate two random numbers r_1, r_2 in $\mathcal{U}(0, 1)$
select μ such that $\sum_{j=1}^{\mu} a_j \leq r_1 a_0$
compute $\tau \leftarrow \frac{1}{a_0} \ln(1/r_2)$
update $X \leftarrow X + v_\mu$
set $t \leftarrow t + \tau$

In particular, two random numbers r_1 and r_2 are used to compute mu and τ, respectively

- r_1 is used to calculate mu by cumulatively adding the values of the propensity function to each other. mu gets the largest value which is smaller than r_1
- r_2 is used to calculate τ by taking advantage of the fact that in a Continuous time markov proces $\Delta t \sim \lambda e^{-\lambda t}$, where lambda is the sum of the rates for all possible events.

In the case of our SIR model, the state change vector is given by Table 3.1 column two, and the propensity function vector is given in column 3.We can include the core stochastic event step in a loop which updates the state and time of our SIR model simulate the progression of an epidemic @alg-SIR_CTMC.

As an example, **?@fig-GMelb_SIR_Sim** 100 sample trajectories of an SIR CTMC comprising N = 4976157, the population of the Greater Melbourne region. In this example, $I_0 = 10$, $\beta = 1.4$, and $\gamma = 1.0$, though clearly variation in these parameter values can be explored.

The Greater Melbourne Region presents large geographical region which highlights the important assumption of homogeneous mixing, which we will explore relaxing in the next chapter.

Algorithm 2 Stochastic simulation of SIR Compartmental Model CTMC

Input:N, I_0, β, γ **Output:**

Initialise time $t \leftarrow 0.0$ and $S \leftarrow N - I_0$, $I \leftarrow I_0$
state change vector $v = [(-1, 1), (0, -1)]$
propensity vector $a = [(\beta I \frac{S}{N}), (\gamma I)]$
while $I > 0$ **do**
    for all j **do**
        compute a_j
        update $a_{net} = a_{net} + a_j$
    end for
    generate two random numbers $r_1, r_2 \sim \mathcal{U}(0, 1)$
    select μ such that $\sum_{j=1}^{\mu} a_j \leq r_1 a_0$
    compute $\tau \leftarrow \frac{1}{a_0} \ln(1/r_2)$
    update $X \leftarrow X + v_\mu$
    set $t \leftarrow t + \tau$
end while

4 Metapops

4.0.1 Relaxing the homogeneous mixing assumption

A fundamental assumption of the simple compartmental model presented in **?@sec-Compartmental_models** is that any susceptible individual is equally likely to become infected by any of the infectious individuals (at a rate proportional to the size of each compartment in the population). However, real-world populations are not uniform in their interactions (particularly on a scale like that of **?@sec-Melb_SIR**), and the assumption of homogeneous mixing can be relaxed to produce more realistic models that demonstrate heterogeneous mixing patterns.

4.1 Multi Patch 'Metapopulation' Model

In the simple SIR model of Chapter 2, a single population is divided into a number of compartments, each with their own properties.

We can extend this model to instead consider a set of n populations, each with their own number of residents N_i for that make up the larger population N

$$\sum_{i=1}^{n} N_i = N$$

Each of these sub-populations, which will hereafter be referred to as 'patches', contains compartments which behave analogously to those from Chapter 2, such that

$$S_i + I_i + R_i = N_i$$

While individuals can only transition between compartments of their respective patch (i.e. N_i is constant), patches are coupled such that susceptibles may still contract the disease by coming in to contact with infectious individuals from another patch. The coupling between two patches $i, j \in \{1, \dots, n\}$ is termed the mixing coefficient, denoted m_{ij}, and is defined as the probability that an individual from patch i will next come into contact by an individual from patch j. As such, $0 \leq m_{ij} \leq 1$ and $\sum_j m_{ij} = 1$.

We can now define the `force of infection` in patch i that is exerted by infectious individuals from patch j, as

$$\lambda_{ij} = \beta \cdot I_j \cdot m_{i,j}$$

The total force of infection experience by patch i

$$\Lambda_i = \sum_{j=1}^{n} \lambda_{ij}$$

And thus describe an ODE model of patch i as

$$\begin{aligned} S_i' &= -\Lambda_i \cdot \frac{S_i}{N_i} \\ I_i' &= \Lambda_i \cdot \frac{S_i}{N_i} - \gamma I_i \\ R_i' &= \gamma I_i \end{aligned}$$

Similarly to **?@sec-Stochastic** SIR, we can construct a Continuous Time Markov Chain Metapopulation SIR model with a state space

$$\mathbb{S} = \left\{(s_1, \dots, s_i, \dots, s_n, i_1, \dots, i_i, \dots, i_n) : 0 \geq s_i, i_i; s_i + i_i < N_j\right\} \tag{4.1}$$

and transition rates

$$\begin{aligned} q_{x,x+inf_i} &= \Lambda_i \frac{S_i}{N_i} \\ q_{x,x+rec_i} &= \gamma I_i \end{aligned} \tag{4.2}$$

for $i, j \in \{1, \dots, n\}$, where

- $x = (s_1, \dots, s_i, \dots, s_n, i_1, \dots, i_i, \dots, i_n)$
- $\mathbf{inf}_i = \mathbf{e}_{2i} - \mathbf{e}_i$, where $\mathbf{e}_i$ is a vector of 0s (length $2n$) with a 1 in the i^{th} position,
- $\mathbf{rec}_i = -\mathbf{e}_{2i}$

Following from **?@sec-Stochastic_SIR_sim**, we can simulate sample paths of this CTMC meta-population model using a stochastic simulation algorithm (Algorithm 3). The process outlined in Algorithm 3 is similar to that of Algorithm **??** with the main distinction[ˆmetapop2s-1] being that there are now $2n$ possible events (i.e. an infection and recovery event for each patch). Each element of the state change vector $\mathbf{v}$ now encodes both the location and type of an event

$$\mathbf{v}_i = \begin{cases} \mathbf{e}_{2i} - \mathbf{e}_i & \text{for} i \leq n \\ -\mathbf{e}_{2i}, & \text{for} i > n \end{cases}$$

The propensity vector $\mathbf{a}$ is similarly defined using the rates form Q defined in Equation 4.2 such that

$$a_i = \begin{cases} \frac{S_i}{N_i}\Lambda_i & \text{for} i \leq n \\ \gamma I_{n-i} & \text{for} i > n \end{cases}$$

Note that the event location and type are determined by the same random number r_1.

Also note that a the beginning of a simulation, all patches are composed of entirely susceptible individuals. A number, I_0, of individuals in a randomly selected patch, α, become infected before initial transition rates are computed.

$$S_i(0) = \begin{cases} N_i & \text{if } i \neq \alpha \\ N_i - E_0 & \text{if } i = \alpha \end{cases}$$
$$I_i(0) = \begin{cases} 0 & \text{if } i \neq \alpha \\ I_0 & \text{if } i = \alpha \end{cases}$$

for $i \in \{1, \ldots, n\}$ where $\alpha \sim \mathcal{U}\{1, r\}$.

4.2 Example: Origin-destination Spatial Metapopulation model

Following Moss et al. (2019), we will use as an example a meta-population model of the Greater Melbourne region subdivided into 40 patches according to the (Australian Statistical Geography Standard (ASGS), 2023) SA3 classification system. The population of each patch along with its numeric SA identifier is given in **?@sec-appendix1** and shown in Figure 4.1

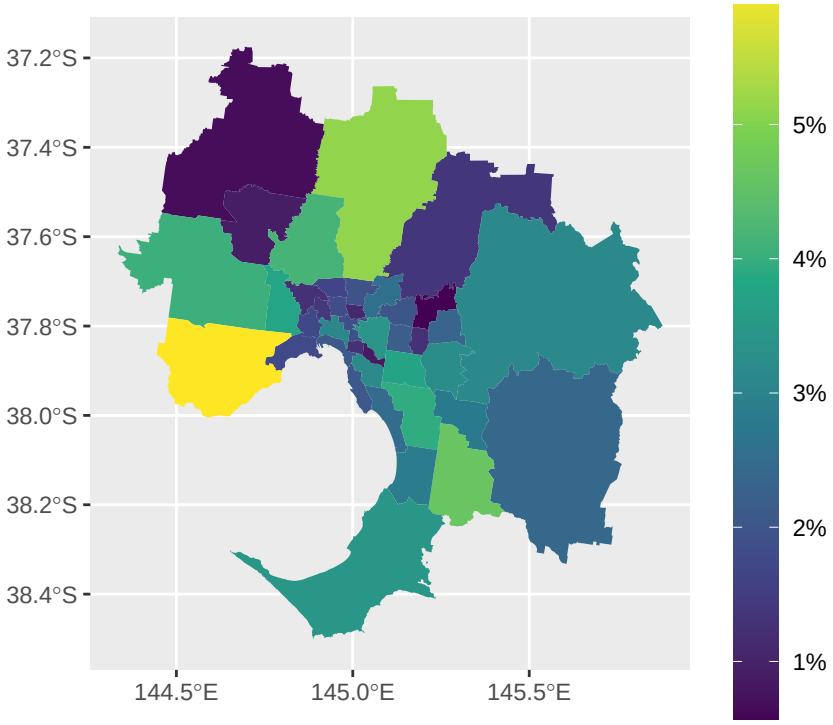


Figure 4.1: **?(caption)**

Algorithm 3 Stochastic simulation of SIR Metapopulation CTMC

Input: $\mathbf{N}$, I_0, β, γ, M

$t \leftarrow 0.0$ ▷ Initialise time
for all i **do** ▷ Initialise susceptibles
    $S_i \leftarrow N_i$
end for
Select $i \sim \mathcal{U}[1, n]$, $I_i \leftarrow I_0$ ▷ Seed infection
while $I \geq 0$ **do**
    for all $i \in \{1, \dots, n\}$ **do**
        $a_i \leftarrow \frac{S_i}{N_i} \cdot \sum_{j=1}^{n} \beta \cdot I_j \cdot M'_{j,i}$ ▷ Update Infection Rates
        $a_{n+i} \leftarrow \gamma I_i$ ▷ Update Recovery Rates
        update $a_{net} = a_{net} + a_i + a_{n+i}$
    end for
    generate two random numbers $r_1, r_2 \sim \mathcal{U}(0, 1)$
    select μ such that $\sum_{j=1}^{\mu} a_j \leq r_1 a_{net}$
    compute $\tau \leftarrow \frac{1}{a_{net}} \ln(\frac{1}{r_2})$
    update $X \leftarrow X + v_\mu$
    set $t \leftarrow t + \tau$
end while

4.2.1 Origin-destination mixing matrix

The mixing matrix was developed after Moss et al. (2019) using an empirically informed origin-destination (OD) matrix derived from 'Place of work' data taken from the Australian Census (**ABS_census2016?**). Rows ('origin') are the 'usual residence', and the columns (destination) are the 'place of work'. The OD matrix

$$F = \begin{pmatrix} f_{1,1} & f_{1,2} & \cdots & f_{1,n} \\ f_{2,1} & f_{2,2} & \cdots & f_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ f_{n,1} & f_{n,2} & \cdots & f_{n,n} \end{pmatrix}$$

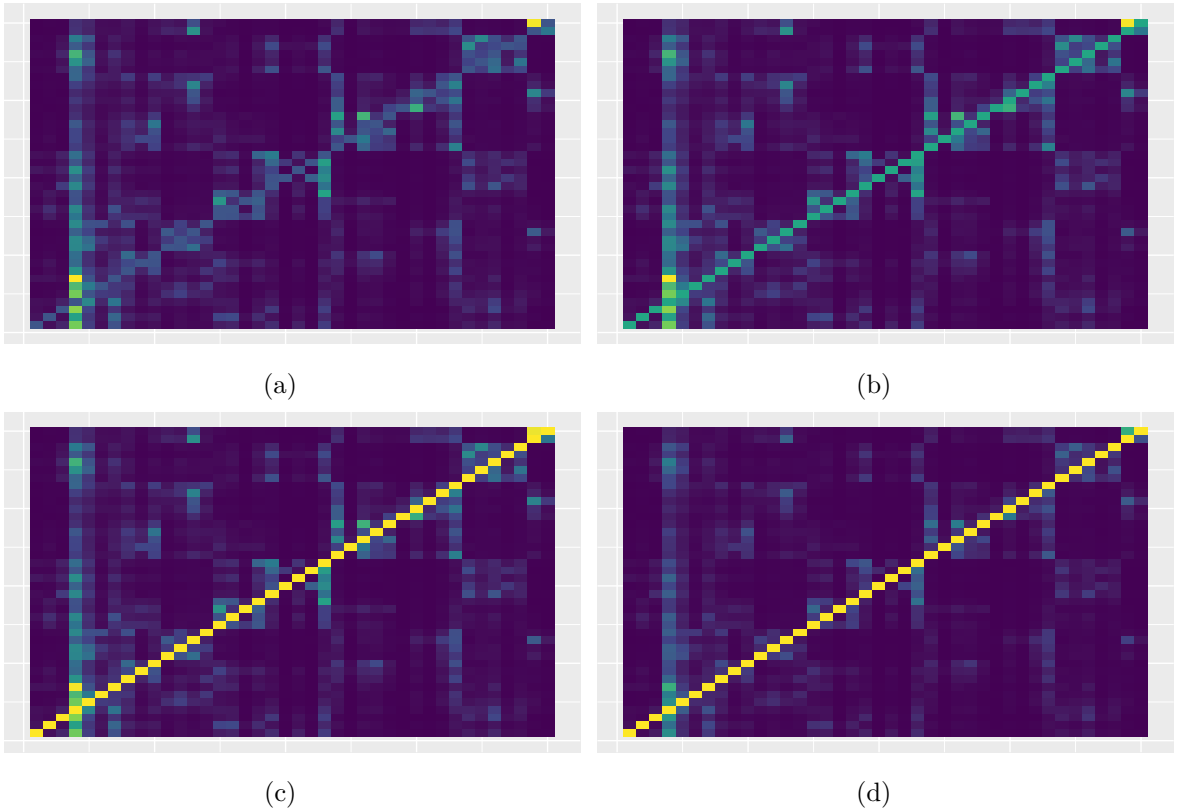
f_{ij} is the proportion of people why usually reside in patch i listing patch j as their place of work. This empirical method is expected to describe contact patterns *between regions*, but contact patterns within *within* the region of residence are expected to result more from contact outside of a work context.Therefore, diagonal elements are set to zero.

$$f_{i,i} = 0$$
$$\sum_{j=1}^{r} f_{i,j} = 1 \quad \forall i \in [1..r]$$

The final mixing matrix is defined using the 'local' mixing is given by a parameter δ^H. Diagonal elements are set to δ^H with the remaining proportion, $\delta^* = 1 - \delta_i^H$ distributed among the non-local patches equally.

$$M = \begin{pmatrix} \delta_1^H & \delta_1^* f_{1,2} & \cdots & \delta_1^* f_{1,r} \\ \delta_2^* f_{2,1} & \delta_2^H & \cdots & \delta_2^* f_{2,r} \\ \vdots & \vdots & \ddots & \vdots \\ \delta_r^* f_{r,1} & \delta_r^* f_{r,2} & \cdots & \delta_r^H \end{pmatrix}$$

OD matrices for several values of δ^H are presented as heatmaps in **?@fig-OD_matrices**



(a) (b)

(c) (d)

Figure 4.2: showing OD mixing matrices with values of δ^H = (a) 0.1, (b) 0.2, (c) 0.3, (d) 0.4

4.2.2 Simulation results

We can observe the metapopulation infection curves for our OD model for a range of R_0 and δ^H values in Figure 4.3, and the key infecton statistics in Figure 6.1

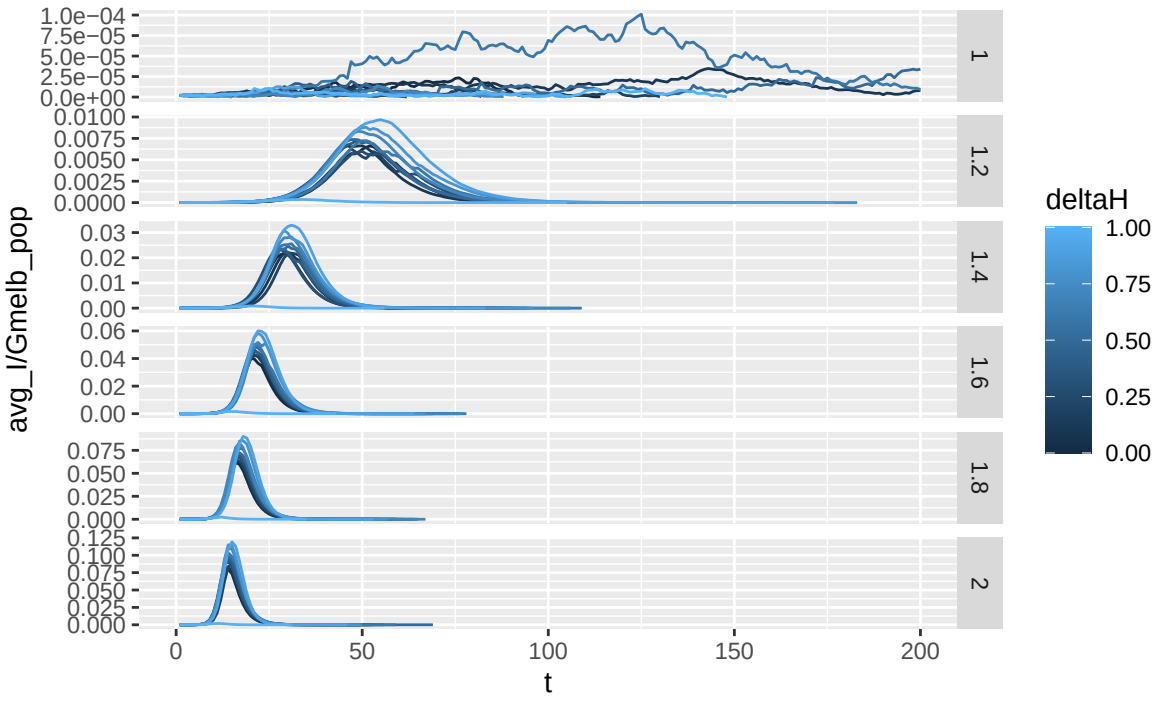


Figure 4.3: Average proportion of the population over time for simulations of a OD mixing metapopulation model of the melbourne metropolitan area with a range of Basic reproduction numbers (R_0), and local mixing coefficients($delta^H$). Note the different y scale in each facet.

However, now we can decompose these metapopulation scale outcomes into those of the underlying subpopulation. For example, **?@fig-OD_patchinf_curve_eg**

And we can summarise over multiple simulations to get a better understanding of the differential influence of the local mixing parameter δ^H in each SA3 population. For example, Figure 4.6 shows the consistent trend toward higher peak infections with greater proportons of local mixing (higher δ^H). However, certain patches demonstrate the opposite effect - the Melbourne City patch for example has the highest peak proportion infection (of any patch) when local mixing is low. This can be explained by the relatively large number of residents from all patches working in the city (this is evident in the OD mixing matrices of Figure 4.2, where the city POW is identifiable as a bright column). High inter-patch mixing means that the city center has a constant supply of infectious contacts.

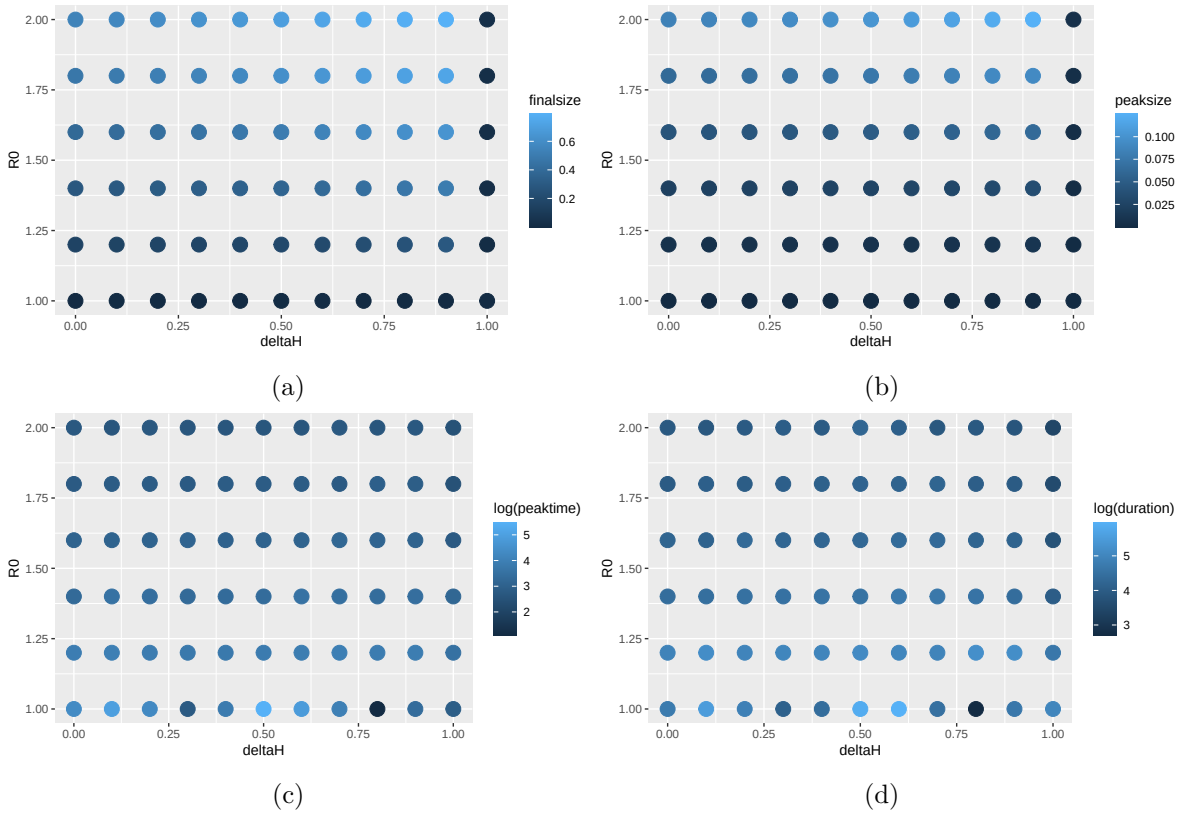


Figure 4.4: Final (a) and peak (b) infection numbers (as a proportion of the entire population), peak time (c) and total duration (d) of a simulated SIR metapopulation model with OD mixing matrix at different values of δ^H and R_0

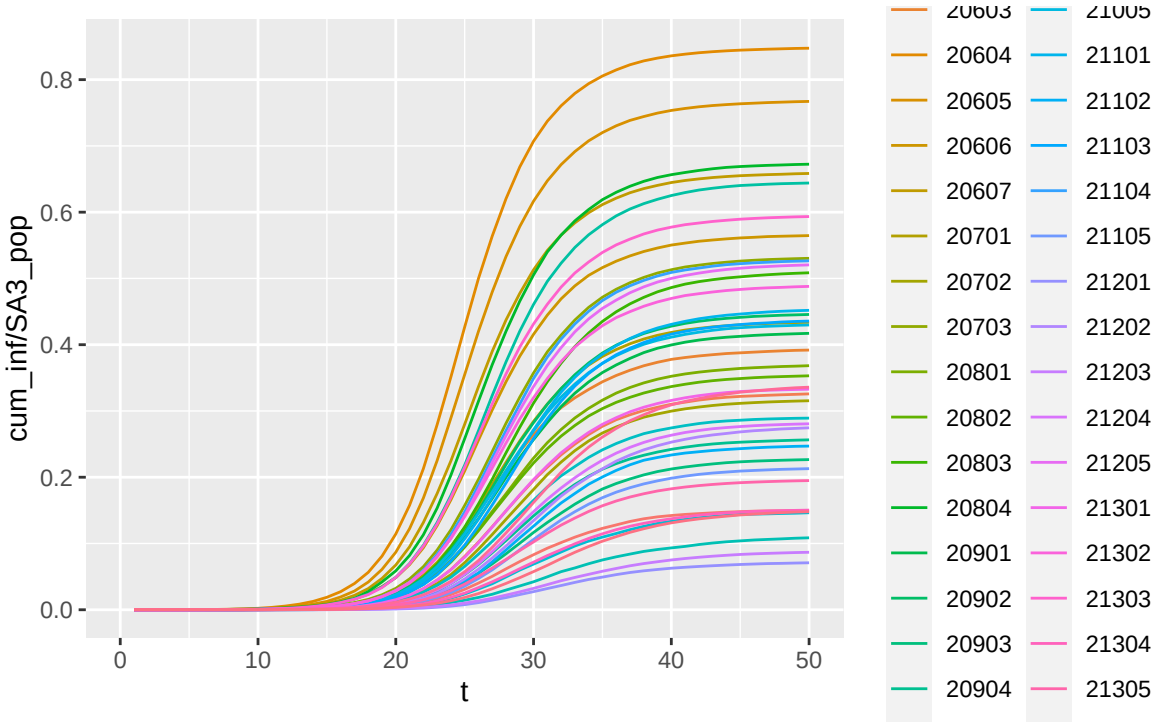


Figure 4.5: Cumulative infections (as a proportion of patch population) for each SA3 scale patch in the first 50 days of single simulation of an OD mixing metapopulation SIR model of the greater Melbourne region. $R_0 = 1.4$, $\delta^H = 0.5$.

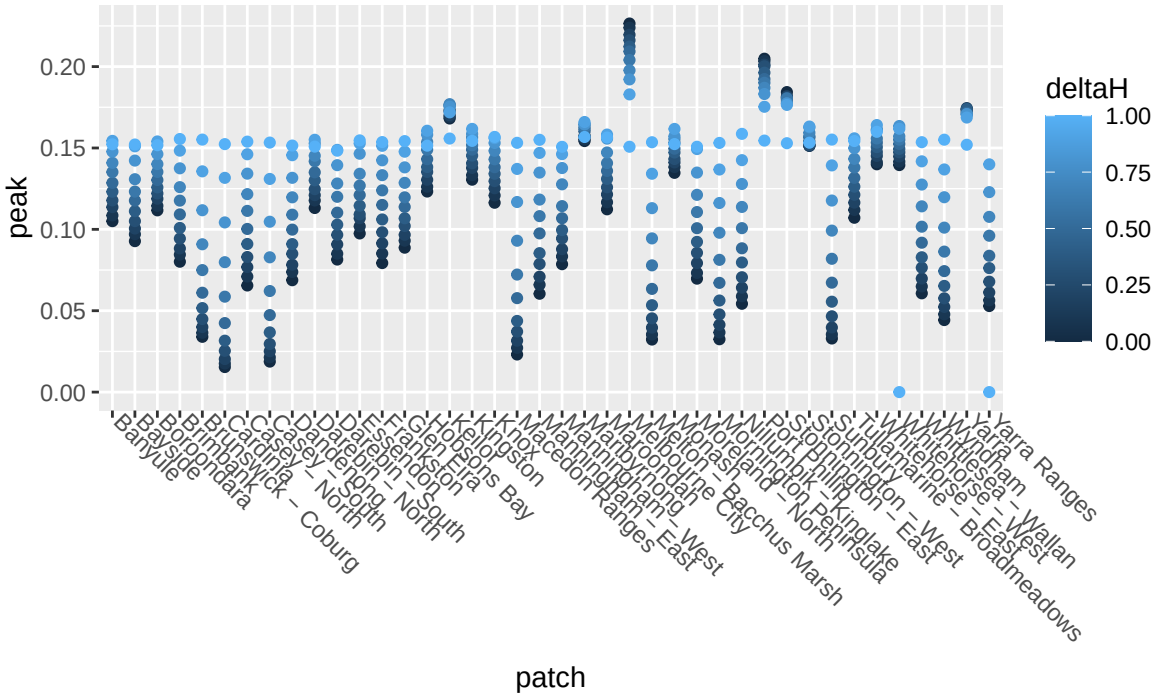


Figure 4.6: showing the peak proportion of infected individuals in each SA3 patch of an OD mixing metapopulation with varying contributions of local mixing(^H)

Part II

Current Study

5 Hierarchical metapopulation models

In **?@sec-MossGmelb**, we observed the construction of a meta-population mixing matrix from empirical origin-destination (OD) data. The patches used in the example were based on the SA3 regions of the Australian Statistical Geography Standard's Statistical Areas (SA) classification (Australian Statistical Geography Standard (ASGS), 2023) (comprising 40 patches for the Greater Melbourne region in that example). In this chapter we will use the larger ASGS SA classification structure to create multiple meta population models of the same geographical region at different spatial resolutions.

5.1 Hierarchical Structure of the SA classification

An important feature of the SA classification structure is that there are multiples scales of classification organised such that lower level SAs are nested within higher level SAs. These levels are denoted by the numbers 1-5, with special groupings for capital city areas (Figure 5.1, blue). For example, several SA3 regions can reside within a single SA4 region, and multiple SA4 regions are contained within the Greater Melbourne Capital City SA (GMCCSA). Moreover, each SA3 region is partitioned into a number of SA2 scale regions, which in turn are partitioned further into a number of SA1 scale regions. Figure 5.2 shows the borders of the 361 SA2, 40 SA3, and 9 SA4 regions of the GMCCSA.

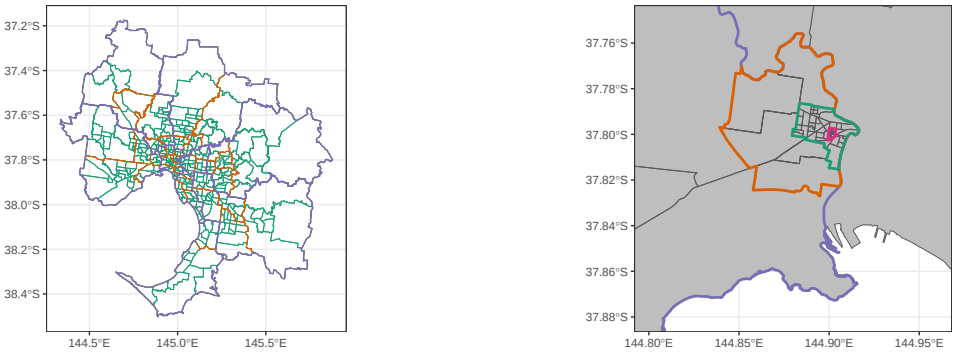
For a specific example, the `Maribyrnong` SA3 region sits inside the `West Melbourne`SA4 region alongside `Essendon` SA3 **?@fig-FootscraySAexample**. Both `West Melbourne` and the neighbouring `Inner Melbourne` SA4 (containing the city center and other SA3 regions), are part of the GSCCSA. Moreover, within `Marybirnong` SA3 there are six SA2 level regions (`Braybrook`, `Footscray`, `Maribyrnong`, `Seddon - Kingsville`, `West Footscray - Tottenham`, `Yarraville`), which can likewise be partitioned in to smaller SA1 level and 'Mesh Block' level regions Figure 5.1. In **?@fig-FootscraySAexample**, we highlight the SA1-SA4 level regions containing the main Footscray CBD (SA1 code '21303134811'; see Figure 5.3).

:::

Note that each red-bounded area represents a higher resolution ('lower' SA level) than the one that encloses it.

Helpfully, besides a common name label, Statistical Areas are also indexed by a structured code representing their classification hierarchy. For example, SA1 regions are denoted by an

Figure 5.1: Structure of The Australian Statistical Geography Standard (ASGS) Statistical Areas (SA) Classification



(a) (b)

Figure 5.2: SA2, SA3 and SA4 borders of the Greater Melbourne Greater Capital City Statistical Area (GMGCCSA) and the the five SA3 blocks that make up the 'Melbourne - West' SA4 region, the six SA2 blocks that make up the Maribyrnong SA3 region, the forty SA1 regionsa that make up the 'Footscray' SA2 region, and the sixteen mesh blocks that make up the central Footscray '21303134811' SA1 region.

11 digit code which can be decomposed into the higher level area codes in which the region sits. Figure 5.3 demonstrates this for the Footscray SA1 region considered above.

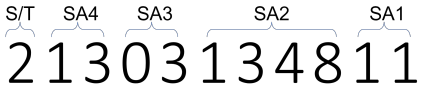


Figure 5.3: Decomposition of theFootscray SA1 region code from **?@fig-FootscraySAexample** into its hierachical SA components. 2 - `Victoria`; 13 - `West Melbourne`; 03 - `Maribyrnong`; 1348 - `Footscray`.

5.2 Comparisons between models of different scales

- Spatial Epidemic models derived from empirical data are limited by the spatial resolution of their source material.
- Comparing models at different scales has been achieved by using different model parameterisations incorporate different sources of empirical spatial information (e.g. Watts et al. (2005), Ajelli et al. (2010), Zachreson et al. (2022)). However, creating equivalent models for controlled testing with different underlying model structures can be difficult.
- We can exploit the hierarchical structure of the ASGS Statistical Areas structure to create meta-population models representing the same overarching spatial structure), but with varying levels of resolution (e.g. by using SA2 scale patches instead of SA3 scale patches).
- To do this, we will parametrically define the mixing matrix M. In doing this, we will explicate two common forms of mixing described in meta-population models.

5.3 Mixing strategies in meta-population models

How can we replicate the homogeneous mixing described for the single patch SIR presented in Chapter 2? We might initially consider a construct a mixing matrix with uniform mixing across patches, i.e.

$$m_{ij} = \frac{1}{n}$$

Where n is the total number of patches. While this mixing matrix would entail homogeneous mixing in a meta-population where $N_i = N_j, \forall i, j$, we have seen in **?@fig-SA3Gmelb_popn**, and it is shown in **?@fig-SA2Gmelb_popn**, that population size is not homogeneous in the GMGCCSA metapopoulation under consideration.

5.3.1 Proportionate mixing

To correct for hetrogeneous patch populations, We can scale mixing coefficients by their patch size i.e. for two patches i and j the mixing coefficient $\phi_{i,j}$:

$$\phi_{ij} = N_j / N_{tot}$$

Which gives mixing matrices shown in **?@fig-GMGCC_PMM**

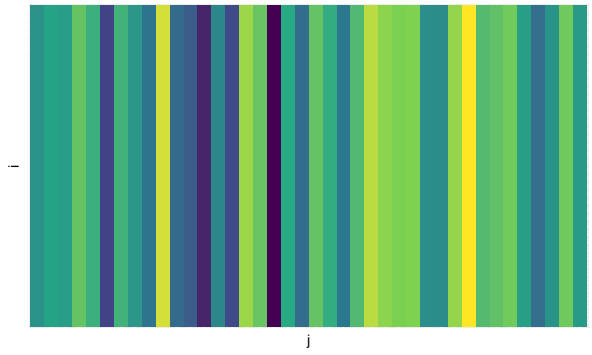


Figure 5.4: SA3

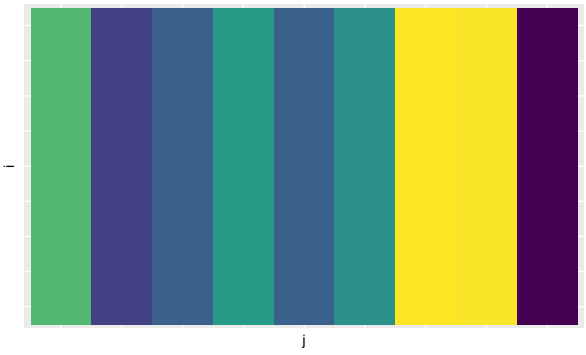


Figure 5.5: SA4

-total size -peak size -Duration of pop prop models with SA2 SA3 SA4 homogeneous patches - Expect equivalence

5.3.2 Statistical Area Mixing structure

We can encode the hierarchical structure of the SA classification in hierarchical block matrices representing within and between SA region mixing. To do so, we specify a set of coefficients $\xi = [\xi_1, ..., \xi_i]$ such that $\sum_i \xi_i = 1$, which determine the proportion of mixing occurs at each level, L, of the spatial hierarchy (i.e. between SA2 regions, between SA3 regions, etc.) . This coefficient is distributed amongst patches occurring in the same level L region, so mixing for any two patches, i and j

$$M_{ij} = \frac{\xi_L}{n^L} \tag{5.1}$$

where

$$j \in S_i^L \ \& \ j \notin S_i^{L-1}$$

and S_i^L is the set of patches in the same level L region as i.

To extend the example from **?@sec-XX**, we can consider a subset of SA2 regions from the GMGCCSA (**?@fig-GMelb_eg_Map**, (**tab-MMeg?**))

Docklands and West industrial are SA2 regions within the 'Melbourne City' SA3 region. Port Melbourne and Port Industrial are SA2 regions within the 'Port Melbourne' SA3 region. Footscray, Seddon-Kingsville, and Yarraville are SA2 regions within the 'Maribyrnong' SA3 region. Newport SA2 is within the Hobsons Bay SA3 region. Port Melbourne and Port Industrial are SA2 regions within the 'Port Melbourne' SA3 region. Furthermore, both 'Melbourne city' and 'Port Melbourne' are within the 'Inner Melbourne' SA4 region, while Maribyrnong and Hobsons Bay SA3 lie inside the 'West Melbourne' SA4 Region.

To model mixing between these SA2 patches (isolated from the rest of the GMGCCSA), we can construct a 8×8 mixing matrix with $\xi = [\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}]$ as follows:

let $i =$ `Footscray`. Since `Footscray` is it's own SA2 region,

$$M_{i,i} = \frac{1}{4}$$

When j is a region in `Maribyrnong` SA3 (alongside `Footscray`), like `Yarraville` or `Seddon-Kingsville`, $M_{i,j}$ will be a proportion of ξ_{SA3}.

$$M_{i,j} = \frac{\xi_{SA3}}{P_i^{SA3} - P_i^{SA2}} = \frac{\frac{1}{4}}{3-1} = \frac{1}{8}$$

When j occurs outside the `Maribyrnong` SA3, but within the `West Melbourne` SA4, like `Newport` SA2

$$M_{i,j} = \frac{\xi_{SA4}}{P_i^{SA4} - P_i^{SA3}} = \frac{\frac{1}{4}}{4-3} = \frac{1}{4}$$

Finally when j is a patch outside the `West Melbourne` SA4, like those in the `Inner Melbourne`SA4

$$M_{i,j} = \frac{\xi_{SA5}}{P_i^{SA5} - P_i^{SA4}} = \frac{\frac{1}{4}}{8-4} = \frac{1}{16}$$

thus for the row M_i, representing the mixing of individuals from Footscray, $\sum_j M_j = \frac{1}{4} + \frac{1}{8} + \frac{1}{8} + \frac{1}{4} + \frac{1}{16} + \frac{1}{16} + \frac{1}{16} + \frac{1}{16} = 1$. Repeating this for all patches i, gives the mixing matrix represented in **?@fig-GMelb_eg_MixMat**

Applying this process to the whole GMGCCSA yields the mixing matrices presented in figure **?@fig-GCC_HMM**

Hierarchical Mixing matrices for the Greater Melbourne Capital City Statistical Area (GMGCCSA) at three levels of spatial resolution.

5.3.3 Population normalised SA mixmat

- in the same way as in **?@sec-proportionatemixing**, the SA structured mixingmatrix given in Section 5.3.2, Equation 5.1, will result in hetrogenous mixing due to the different population per patch **?@fig-gmelbpop**.

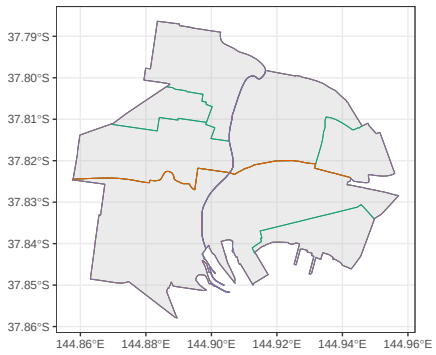


Figure 5.6: **?(caption)**

Figure 5.7: **?(caption)**

- Again, we can rectify this by scaling columns in the mixing matrix by the according population size. Due to the hierarchical structure we can reduce patch counts to population counts

$$M_{ij} = \frac{N_j}{N_i^L - N_i^{L-1}} \xi^L$$

if $j \in S_i^L$ & $j \notin S_i^{L-1}$

where i, j are patches S_i^L is the set of patches in level L N_i is the population of i N_i^L is the population of S_i^L $\xi^L(\delta)$ is the proportion of mixing that occurs within S_i^L but notS_i^{L-1}

5.3.4 blending between population and hiereachy

parameter μ ### Simulation

We can now simulate the spread of disease through the GMGCCSA using the mixing matrices presented in Section 5.3.2 and **?@sec-HPMM**. We will use the same parameters as in **?@sec-sim**, but now with mixing determined by the population normalised mixing matrices defined above.

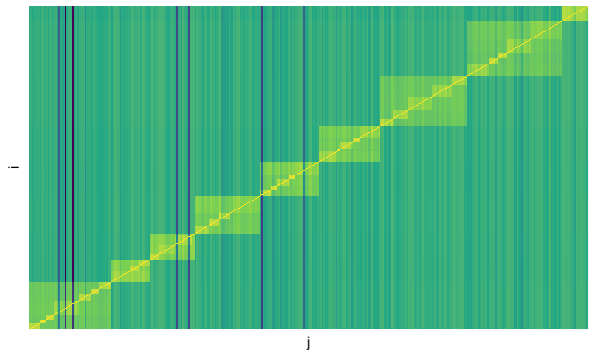


Figure 5.8: fig-gmelbpop_SA1

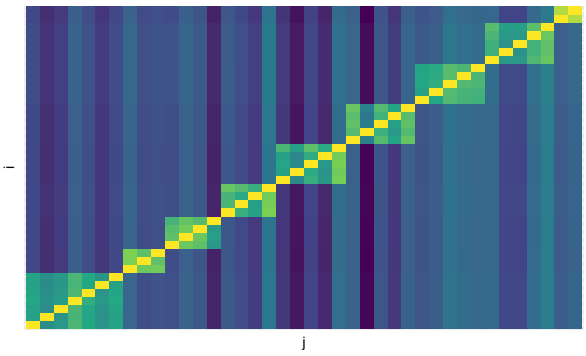


Figure 5.9: fig-gmelbpop_SA2

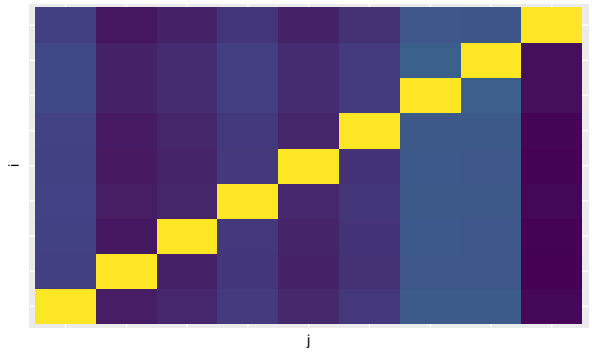
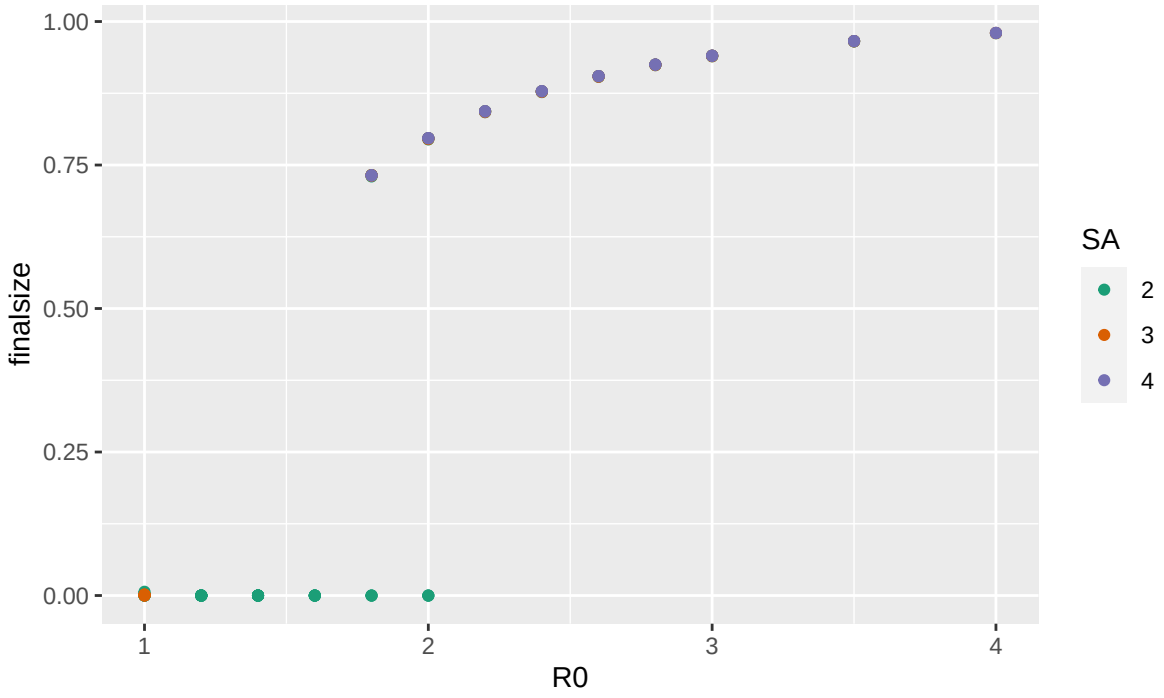


Figure 5.10: fig-gmelbpop_SA3



However, now we can decompose these metapopulation scale outcomes into those of the underlying subpopulation. For example, **?@fig-OD_patchinf_curve_eg**

5.3.5 Note that we can also apply a similar protocol to OD mixing matrices from ?@sec-OD_

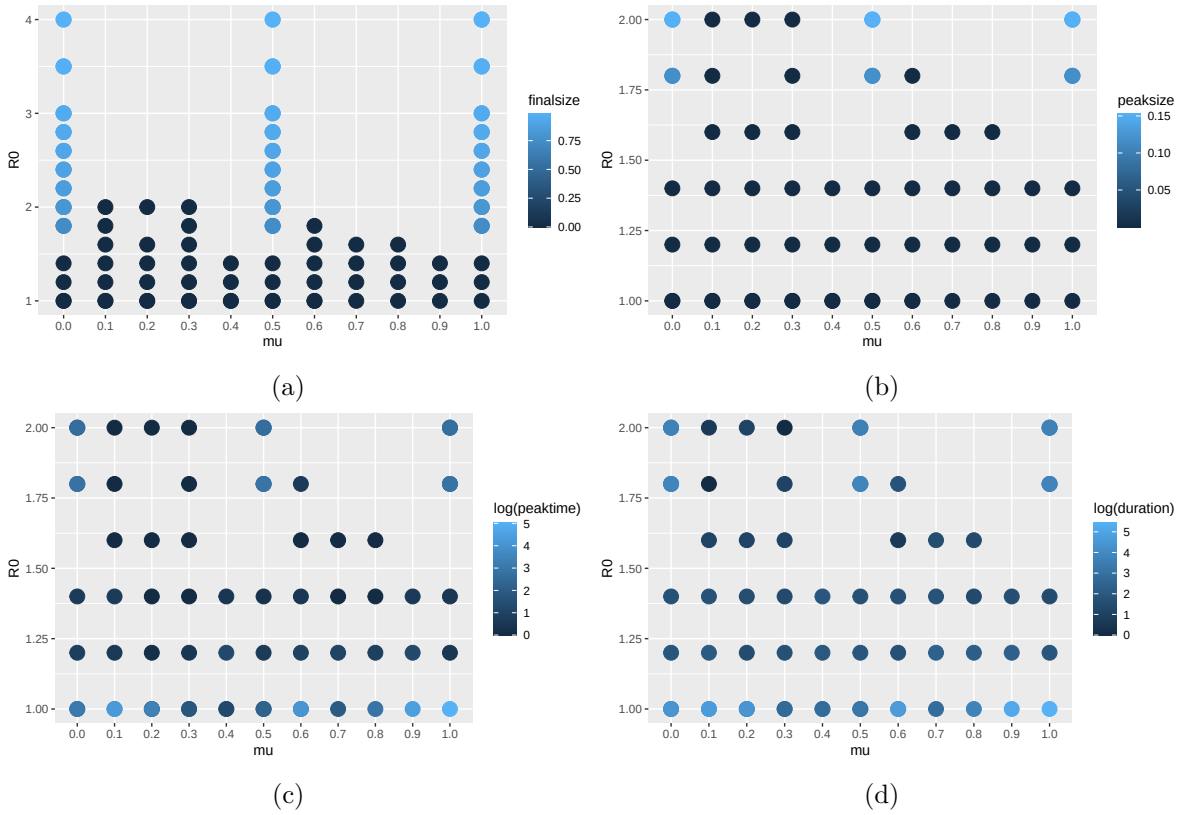


Figure 5.11: Final (a) and peak (b) infection numbers (as a proportion of the entire population), peak time (c) and total duration (d) of a simulated SIR metapopulation model with OD mixing matrix at different values of δ^H and R_0

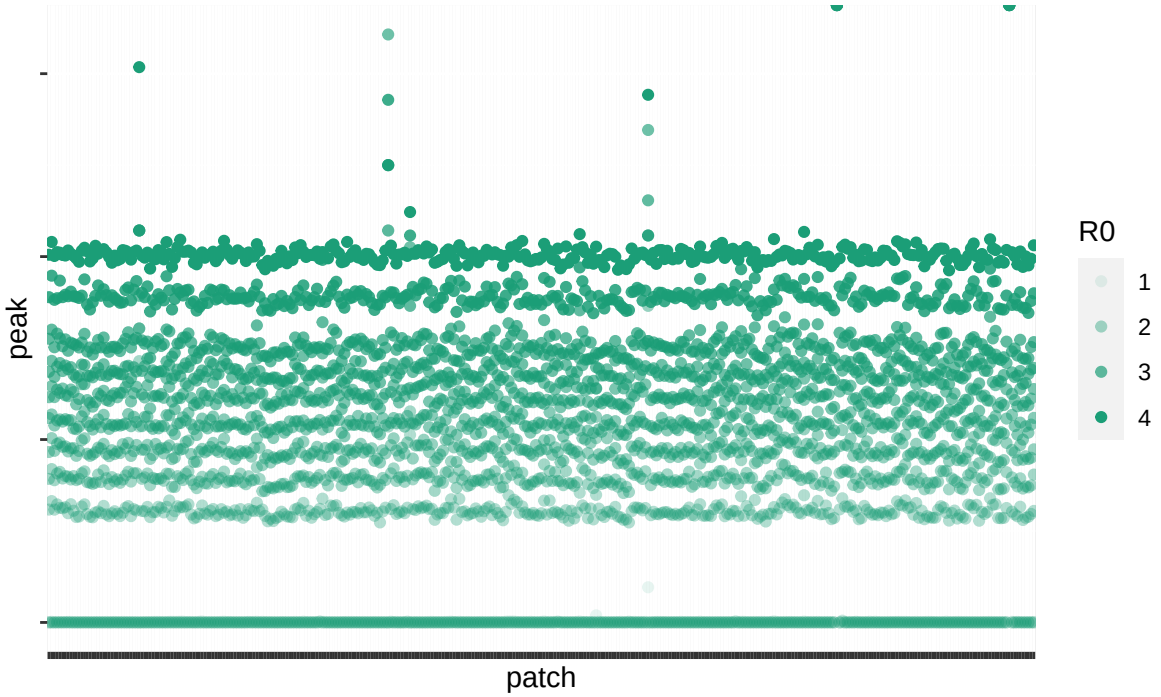


Figure 5.12: showing the peak proportion of infected individuals in each SA3 patch of an OD mixing metapopulation with varying contributions of local mixing(^H)

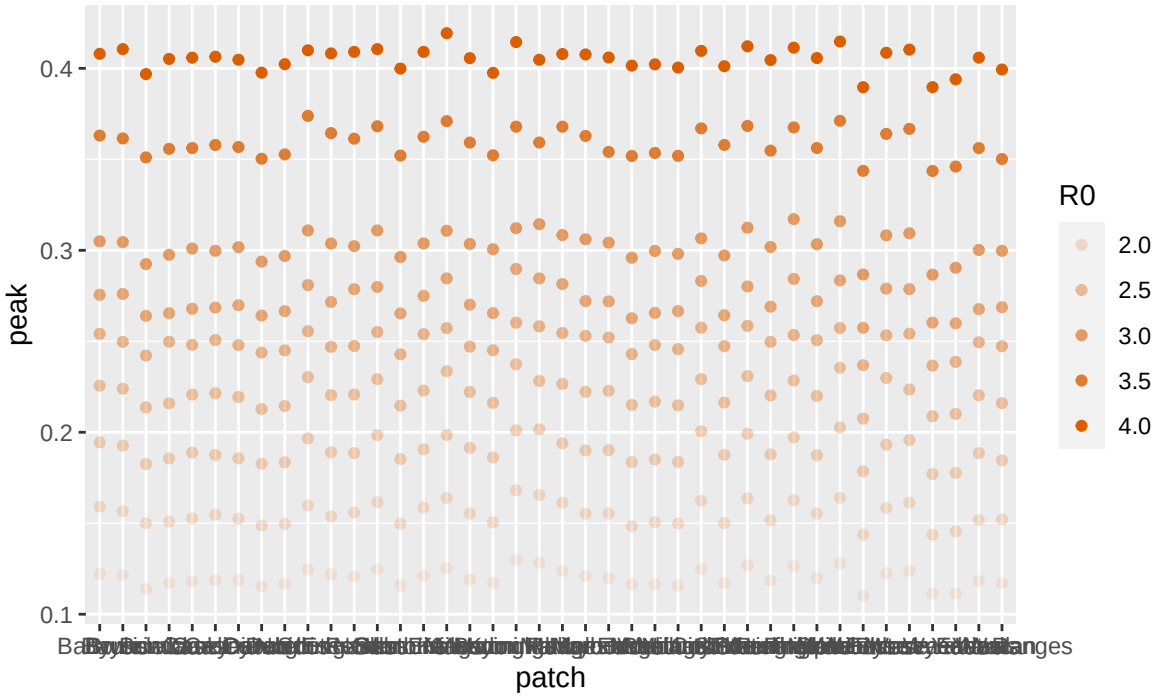


Figure 5.13: showing the peak proportion of infected individuals in each SA3 patch of an OD mixing metapopulation with varying contributions of local mixing(^H)

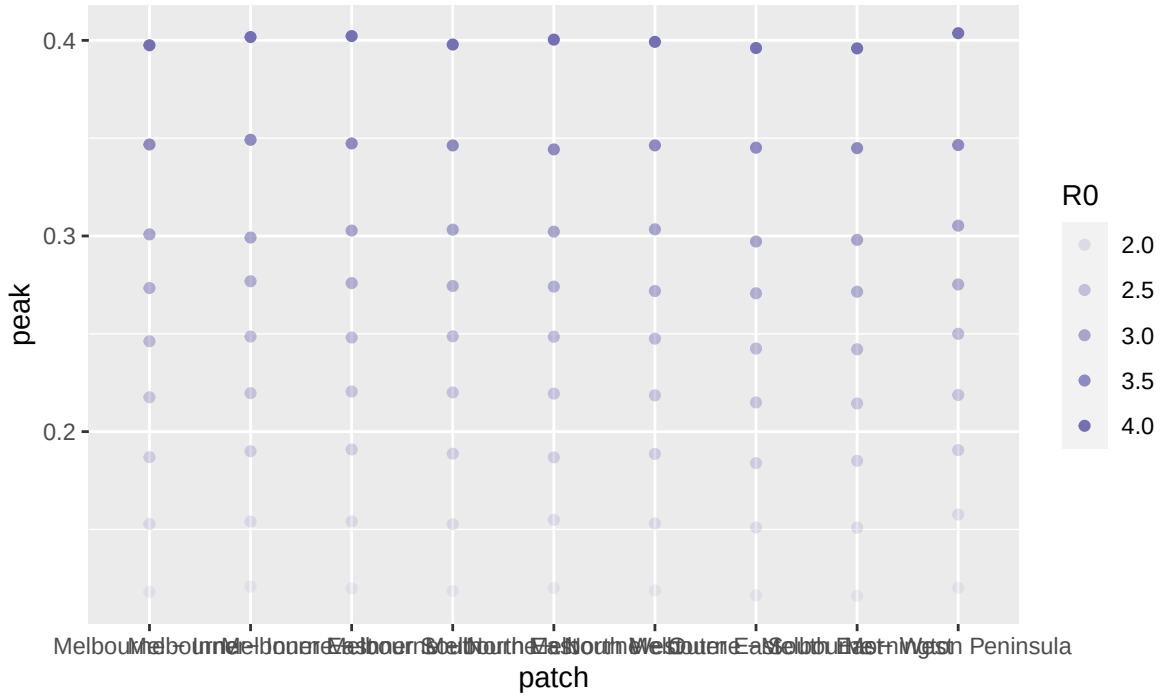


Figure 5.14: showing the peak proportion of infected individuals in each SA3 patch of an OD mixing metapopulation with varying contributions of local mixing(^H)

6 Mixing Matrix Interventions

6.1 Motivation

- Reduce the total epidemic size/effects by intervening in strategic locations
- Spatial resolution allows for precise targeting. Behavioural intervention (restriction of mixing) Mixing matrix is modified during simulation

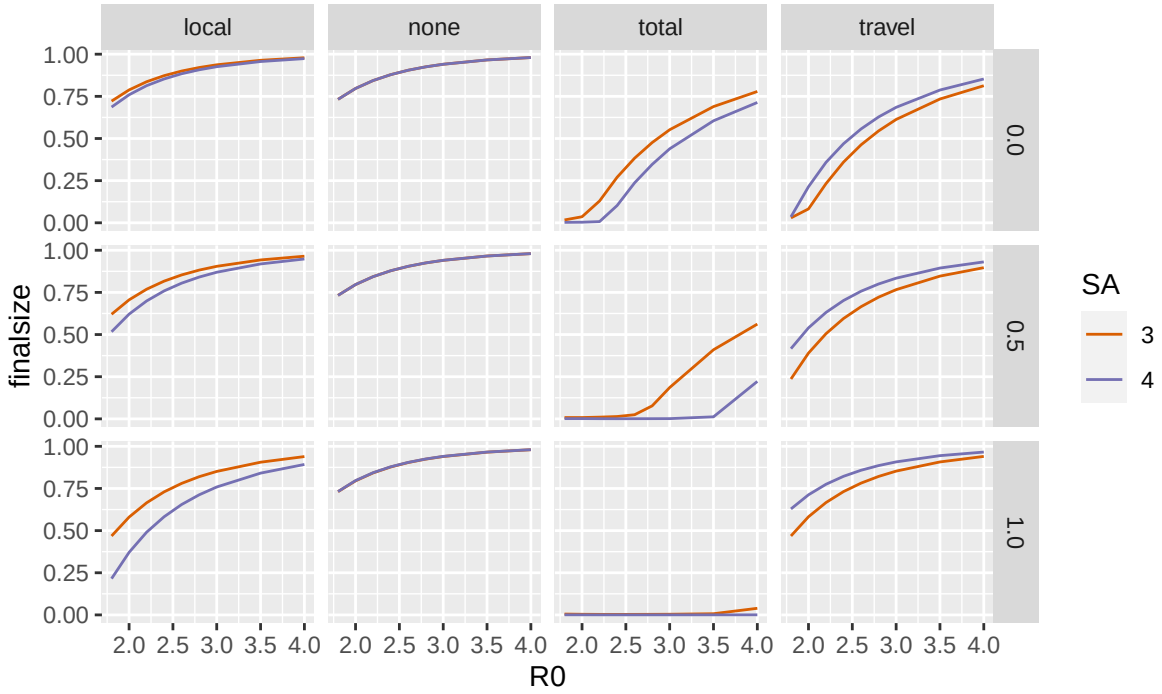
6.2 Formulation

- Parameters
  - Trigger: condition at which intervention occurs
  - Effect:
    - `local` reduce the intra-patch mixing
    - `travel` reduce inter-patch mixing
    - `total` reduce intra- and inter-patch mixing

6.3 Implimentation

6.3.1 local intervention

6.4 Simulation



However, now we can decompose these metapopulation scale outcomes into those of the underlying subpopulation. For example, **?@fig-OD_patchinf_curve_eg**

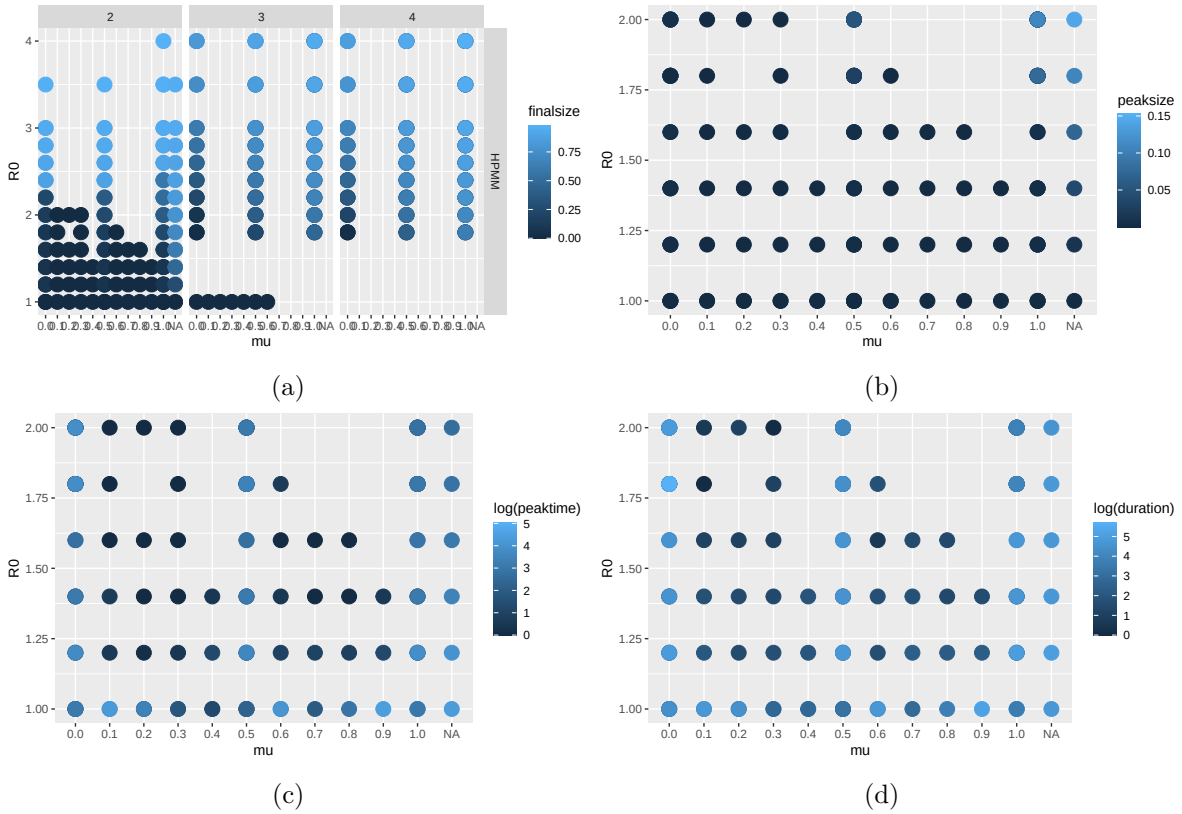


Figure 6.1: Final (a) and peak (b) infection numbers (as a proportion of the entire population), peak time (c) and total duration (d) of a simulated SIR metapopulation model with OD mixing matrix at different values of δ^H and R_0

Bibliography

Australian statistical geography standard (ASGS), (2023).

Ajelli, M., Gonçalves, B., Balcan, D., Colizza, V., Hu, H., Ramasco, J. J., Merler, S., & Vespignani, A. (2010). Comparing large-scale computational approaches to epidemic modeling: Agent-based versus structured metapopulation models. *BMC Infectious Diseases*, *10*(1), 190. https://doi.org/10.1186/1471-2334-10-190 QID: Q30390849

Black, A. J., & Ross, J. V. (2015). Computation of epidemic final size distributions. *Journal of Theoretical Biology*, *367*, 159–165. https://doi.org/10.1016/j.jtbi.2014.11.029 QID: Q41604696

Dobrow, R. P. (n.d.). *Introduction to Stochastic Processes with R*.

Gillespie, D. T. (1976). A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of Computational Physics*, *22*(4), 403–434. https://doi.org/10.1016/0021-9991(76)90041-3

Moss, R., Naghizade, E., Tomko, M., & Geard, N. (2019). What can urban mobility data reveal about the spatial distribution of infection in a single city? *BMC Public Health*, *19*(1), 656. https://doi.org/10.1186/s12889-019-6968-x

Simoni, G., Reali, F., Priami, C., & Marchetti, L. (2019). Stochastic simulation algorithms for computational systems biology: Exact, approximate, and hybrid methods. *WIREs Systems Biology and Medicine*, *11*(6), e1459. https://doi.org/10.1002/wsbm.1459 _Eprint: https://wires.onlinelibrary.wiley.com/doi/pdf/10.1002/wsbm.1459 QID: Q91506994

Teo, M. (2017). *Optimal allocation of vaccines in metapopulations* [PhD thesis].

Teo, M., Bean, N., & Ross, J. V. (2021). Optimised prophylactic vaccination in metapopulations. *Epidemics*, *34*, 100420. https://doi.org/10.1016/j.epidem.2020.100420 QID: Q104571690

Ushey, K. (2022). *renv: Project environments*. https://CRAN.R-project.org/package=renv R package version 0.16.0

Watts, D. J., Muhamad, R., Medina, D. C., & Dodds, P. S. (2005). Multiscale, resurgent epidemics in a hierarchical metapopulation model. *Proceedings of the National Academy of Sciences*, *102*(32), 11157–11162. https://doi.org/10.1073/pnas.0501226102

Zachreson, C., Chang, S., Harding, N., & Prokopenko, M. (2022). The effects of local homogeneity assumptions in metapopulation models of infectious disease. *Royal Society Open Science*, *9*(7), 211919. https://doi.org/10.1098/rsos.211919

A Additional stuff

You might put some computer output here, or maybe additional tables. It is possible to have multiple appendices. Just list them in the appropriate place within `_quarto.yml`.