University Library

MINERVA ACCESS

A gateway to Melbourne's research publications

Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:
Zachreson, C;Chang, S;Harding, N;Prokopenko, M

Title:
The effects of local homogeneity assumptions in metapopulation models of infectious disease

Date:
2022-07-13

Citation:
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), pp.211919-. https://doi.org/10.1098/rsos.211919.

Persistent Link:
https://hdl.handle.net/11343/316965

License:
CC BY

ROYAL SOCIETY OPEN SCIENCE

royalsocietypublishing.org/journal/rsos

Research

Cite this article: Zachreson C, Chang S, Harding N, Prokopenko M. 2022 The effects of local homogeneity assumptions in metapopulation models of infectious disease. *R. Soc. Open Sci.* **9**: 211919.
https://doi.org/10.1098/rsos.211919

Received: 12 December 2021
Accepted: 23 June 2022

Subject Category:
Computer science and artificial intelligence

Subject Areas:
health and disease and epidemiology/computer modelling and simulation

Keywords:
epidemic models, compartmental models, agent-based models, heterogeneity

Author for correspondence:
Cameron Zachreson
e-mail: cameron.zachreson@unimelb.edu.au

The effects of local homogeneity assumptions in metapopulation models of infectious disease

Cameron Zachreson[1,2], Sheryl Chang[2], Nathan Harding[2] and Mikhail Prokopenko[2,3]

[1]School of Computing and Information Systems, The University of Melbourne, Melbourne, Australia
[2]Centre for Complex Systems, Faculty of Engineering, The University of Sydney, Sydney, New South Wales 2006, Australia
[3]Sydney Institute for Infectious Diseases, The University of Sydney, Westmead, New South Wales 2145, Australia

CZ, 0000-0002-0578-4049; NH, 0000-0003-1707-9248; MP, 0000-0002-4215-0344

Computational models of infectious disease can be broadly categorized into two types: individual-based (agent-based) or compartmental models. While there are clear conceptual distinctions between these methodologies, a fair comparison of the approaches is difficult to achieve. Here, we carry out such a comparison by building a set of compartmental metapopulation models from an agent-based representation of a real population. By adjusting the compartmental model to approximately match the dynamics of the agent-based model, we identify two key qualitative properties of the individual-based dynamics which are lost upon aggregation into metapopulations. These are (i) the local depletion of susceptibility to infection and (ii) decoupling of different regional groups due to correlation between commuting behaviours and contact rates. The first of these effects is a general consequence of aggregating small, closely connected groups (i.e. families) into larger homogeneous metapopulations. The second can be interpreted as a consequence of aggregating two distinct types of individuals: school children, who travel short distances but have many potentially infectious contacts, and adults, who travel further but tend to have fewer contacts capable of transmitting infection. Our results could be generalized to other types of correlations between the characteristics of individuals and the behaviours that distinguish them.

© 2022 The Authors. Published by the Royal Society under the terms of the Creative Commons Attribution License http://creativecommons.org/licenses/by/4.0/, which permits unrestricted use, provided the original author and source are credited.

THE ROYAL SOCIETY PUBLISHING

1. Introduction

Computational models of infectious disease dynamics come in two general forms: those that explicitly model each member of the population, and those that treat the population as a statistical ensemble of indistinguishable entities. The most well-established techniques are of the latter type, and compute the dynamics of susceptible and infected populations using compartmental models that specify time-dependent infection rates based on the current numbers of susceptible and infected individuals in the population. A well-known model of this type is the SIR model of Kermack & McKendrick [1]. Such models are advantageous due to their simple formulation and small parameter space, which facilitate the explicit evaluation of the assumptions implicit in the rate equations and make fitting the model to data straightforward for cases in which those assumptions are approximately valid.

There are two primary critiques of these compartmental approaches. The first relates to the assumptions made about the type of dynamics, based on average rate equations, which can be interpreted explicitly in stochastic implementations of compartmental models that effectively express deterministic systems of ordinary differential equations as discrete counting processes. Stochastic implementations offer the capacity for increased realism when the number of infected individuals is small, as it is during the initial growth and final extinction phases of epidemics. These formulations assume exponentially distributed dwell times within compartments (i.e. recovery occurs at a constant average rate). In the SIR model, this assumption produces exponentially distributed *generation intervals* (with mean T_g), which defines the relationship between the basic reproductive number R_0 (the average number of secondary infections produced by a primary case in a completely susceptible population), and the initial exponential growth rate (r) of the infected population [2]. The critique is that there is a universal lack of empirical evidence for these types of dynamics in documented processes of disease transmission, with real generation interval distributions differing markedly from the exponential form. That is, the individual-level disease natural history implied by the SIR model is not justified based on any empirical evidence. More generally, observed dwell times in various disease compartments (i.e. incubation period distributions, or periods of symptom expression) are not exponentially distributed in epidemics of real-world pathogens [3–5]. To address these issues, extensions of the SIR model can provide more realistic representations of generation interval distributions by splitting the infected state into multiple compartments, each with exponentially distributed dwell times (see, e.g. [6,7]). While such extended compartmental models offer the capacity for increased realism, they also contain more parameters that must be calibrated and constrained.

A second common critique is that compartmental models are limited in their capacity to account for heterogeneity in population properties and structure. The compartmental approach assumes each individual in a compartment has identical disease susceptibility, infectiousness and contact frequency with others. Furthermore, in its simplest form, the approach assumes that each individual has a uniform probability of encountering any other individual in the population, an assumption which neglects social behaviours that are known to be important in the spread of infectious disease (i.e. recurrent mobility patterns and social clustering in households or schools). This critique regarding heterogeneity is less fundamental because it can be addressed with implementations that introduce added layers of complexity (additional compartments), and modulate the interaction rates between them to approximate population heterogeneity on a chosen scale.

The most direct approach to addressing both of the above critiques is to simulate the disease natural history and population heterogeneity explicitly with agent-based models (ABMs, also referred to frequently as individual-based models) in which each individual of the population is treated as a discrete, unique entity. While ABMs can, in principle, account for all known heterogeneity in population structure and behaviour, in practice only a small subset of known factors can be included in any given framework. This limitation is due both to the high computational burden of large-scale agent-based simulations, and also to the combinatoric explosion of parameter combinations that occurs as model complexity increases. The choice of which factors to include is typically constrained by the available data, required for model specification. Due to data constraints, it is difficult or impossible to faithfully incorporate all aspects of population heterogeneity that could be important for disease dynamics. On the other hand, the more factors that are included in such models, the more difficult it is to establish which of them are responsible for any observed departures from the dynamics of compartmental models.

Over the last several decades, there has been a sustained effort in the infectious disease modelling community to understand the role of population heterogeneity. This has led to hybrid models that

capture some aspects of population structure explicitly while treating others with homogeneous approximations. For example, the popular metapopulation approach treats discrete subpopulations as homogeneous entities that interact with one-another through coupling equations of various types [8–12]. Another approach, the degree-based mean-field approximation, attempts to incorporate heterogeneous contact patterns into continuum models by making assumptions about the relative timescales of social network dynamics and infectious disease spread [13]. Other work has focused on incorporating nonlinear terms into compartmental frameworks to directly account for the global consequences of heterogeneity on the infectious disease dynamics [14–17].

While many studies have addressed these questions of heterogeneity, it is rare to find explicit comparisons between different modelling approaches (but see e.g. [15,18]). In this work, we compare the results of a complex ABM with a metapopulation model (MPM) derived from the same population data. By aggregating the individuals into subpopulations, we explicitly remove demographic heterogeneity, the nuances of disease natural history progression, and the effects of social clustering. However, the MPM still accounts for medium-scale recurrent mobility patterns, and the basic transmission characteristics of the modelled pathogen (i.e. the basic reproductive ratio R_0, and mean generation interval). After comparing the results of the ABM with both stochastic and deterministic implementations of the MPM, we explore some generic approaches of correcting for observed discrepancies and discuss the results in the context of which factors, included in the ABM, are responsible for the qualitative differences.

Our results indicate that incorporation of nonlinear terms into transmission equations is helpful, but not sufficient to account for discrepancies related to coupling strength (contact rates) between different subpopulations. However, when we systematically decouple the subpopulations of the MPM by decreasing the rate of contact between them, we are able to recover important qualitative aspects of the ABM dynamics. In this case, our results indicate that behavioural heterogeneity manifests both as a weakening of transmission strength with depletion of the susceptible population, as well as a reduction in coupling strength between communities. This is likely due to the different travel behaviours of children and adults. Children, who tend to have higher susceptibility, infectiousness and contact rates, contribute substantially less to long-range population fluxes associated with commuting behaviour. For cases in which individual properties correlate with different types of behavioural tendencies, heterogeneity will likely alter both the dynamics within local regions and the coupling between regions.

2. Methods

2.1. Agent-based model

In general, ABMs attempt to explicitly capture the salient features of individual behaviour in order to simulate non-equilibrium dynamics on the level of heterogeneous collectives.

The ABM of influenza transmission in Australia used in this work has been described in detail in several previous publications [19,20], so only a brief overview will be given here. The aspects of the modelling framework that are most relevant to the present study are the mid- to long-range mobility network [21] and the heterogeneity between students and adults which constitutes the most significant classification in terms of the individual behaviour captured by the model. Notably, the ABM individually simulates approximately 23 M agents, and effective use of it requires high-performance computing facilities.

2.1.1. Population model

The population model is built from the Australian Census of 2016 (Australian Bureau of Statistics). It incorporates local heterogeneity and connectivity at the level of households, household clusters, neighbourhoods and communities. Workplaces and schools provide settings for mixing outside of the local community context, and give rise to the large-scale connectivity patterns captured by the MPM derived here.

Individuals are assigned to households based on the local distribution of household sizes and compositions. The ages of individuals are assigned based on the composition of their households, which effectively reproduces the age distribution computed independently from the census [19]. This procedure produces a calibrated, parsimonious model of the Australian residential population.

Agents between the ages of 18 and 65 years are assigned to workplace groups of 10 individuals based on commuting data provided in the Census [21]. Children between the ages of 5 and 18 years are assigned to schools based on proximity and capacity (they are randomly assigned to the nearest school with available capacity). School locations and capacities were determined from a dataset provided by the Australian Curriculum Assessment and Reporting Agency as described previously [20].

The dynamics of the ABM are computed in time steps of 12 h, which correspond to the 'nighttime' and 'daytime' sets of contact groups in which disease transmission may occur. During the 'daytime' phase, disease transmission events from interactions within workplaces and schools are computed. Individuals who are not assigned a workplace or school mixing environment (approx. 45% of the total population) do not participate in disease transmission during the 'daytime' simulation phase. During the 'nighttime' phase, interactions within local households, household clusters, neighbourhoods and communities are computed, with interactions confined to the scale of statistical area level two (SA2, the community scale, each containing between 1000 and 10 000 individuals [22]). All individuals are assigned residence locations and participate in local disease transmission during the 'nighttime' phase. The timestep of 12 h was selected due to the difference in behavioural features during 'daytime' and 'nighttime' intervals, and is sufficiently short to produce plausible epidemic dynamics. While a shorter timestep would change the disease dynamics produced by the model for a given set of inputs, the comparison we perform in this work is based on the parameters derived from the observed ABM outputs, not from the ABM inputs governing disease transmission. Therefore, we do not expect the choice of ABM timestep to substantially affect the comparison we present in this work.

2.1.2. Disease model

The disease model used in the ABM is designed to simulate influenza infection within an individual and consists of the following agent states: susceptible, latent (exposed, non-infectious), infected (infectious) and recovered. Additionally, 33% of agents are asymptomatic if infected, with infectiousness reduced by a factor of two relative to symptomatic agents. Symptomatic individuals become symptomatic in the days following the end of their latent period, and this occurs in fixed fractions with 30% becoming symptomatic on day 2 of their infection (the day immediately following the latent period), 50% expressing symptoms on day 3 and the remaining 20% becoming symptomatic on day 4 of their infection. The trajectory of asymptomatic or presymptomatic infectiousness follows a bi-linear model that increases for 12 h following the end of the latent period and subsequently declines until reaching zero after 5.5 days. At the moment of symptom onset, infectiousness increases by a factor of two while following the same bi-linear trajectory. This semi-deterministic progression of contagiousness and symptom onset produces a non-trivial generation interval distribution that cannot be captured exactly in standard compartmental frameworks that assume time-independent transition probabilities. In the ABM, no correlations exist between the descriptive properties of an individual and their tendency to become symptomatic (which is assigned homogeneously at random), so we do not explicitly include symptom expression within the compartmental formulation of disease progression used in our MPM.

In the ABM, disease transmission rates are dependent on both age and mixing context, with contacts between children in the home and classroom environments producing the highest chance of transmission between infected and susceptible individuals. The model of disease progression and the age dependence of transmission strength is described in detail elsewhere (see Cliff *et al.* [19]).

2.1.3. Scenario

The ABM simulates pandemic influenza in Australia, and initiates the Australian epidemic with importations (exogenous introduction of index cases) within 50 km of international airports. Relative importation rates are determined by the international arrival numbers at each airport, which produces the initial (seeding) conditions of each epidemic simulation which then proceeds through stochastic transmission of disease using binomially distributed random variables that determine the infection status of each agent, at each 12 h step of the simulation. Each scenario terminates when the infected population reaches 0, or the pre-determined time limit is exceeded.

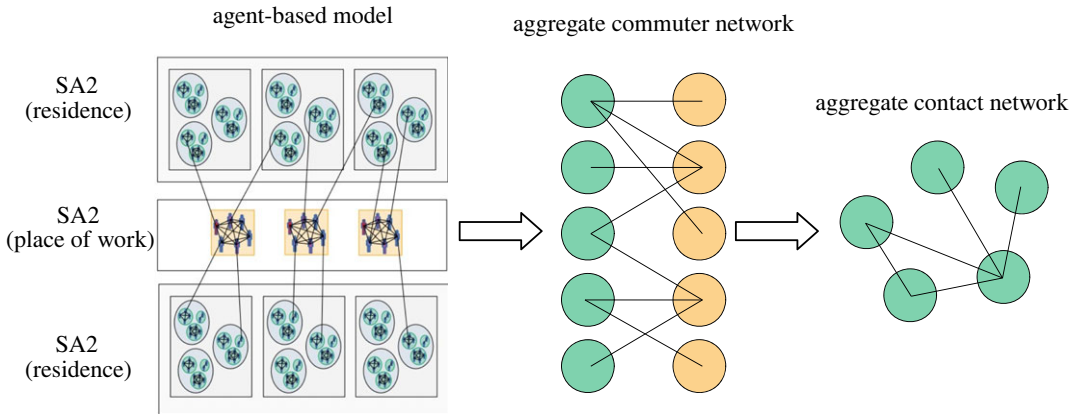


Figure 1. Schematic of the aggregation procedure used to convert the detailed travel patterns in the agent-based model into the large-scale contact network used in the metapopulation model. Travel of individuals from home to workplaces or schools is aggregated into larger regions for a bipartite origin-destination matrix representation. Mass action expressions are then applied in order to generate a mixing matrix from those travel volumes.

2.2. Metapopulation model

The goal in creating our MPM is to preserve the large-scale mobility patterns of the population, while averaging over all local heterogeneity. The basic process is illustrated in figure 1. We do this by systematically building up a transport matrix on the scale of SA2. Each of these n partitions contains approximately 1×10^4, to 3×10^5 individuals. In the ABM, individuals either stay within their home SA2 during the day or they commute to school or work in another area. The MPM consists of a matrix representing the mixing patterns between these locations. The first step in creating such a model is to build a bipartite adjacency matrix $\boldsymbol{\phi}$ describing the proportion of individuals resident in each area i who make regular trips to destination regions j

$$\boldsymbol{\phi}_{ij} = \frac{\mathrm{OD}(\mathrm{SA2}_i \to \mathrm{SA2}_j)}{N_i}, \tag{2.1}$$

where $\mathrm{OD}(\mathrm{SA2}_i \to \mathrm{SA2}_j)$ is the number of travellers residing in area i who commute between their home in area i and their destination in area j, on a typical day (here, N_i is the population of individuals living in area i). This bipartite structure is illustrated on the left-hand side of figure 2. Building the commuter matrix (OD) is a simple process of iterating through each individual in the ABM population, and incrementally increasing the entry of the OD matrix corresponding to that individual's home-destination pair.

To convert the resulting bipartite mobility network into a coupling matrix, we apply standard methods involving simple mass-action expressions. For a given pair of locations, the coupling between them is computed as

$$\boldsymbol{\Phi}_{ij} = \sum_{k=1}^{n} \boldsymbol{\phi}_{ik} \boldsymbol{\phi}_{jk} \frac{1}{N_k^p}, \tag{2.2}$$

where

$$N_j^p = \sum_{i=1}^{n} \boldsymbol{\phi}_{ij} N_i \tag{2.3}$$

is the population *present* in location j during work hours.

The mixing matrix $\boldsymbol{\Phi}$ is a weighted, symmetric, unipartite adjacency matrix representing mixing probabilities between individuals residing in each respective location, so that multiplication of $\boldsymbol{\Phi}_{ij}$ by subpopulation numbers from areas i and j gives the number of contacts between those subpopulations per unit time. Several implicit assumptions about human interaction behaviour are encoded into this mass action derived representation of contact patterns. These can be expressed as the assumption of spatial homogeneity on the scale of SA2, and a time-scale separation assumption that assumes averaging of temporal fluctuations in travel patterns.

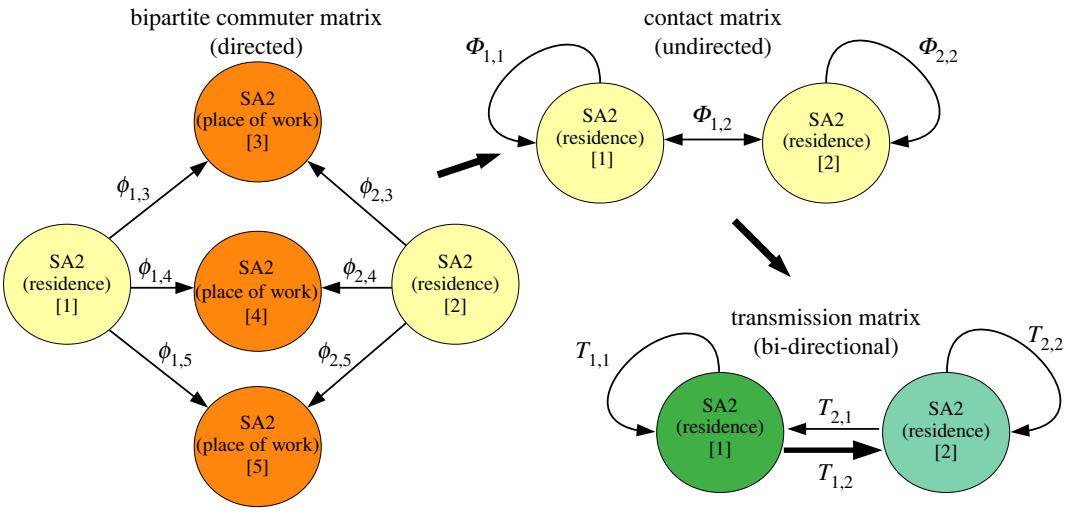


Figure 2. A schematic of the projection process used to build the matrix of coupled metapopulations. From a bipartite commuter matrix, mass-action equations give a symmetric (undirected) contact matrix describing the strength of potential disease transmission between and within areas. Once the epidemic is seeded into this network of metapopulations, a disease transmission matrix describes the rates at which infection may spread between coupled areas with transmission computed based on the populations of infected and susceptible individuals, and the strength of the coupling between each location.

To convert this model of coupled metapopulations into a full infectious disease model, we first need to define our disease dynamics. For simplicity, generality and approximate correspondence to the ABM, we will use the canonical SIR dynamics which, for a single homogeneously mixing population, are described by

$$\frac{\mathrm{d}S}{\mathrm{d}t} = -\beta SI, \quad \frac{\mathrm{d}I}{\mathrm{d}t} = \beta SI - \gamma I, \quad \frac{\mathrm{d}R}{\mathrm{d}t} = \gamma I, \tag{2.4}$$

in which S is the proportion of susceptible individuals, I is the proportion of infected individuals and R is the proportion of recovered individuals, while β and γ are the transmission and recovery rates, respectively. Strictly speaking, β incorporates also the rate at which contacts capable of transmitting infection occur (in some descriptions, the contact rate is included as a separate parameter).

The homogeneous transmission dynamics described by equation (2.4) can be applied to a network of coupled subpopulations by using Φ to describe the contact rates between them. Computing transmission between communities requires multiplication by infected and susceptible populations in each region, a transmission rate per contact, and a contact rate per unit time (here, these two terms are absorbed into the transmission rate β). The rate of transmission from a single infected individual residing in area j to the susceptible population of area i is computed as

$$T_{ij} = \beta \Phi_{ij} S_i N_i, \tag{2.5}$$

in which S_i is the proportion of susceptible individuals in subregion i, and N_i is the total population of subregion i. This formulation allows us to build several different implementations of MPMs, the results of which can then be compared to those of the ABM. Here, we discuss a deterministic discrete-time implementation, a stochastic discrete-time implementation and a stochastic continuous-time implementation.

2.2.1. Deterministic, discrete time

To compute the dynamics of the deterministic version of this MPM, all that is required is the combination of the transmission and recovery terms, allowing numerical integration of the coupled rate equations. The vector of infected populations $\boldsymbol{I}$ at time $t + \Delta t$ is computed as

$$\boldsymbol{I}(t + \Delta t) = [\mathbb{I} + \Delta t(T - \gamma \mathbb{I})]\boldsymbol{I}(t), \tag{2.6}$$

where $\mathbb{I}$ is the identity matrix of size n (note that we have used I to represent the fraction of infected individuals in equation (2.4), while $\boldsymbol{I}$ is a vector of infected populations in each subregion). The infection numbers in each location depend on the incoming transmissions from all other areas, local transmissions and local recovery of individuals. Since these are continuous dynamics, non-integer

values of infected populations are possible. While this condition may be acceptable when the infected and susceptible populations are both large, it makes interpretation difficult in the initial stages of the epidemic, and leads to substantial discrepancies in the results of numerical simulations, as we will discuss in the results sections.

2.2.2. Stochastic, discrete time

In the discrete-time stochastic description of the structured SIR dynamics, the force of infection on the subset of individuals living in region i is computed from the mixing terms given in equation (2.2), aggregated over each location's neighbours j, multiplied by the number of ill individuals in each neighbouring region, the transmission rate β and the chosen time interval Δt. This gives the individual-level probability of transitioning from the susceptible to infected state, computed independently for each SA2 subregion i

$$p(S \to I \mid S,\ \mathrm{SA2}_i) = 1 - \exp\left(-\beta \Delta t \sum_j \Phi_{ij} \boldsymbol{I}_j\right), \tag{2.7}$$

while recovery only depends on the rate γ and the time interval

$$p(I \to R \mid I) = 1 - \exp(-\gamma \Delta t). \tag{2.8}$$

2.2.3. Stochastic, continuous time

To ensure that the dynamics computed by the discrete-time implementation were sufficiently free of artefacts due to time discretization, we implemented a continuous stochastic version of the model using the Gillespie algorithm. To do so, at each time t the probability of each possible next event was computed, and the next event chosen by inverse-transform sampling of the corresponding cumulative distribution function. The delay δt between each subsequent event was computed in the standard way by sampling from the exponentially distributed inter-event times $\delta t \sim \rho\, \mathrm{e}^{-\rho}$ where ρ is the sum over all individual event rates. In the system described here, there are $2n$ possible events, consisting of either recovery or infection events, in each of the n subregions. By comparing the results of the continuous-time model to those of the discrete-time stochastic model, we determined that a discrete-time step of $\Delta t \leq 0.1$ d was sufficient for convergence of the discrete dynamics. Because we did not identify any significant differences between the results of the continuous-time simulation and those of the stochastic discrete-time model, we do not report the results of the continuous-time model, but have included the source code in the online Zenodo database (DOI:10.5281/zenodo.6486795).

2.3. Extensions to the SIR framework

After discussing the differences between the results produced by the ABM and those produced by a comparable MPM, we introduce two extensions to the standard SIR modelling framework. These extensions modify two aspects of the dynamics that are not captured in the MPM framework described above.

2.3.1. The saturation parameter λ

The first of these aspects is the tendency for local depletion of the susceptible population [16]. In tightly connected clusters of individuals such as households or workplaces, disease spread is enhanced by the higher-than-average contact rates between the individuals in the cluster. However, such clustering also produces lower-than-average transmission rates between such groups, and this leads to depletion of the local susceptible population during disease transmission. Mathematically, we attempt to capture such effects in a phenomenological way by introducing a saturation parameter $\lambda > 0$ that decelerates transmission as the susceptible population decreases

$$T_{ij} = \beta_i \Phi_{ij} S_i, \tag{2.9}$$

where

$$\beta_i = \beta [S_i]^{\lambda}, \tag{2.10}$$

which produces a system in which the dynamics are initially identical to those of the base model but transmission strength decreases more quickly as the susceptible population is depleted.

2.3.2. The decoupling parameter σ

In addition to this saturation parameter, we explore the effects of altering the properties of the mixing matrix to enhance local mixing and inhibit inter-regional transmission. By doing so, transmission within regions will accelerate and transmission between regions will slow down. Our rationale for introducing such an effect is that we would like to understand the phenomena produced by correlation between decreased travel behaviour and increased social mixing behaviour of school children, without explicitly simulating the underlying heterogeneity. In the ABM, school children have larger mixing groups (classroom groups, grade groups and school groups), but typically remain local to their home region during the day. To modify our MPM in order to adjust for these effects, we alter the commuter flows described in the OD matrix ϕ, by uniformly moving commuters from off-diagonal components into the diagonal component (those who remain in their local areas). To manipulate the degree to which this occurs, we introduce the parameter σ, which modulates the original OD matrix as follows:

$$\left.\begin{aligned} \tilde{\phi}_{ij} &= \phi_{ij}(1-\sigma)\,,\ i \neq j \\ \text{and} \quad \tilde{\phi}_{ii} &= \phi_{ii} + \sum_{i \neq j} \sigma \phi_{ij}, \end{aligned}\right\} \tag{2.11}$$

so that a fraction $\sigma \in [0, 1]$ of travellers remain within their local regions instead of transiting outside of them. These traveller flows are used to compute the mixing matrix Φ, which determines the transmission strength between regions. The decoupling parameter σ attenuates the strength of disease transmission between regions by reducing the rate of contact between individuals residing in different locations.

While we use a static representation of population mobility, and adjust the traveller volumes directly before computing mixing rates, the extension we propose is directly analogous to dynamic representations that model stochastic travel patterns based on average departure and return rates. For example, in the formulation proposed by Sattenspiel *et al.* [23], if spatially homogeneous departure and return rates are assumed, then our correction factor σ can be mapped to a global attenuation of departure rates from home locations (s) as

$$s^* = \frac{s\rho}{(z-1)s + z\rho}, \tag{2.12}$$

where s^* is the attenuated departure rate from home locations, ρ is the spatially homogeneous return rate from destination regions, and $z = (1-\sigma)^{-1}$. Here, z alters average home and destination populations by changing the relative rates of departures and arrivals, rather than by adjusting a static snapshot of the travel topology.

While it is useful for generality to connect our decoupling parameter σ to the alternate formulation above, we note here that periodic daily travel patterns in which approximately half of the working population's waking hours are simultaneously spent in a destination region and half in a home region are assumed in the ABM. These types of mobility patterns cannot be captured under the time-scale separation assumption used in our MPM or in more explicit steady-state formulations such as the one developed by Sattenspiel *et al.* [23].

In order to develop a first approximation of the decoupling due to time-use assumptions of the ABM, we first note that the 'nighttime' phase of the ABM corresponds to a travel topology with $\sigma_{\text{night}} = 1.0$ (interactions occur only within local regions), while the 'daytime' phase corresponds to $\sigma_{\text{day}} = 0$ (interactions occur in destination regions). Therefore, it is plausible that a correction of $\sigma = 0.5$ could account for this model of time use, as a time-average of the periodically varying transmission dynamics of the ABM. However, the ABM additionally assumes that individuals who are not assigned working groups (45% of the total population) do not contribute to disease transmission during the daytime phase. Therefore, if $h_{\text{tot}} = h_{\text{night}} + h_{\text{day}}$ is the total number of person-hours accounted for in the ABM per day, where $h_{\text{night}} = 12N$ and $h_{\text{day}} = 0.55 \times 12N$, then the effective decoupling due to the time-use assumptions of the ABM is

$$\sigma_{\text{ABM}} = \sigma_{\text{night}} \frac{h_{\text{night}}}{h_{\text{tot}}} + \sigma_{\text{day}} \frac{h_{\text{day}}}{h_{\text{tot}}}, \tag{2.13}$$

which gives a value of $\sigma_{\text{ABM}} = 0.645$, a substantial departure from the coupling strength assumed by the

MPM formulation. This first approximation does not account for correlations between mobility range and disease transmission rate, to which we attribute additional decoupling, up to $\sigma \approx 0.9$ (see results).

With these two degrees of freedom (σ and λ) added to the base MPM configuration, we can tune the relatively simple MPM model to closely match the dynamics produced by the complex ABM.

2.3.3. Adding a latent period (SEIR)

In addition to the extensions described above, we perform an analysis of model uncertainty with respect to the SIR compartmental formulation. As the ABM includes a latent period of one day between infection and the onset of infectiousness, we explore an alternative formulation that includes an exposed (E) compartment into which newly infected individuals in the MPM reside before stochastically transitioning to the infectious (I) compartment with an average dwell time of 1 day (exponentially distributed). Similar to SIR dynamics, in the canonical SEIR model individuals progress between the compartments $S \to E \to I \to R$, with the compartment populations following:

$$\frac{\mathrm{d}S}{\mathrm{d}t} = -\beta SI, \quad \frac{\mathrm{d}E}{\mathrm{d}t} = \beta SI - \kappa E, \quad \frac{\mathrm{d}I}{\mathrm{d}t} = \kappa E - \gamma I, \quad \frac{\mathrm{d}R}{\mathrm{d}t} = \gamma I, \tag{2.14}$$

where the new parameter κ determines the rate at which individuals transition from the latent infected (E) state, to the infectious (I) state. While residing in the exposed compartment, infected individuals cannot transmit the contagion to others, neither can they become infected. Here, we employ a stochastic, discrete-time implementation of the SEIR model in order to assess the robustness of our results to the choice of SIR compartmental dyanamics in our MPM formulation.

2.4. Matching the models

There are several fundamental differences between the ABM and the MPM. These relate to the details of disease natural history (i.e. time-dependent recovery probabilities in the ABM), heterogeneity of individual behaviours which determine structured mixing patterns within SA2 regions (which are considered as homogeneous populations in the metapopulation framework), and social clustering across spatial regions due to habitual contact patterns between travellers (i.e. the interactions of co-workers).

In order to approach the question of how these fundamental factors affect the results of simulations, we first calibrate the two frameworks to match important quantities that must be conserved across models to facilitate a fair comparison. Despite the different levels of model complexity, we establish fairness by matching the initial conditions, basic reproductive ratios R_0 and global exponential growth rate of incidence trajectories. Importantly, this choice of compartmental framework and method of matching the models produces differently shaped generation interval distributions, but only slightly different mean generation intervals.

2.4.1. The basic reproductive number R_0, generation interval and initial growth rate r (SIR)

An essential quantity describing any model of contagion spread is the basic reproductive number R_0, the average number of secondary cases produced by a typical primary infection (index case), in a completely susceptible population. For the ABM, this quantity is difficult to compute analytically due to the complex nature of the population structure and disease natural history model. However, it is straightforward to estimate R_0 from simulations. This allows calibration with respect to the single independent parameter κ, the transmission multiplier, in order to approximate a set value of R_0. We used this technique to calibrate the ABM so that $R_0 = 2.23$, that is, for typical index cases in an entirely susceptible population, there occur 2.23 secondary infections on average. The calibration procedure is described in detail in our previous paper, and takes into account the age-structured distribution of primary infections [24]. In the ABM, the duration of illness is 5.5 days, with a constant one-day latency period applied to all infections. The resulting generation interval, T_g, the average time for the infection to propagate between infectious and susceptible individuals, was computed as $T_g \approx 3.4$ days (the ensemble of index cases and transmission events used to generate our estimate of R_0 was also used to estimate generation interval). In SIR dynamics, T_g is related to the reproductive ratio and the initial exponential growth rate r as

$$T_g = \frac{R_0 - 1}{r}. \tag{2.15}$$

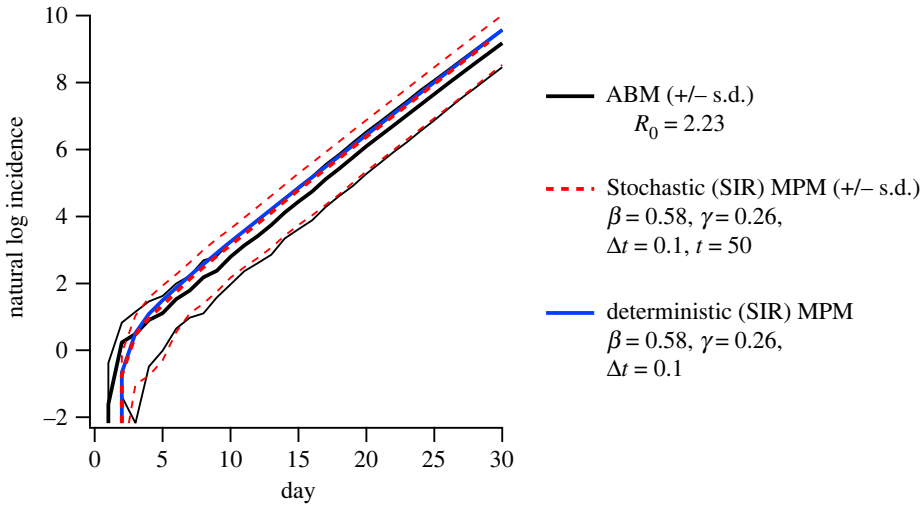


Figure 3. Calibration of the initial conditions of the metapopulation models to match those of the agent-based models. The plot shows the natural logarithm of incidence (number of new illnesses) counted on each day of the simulation. By calibrating the recovery rate of the SIR model to the generation time expected to match the exponential growth rate observed in the ABM, the initial growth of the incidence rate in the ABM is well matched to both the deterministic and stochastic implementations of the metapopulation model.

An average growth rate $r = 0.32$ day^{-1} was estimated by calculating the mean of the incidence from the first 30 simulation days of 30 independent realizations of the ABM, and fitting to an exponential growth function. Using this growth rate, and the estimated value $R_0 = 2.23$, equation (2.15) gives $T_g = 3.84$ days, which is close to the generation interval estimated from the ABM results. In the SIR dynamics, the inverse generation interval $1/T_g$ is equivalent to the recovery rate γ. In the SIR model, $R_0 = \beta/\gamma$, so initializing the SIR MPM with $\gamma = 0.26$ d^{-1} and $\beta = 0.58$ d^{-1} should give an initial dynamics approximately equivalent to those of the ABM (note that because the matrix Φ is row-stochastic, the R_0 expression for the MPM is equivalent to that of the homogeneous SIR model). These parameters indeed produce almost identical initial growth dynamics to the average over ABM trajectories, in both deterministic and stochastic MPM implementations, as shown in figure 3.

2.4.2. The basic reproductive number R_0, generation interval and initial growth rate r (SEIR)

In the case of the SEIR model extension, the process for calibrating the initial conditions is the same as for the SIR model, except that the analytic formulation of the generation interval differs due to the inclusion of the additional exposed compartment (see, e.g. [25]). For the same $R_0 = 2.23$, initial growth rate $r = 0.32$ and an average latent (exposed) period of $1/\kappa = 1$ day, the SEIR formulation produces a transition rate from the infectious (I) to recovered (R) compartments of

$$\gamma = \frac{[r^2/\kappa] + r}{R_0 - [r/\kappa] - 1}, \tag{2.16}$$

for an infectious period of $1/\gamma = 2.15$ days and an average generation interval of

$$T_g = \frac{1}{\kappa} + \frac{1}{\gamma} = 3.15 \text{ days}, \tag{2.17}$$

the sum of the average latent and infectious periods (recalling that κ, the rate of transition from exposed to infectious, is set to 1 d^{-1} to match the latency period of the ABM). We note that the SIR and SEIR formulations both produce generation intervals that are near that of the ABM, with the SIR and SEIR models over- and underestimating the generation interval of the ABM (respectively). While either model could provide a fair comparison for the purposes of this study, we apply the SIR formulation to the main analysis, including the SEIR extension as a test of the robustness of our comparison to the choice of compartmental framework.

2.4.3. Initial conditions

The results of infectious disease models in heterogeneous populations can be sensitive to the distribution of index cases. We took the following steps to ensure a fair comparison between models. In the ABM, the international pandemic scenario corresponds to a steady influx of index cases near international airports. Specifically, each international arrival produces a 0.004% chance of a new index case occurring within each member of the susceptible population residing in the SA2 regions within 50 km of the corresponding airport. This incoming force of infection is continuously applied throughout the simulation. Mimicking these conditions in the metapopulation framework is straightforward, because the seeding procedure is carried out on the same scale (SA2) in the ABM. In the discrete-time stochastic implementation, the procedure is identical to that of the ABM, while in the deterministic implementation the incoming force of infection is interpreted as a local rate of infection, applied in addition to the transmission terms described by equation (2.9). In the continuous-time stochastic implementation, the seeding rates are added to the transmission and recovery rates when computing the global event rate, and appended to the infection rate terms for each of the affected SA2 regions when computing the next-event CDF. After matching the initial conditions and growth dynamics, the results of the ABM can be compared to those of the MPM.

2.4.4. Censoring of case counts

To facilitate a fair comparison of incidence trajectories, we note that the one-day latency period and stochastic symptom expression in the ABM produces right-censoring of case counts (counted at symptom onset) with an observed incidence of

$$I_{\text{obs}}(t) = 0.2I(t-4) + 0.5I(t-3) + 0.3I(t-2), \tag{2.18}$$

which we apply as a sliding observation kernel to the raw infection incidence counts produced by the MPM in order to compare the trajectories produced by the models. We additionally apply a global attenuation of 0.67 (the symptomatic fraction) to all MPM infection incidence counts.

3. Results

3.1. Initial comparison of metapopulation and agent-based models

Without correcting for population heterogeneity, the MPM calibrated to the same initial growth rate and initial conditions as the ABM overestimates peak prevalence and underestimates epidemic duration. This general finding is true for both the deterministic and stochastic implementations of the MPM. Of the two implementations, the stochastic MPM is qualitatively more consistent with the ABM (figure 4).

3.2. Comparison of extended metapopulation models and agent-based model

To account for the effects of heterogeneity, equations (2.10) and equation (2.11) introduce the correction factors λ and σ, respectively. Varying the transmission attenuation parameter λ decreases peak incidence without altering the initial growth rate, as shown in figure 5*a*. On the other hand, altering the inter-region mixing parameter σ marginally reduces peak incidence and slightly increases peak timing, while substantially broadening the incidence curve and, for large enough values ($\sigma \approx 0.9$), produces observable bi-modality, a feature of the ABM produced by weak connections between urban and regional areas, and discussed at length in our previous work (figure 5*b*) [20].

By tuning the attenuation parameter λ and the inter-region decoupling parameter σ, the mean incidence trajectories of the MPM can be made to closely match those of the ABM (figure 6). For attenuation $\lambda = 0.6$ and inter-region mixing reduction of $\sigma = 0.9$, the growth rate, peak incidence (and its run-to-run standard deviation), and the bimodal character of the ABM are reproduced in the MPM implementation. The inter-region mixing parameter σ produces a delay in the peak incidence. This is likely due to the limitation of our homogeneous decoupling procedure that applies uniform relative alteration in diagonal versus off-diagonal travel matrix elements, over all regions (equation (2.11)). That is, the decoupling procedure used in this work applies the same function between tightly coupled urban regions as it does to weakly coupled rural–rural, urban–rural and inter-city connections. In the ABM, bimodality of incidence trajectories results from heterogeneous decoupling in which intra-urban locations remain tightly coupled. As these regions are responsible for the first

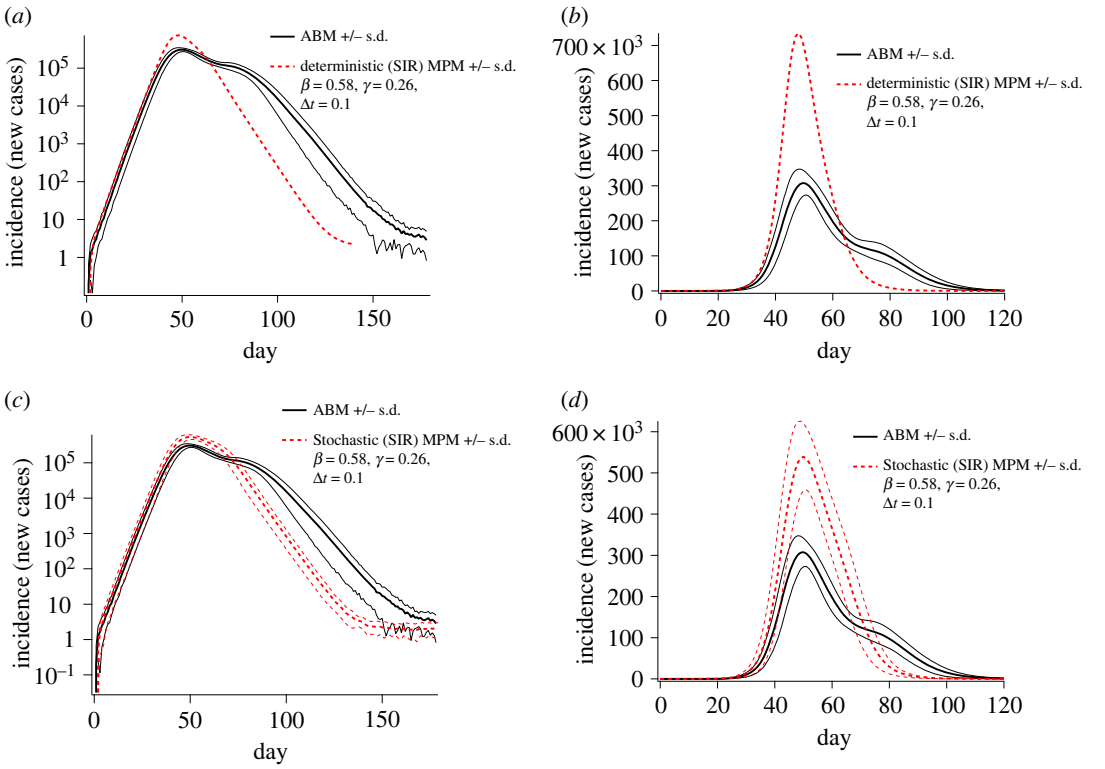


Figure 4. Epidemic incidence curves comparing the results of the agent-based model with those of the deterministic and stochastic implementations of the metapopulation model. Black curves correspond to the ABM results (upper and lower curves show ± s.d.). Red curves in subplots (*a*) and (*b*) correspond to log-scaled and linearly scaled incidence (respectively) for the deterministic implementation, while red curves in subplots (*c*) and (*d*) show log-scaled and linearly scaled incidence computed by the stochastic implementation. For the stochastic implementation, the average and standard deviation were computed from 50 independent runs.

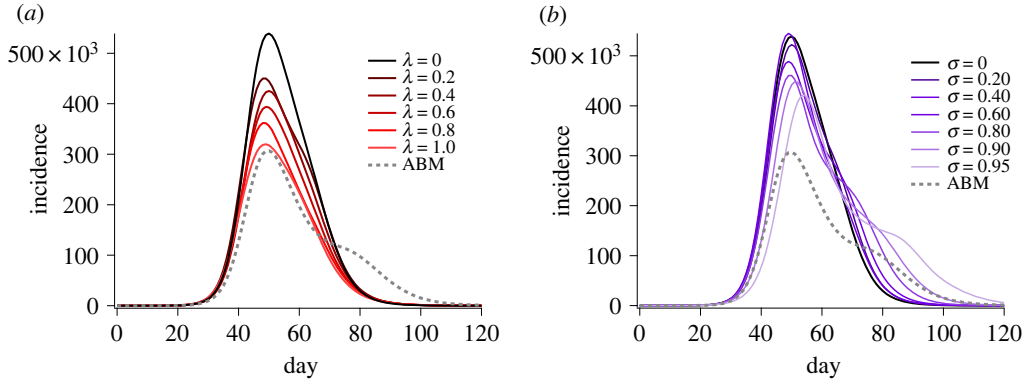


Figure 5. Incidence curves produced by independently varying (*a*) the susceptibility attenuation parameter λ and (*b*) the inter-community decoupling parameter σ, in the extended stochastic metapopulation model.

incidence peak (see [20]), decoupling them in the MPM shifts the global peak to later time points. This global peak shift limits the degree of decoupling that can be applied while still producing a satisfactory match to the ABM incidence trajectories.

Finally, to investigate the qualitative correspondence between the dynamics produced by the ABM and adjusted MPM, we visualized the spatio-temporal distribution of illness prevalence as computed by both models. Examples are shown in figure 7. In both models, outbreaks begin in densely populated urban areas, where international arrivals are concentrated and epidemic seeding is more likely. This leads to an initial peak concentrated in the larger urban centres, followed by a later second peak corresponding to urban regions and smaller urban areas with lower levels of international air traffic. As expected based on the correspondence shown in figure 6, the dynamics of the ABM are born out in the adjusted MPM.

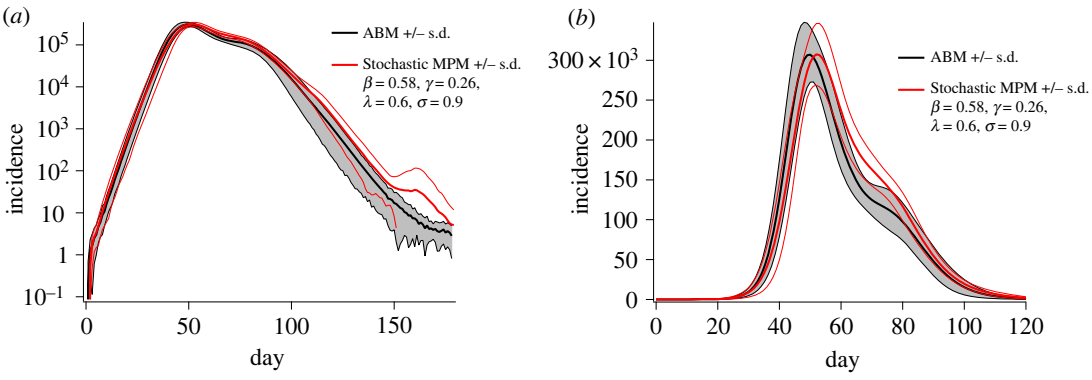


Figure 6. Log-scaled (*a*) and linear-scaled (*b*) incidence curves comparing the ABM with the extended stochastic MPM after fine-tuning the saturation parameter λ and inter-community decoupling parameter σ to qualitatively match the ABM results. Initial growth rates and peak incidence are very closely matched between the models. Peak timing is slightly over-estimated by the adjusted MPM.

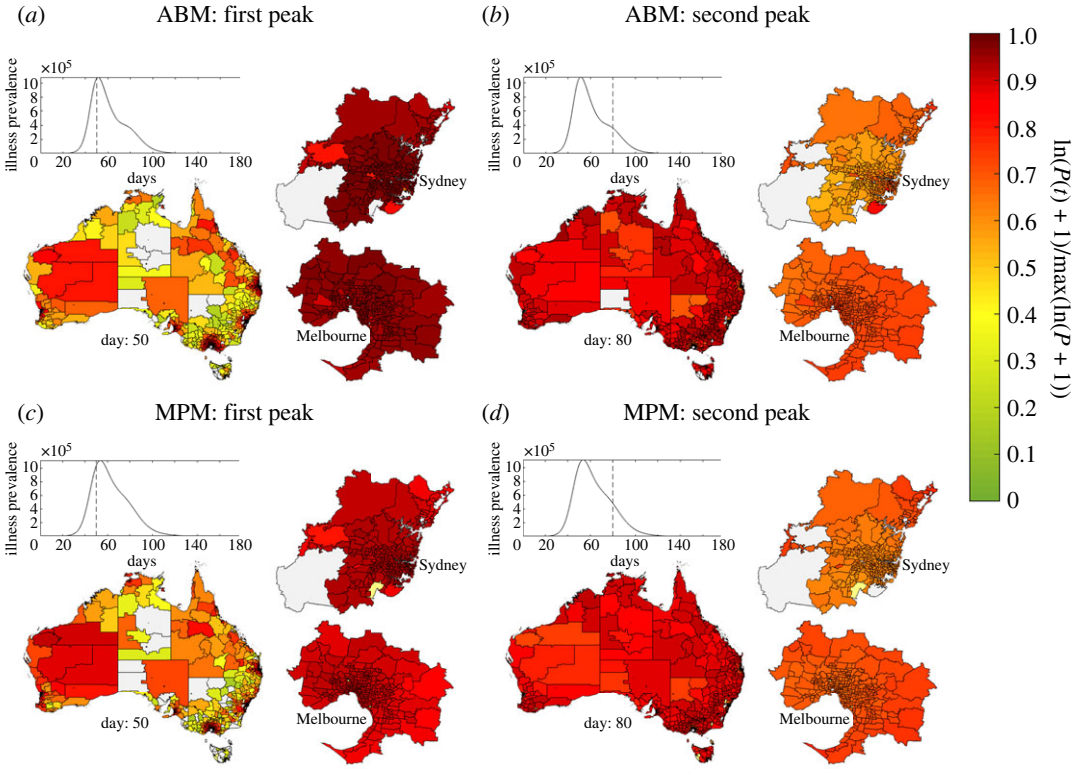


Figure 7. Spatio-temporal patterns of illness prevalence generated by the ABM (*a*,*b*) and MPM (*c*,*d*). Choropleth maps of log-scaled disease prevalence (the fraction of concurrently ill individuals in each SA2) show the spatial distribution of illness averaged over 30 instances of the ABM (*a*,*b*) and 10 instances of the extended MPM (*c*,*d*). The selected snapshots from days 50 and 80 of each simulation show the average distribution of illness during the first prevalence peak (*a*,*c*), and the second peak (*b*,*d*), respectively. The MPM results in (*c*) and (*d*) were generated using the tuned parameters $\sigma = 0.9$, $\lambda = 0.6$. Colours correspond to natural log prevalence (proportion currently ill at time t) scaled relative to the largest prevalence value recorded over all SA2 regions. In (*c*,*d*), the raw prevalence values from simulations shown in the plots of prevalence versus time were adjusted by 0.67 (the symptomatic proportion in the ABM) for comparison to the ABM prevalence plots shown in (*a*,*b*), however, no observation kernel was applied to prevalence counts.

3.3. Sensitivity analysis: SEIR formulation

To test the robustness of our results to the choice of compartmental representation, we performed a calibration and comparison of an SEIR model designed to produce equivalence to the ABM both in terms of the initial growth rate r, the basic reproductive ratio R_0 and the latency period $1/\kappa$ (1 day on average). Due to the tightly constrained calibration conditions, this comparison demonstrates nearly

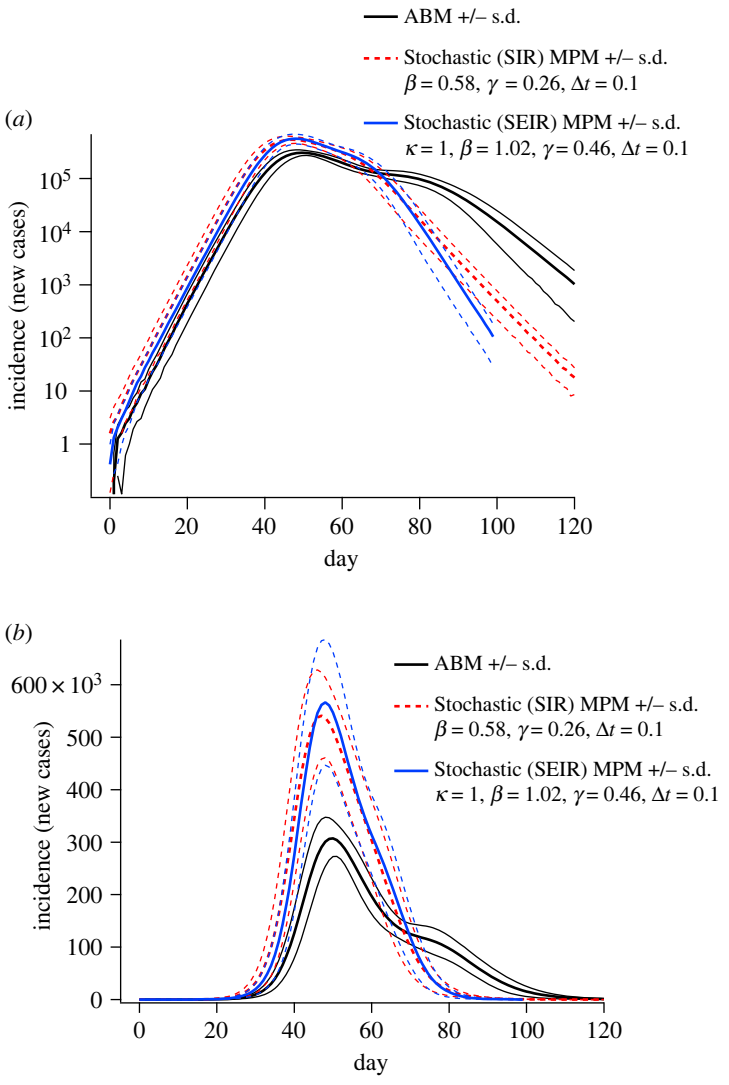


Figure 8. Comparison of SIR and SEIR dynamics, using an average latency period of one day. Note that for direct comparison between the SIR and SEIR outputs, no observation kernel has been applied to the incidence numbers in this figure. Here, the incidence numbers for SIR and SEIR formulations correspond to the number of individuals entering the infectious (I) compartment.

identical initial incidence dynamics to those of the SIR formulation (figure 8*a*), and very similar peak incidence, with a slightly later peak and slightly faster incidence decay during the epidemic tail (figure 8*b*). Overall, this analysis demonstrates that our results regarding the divergence of the compartmental MPM and ABM dynamics is likely to be robust unless the heterogeneity factors we correct for in our extended formulations are taken into account directly.

4. Discussion

A comparison between an agent-based and a structured metapopulation stochastic model, using a pandemic event in Italy simulated within the GLobal Epidemic and Mobility (GLEaM) framework [18], pointed out that the MPM consistently yielded a larger incidence than the ABM. This bias was explained by the assumption of homogeneity used in the MPM, as well as different structures in the intra-population contact patterns captured by the respective approaches. For example, the largest discrepancy between the two models was reported for the 60 + age class—a 'class with the most marked difference in household structure and workplace habits that cannot be taken into consideration in the metapopulation level' [18]. Our comparative analysis also indicated that aggregation of heterogeneity in household composition and social mixing behaviour produced a similar result. These discrepancies were reported to increase with higher values of the basic reproductive number R_0 [18],

and hence, this tendency forms an elevated concern in modelling more infectious diseases, such as COVID-19. Indeed, pandemic spread of SARS-CoV-2, and its more infectious Delta (B.1.617.2) variant, was often driven by structural factors, especially in areas characterized by socioeconomic disadvantage profiles, with high-density housing, multi-generational and shared households, higher concentrations of essential workers [26], increasing the risks of exposure within overcrowded living spaces [27], and disproportionately affecting rural and urban areas [28].

ABMs manifest both behavioural and mechanical adequacy of the disease transmission mechanism [29,30]. While 'behavioural adequacy' ensures that simulated epidemic patterns agree with the observations, 'mechanical adequacy' guarantees a concordant natural history of the disease within an individual, representing stages of the pathogenesis from exposure and incubation to the infectivity peak and then to recovery or death. In a large-scale ABM, each infected agent follows a profile which is stochastically generated from an explicitly defined distribution, thus intrinsically representing diversity of possible histories. When the natural history of disease exhibits non-trivial dependencies (e.g. with respect to pre-symptomatic or asymptomatic infectivity), an aggregation of individual agent profiles within an MPM is likely to average out nuances of the time-dependent infectiousness profile. Aggregation of these individual-level factors impedes the design of intricate targeted interventions based on nuanced timing considerations (e.g. the isolation of infectious contacts before they can infect others).

In general, one may need to consider a balance between (i) increased realism of ABMs, achieved at the cost of a more laborious calibration of internal parameters [30,31] and a detailed reconstruction of mobility patterns [21], and (ii) coarse-grained compartmental approximations of the overall epidemic dynamics, attained with substantially reduced design, calibration and computational costs. To re-iterate, there is much promise for hybrid models which enrich homogeneous approximations with some aspects of population heterogeneity [32,33], or that use individual-based models to define sensitive initial conditions [34].

However, modelling targeted interventions remains a challenge even for these hybrid approaches, since the intensity of interactions within mixing groups dynamically changes in response to various spatio-temporal constraints (social distancing, vaccinations, school closures etc.). These fine-grained changes may generate local and transient effects that can be lost during an aggregation, complicating modelling of intervention efforts focused on super-spreading events [35], epidemic risk assessment for vulnerable communities [36], and estimation of age-dependent hospitalization and fatality rates [37,38].

This work demonstrated that by introducing two global correction factors, for a total of four free parameters (β, γ, σ, λ), a simple MPM can be tuned to closely approximate the dynamics observed from a complex ABM tailored to simulate pandemic influenza H1N1 in the Australian population. The base MPM was derived from a detailed accounting of mobility between regions, to match the mixing patterns captured by the ABM. To match the ABM results, the base MPM was corrected to globally attenuate inter-region transmission, and account for local depletion of susceptible populations. The final, adjusted MPM combines a detailed accounting of mobility with a phenomenological model of disease transmission. This approach balances a key advantage of the ABM (detailed mobility patterns), with the capacity for efficient calibration and low computational cost.

In the ABM, the demographic properties of each area influence the transmission strength between pairs of regions, which suggests that a more accurate correction procedure should take into account the proportions of different sub-populations in each region (i.e. children and adults). If these demographic distributions were taken into account explicitly, the phenomenological correction factor σ could be replaced element-wise with a data-derived $n \times n$ correction matrix, modifying transmission strength between each pair of regions based on their local demographics. On the other hand, the MPM developed here requires only mobility information and global observations of disease incidence for its formulation, and could be a valuable approach when demographic information (and its relationship to disease spread) is uncertain or unavailable. The MPM formulation here uses 'perfect' mobility data, for a one-to-one match with the ground truth as specified by the ABM. In real outbreaks, mobility patterns are uncertain, but can be inferred from various sources on scales of aggregation appropriate for maintaining the privacy of individuals. Future applications of the model developed here could import near real-time mobility and case data, to account for sensitive initial conditions and population mixing patterns that change dynamically during an epidemic [39,40].

A key limitation to application of the MPM approach demonstrated here to forecasting of epidemics is the late divergence of the trajectories, a known limitation of sub-exponential growth models [41]. The calibrated base MPM closely reproduced the early dynamics of the ABM, but dramatically over-estimated peak epidemic prevalence (which translates to peak public health burden). The trajectories and standard deviation bands shown in figure 4*c*,*d* suggest that the initial MPM trajectory deviates

substantially from the ABM by approximately day 40. While still preceding the peak, this divergence is late relative to the timescale of epidemic growth, with day 40 corresponding to incidence in excess of 10^5 cases per day. Additionally, figure 5 suggests that the initial divergence from exponential growth could be accounted for only with the saturation parameter λ, but that doing so would produce an underestimate of the epidemic duration and attack rate without also adjusting the decoupling parameter σ. Unfortunately, σ alters the global incidence trajectory only after the first epidemic peak and cannot be estimated based on initial growth behaviour. Therefore, it remains unclear how to prospectively calibrate the correction factors σ and λ early in an epidemic, during the exponential growth phase when intervention decisions need to be made. To do so, a clearer understanding of the behavioural phenomena underlying these corrections is required, aided by detailed observations of local epidemic trajectories and synchronization between regions. Other possible future extensions of this work include application of the comparison to different disease models (varying R_0, generation interval distribution, incubation periods etc.), and to more detailed compartmental behavioural models including aspects such as age structure.

Unlike previous explorations that have introduced similar corrections to traditional models [14,16,17,42], this study used a detailed ABM as ground-truth. This simulated ground-truth offers the advantage that aspects of heterogeneity accounted for by the correction factors are known. The disadvantage of using simulated ground-truth is that other potentially important factors, such as risk-driven behavioural feedback, were not incorporated into the ABM. It is not clear whether the correction factors introduced here would be sufficient to capture the effects of such phenomena. What our work demonstrates clearly is that salient heterogeneity can be accounted for through simple, phenomenological corrections to traditional modelling approaches. As part of a hybrid, multi-scale modelling framework, these types of methods may prove useful in forecasting epidemic dynamics and efficiently providing evidence for or against intervention policies during future epidemics.

Data accessibility. All simulated data, simulation code and processing scripts necessary to reproduce the results reported here (including the MPM source code and inputs produced by the ABM population generator) are freely available and may be found on the associated Zenodo database (doi:10.5281/zenodo.6486795) [43]. The source code and input data used for the agent-based model population generator and epidemic simulations (ACEMod) is available in a separate database (doi:10.5281/zenodo.5773908) [44].

Authors' contributions. C.Z.: conceptualization, data curation, formal analysis, investigation, methodology, software, visualization, writing—original draft; S.C.: conceptualization, methodology, writing—review and editing; N.H.: conceptualization, methodology, writing—review and editing; M.P.: conceptualization, methodology, resources, supervision, writing—original draft, writing—review and editing.

All authors gave final approval for publication and agreed to be held accountable for the work performed therein.

Conflict of interest declaration. We declare we have no competing interest.

Funding. This work was partially supported by the Australian Research Council grants no. DP200103005 and DP220101688 (M.P. and S.C.). C.Z. is supported in part by National Health and Medical Research **Council project** grant no. APP1165876. The agent-based model used in this work (ACEMod) is registered under The University of Sydney's invention disclosure CDIP ref. 2019–123.

References

1. Kermack WO, McKendrick AG. 1927 A contribution to the mathematical theory of epidemics. *Proc. R. Soc. Lond. A* **115**, 700–721. (doi:10.1098/rspa.1927.0118)
2. Wallinga J, Lipsitch M. 2007 How generation intervals shape the relationship between growth rates and reproductive numbers. *Proc. R. Soc. B* **274**, 599–604. (doi:10.1098/rspb.2006.3754)
3. Nishiura H. 2007 Early efforts in modeling the incubation period of infectious diseases with an acute course of illness. *Emerg. Themes Epidemiol.* **4**, 1–12. (doi:10.1186/1742-7622-4-2)
4. Virlogeux V *et al.* 2015 Estimating the distribution of the incubation periods of human avian influenza A (H7N9) virus infections. *Am. J. Epidemiol.* **182**, 723–729. (doi:10.1093/aje/kwv115)
5. Ferretti L, Wymant C, Kendall M, Zhao L, Nurtay A, Abeler-Dörner L, Parker M, Bonsall D, Fraser C. 2020 Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. *Science* **368**, eabb6936. (doi:10.1126/science.abb6936)
6. Champredon D, Dushoff J, Earn DJ. 2018 Equivalence of the Erlang-distributed SEIR epidemic model and the renewal equation. *SIAM J. Appl. Math.* **78**, 3258–3278. (doi:10.1137/18M1186411)
7. Marziano V *et al.* 2021 Retrospective analysis of the Italian exit strategy from COVID-19 lockdown. *Proc. Natl Acad. Sci. USA* **118**, e2019617118. (doi:10.1073/pnas.2019617118)
8. Balcan D, Gonçalves B, Hu H, Ramasco JJ, Colizza V, Vespignani A. 2010 Modeling the spatial spread of infectious diseases: the GLobal Epidemic and Mobility computational model. *J. Comput. Sci.* **1**, 132–145. (doi:10.1016/j.jocs.2010.07.002)
9. Keeling MJ, Danon L, Vernon MC, House TA. 2010 Individual identity and movement networks for disease metapopulations. *Proc. Natl Acad. Sci. USA* **107**, 8866–8870. (doi:10.1073/pnas.1000416107)
10. Colizza V, Vespignani A. 2008 Epidemic modeling in metapopulation systems with heterogeneous coupling pattern: theory and simulations. *J. Theor. Biol.* **251**, 450–467. (doi:10.1016/j.jtbi.2007.11.028)
11. Diekmann O, Heesterbeek H, Britton T. 2012 *Mathematical tools for understanding infectious disease dynamics*, vol. 7. Princeton, NJ: Princeton University Press.
12. Gatto M, Bertuzzo E, Mari L, Miccoli S, Carraro L, Casagrandi R, Rinaldo A. 2020 Spread and

dynamics of the COVID-19 epidemic in Italy: effects of emergency containment measures. *Proc. Natl Acad. Sci. USA* **117**, 10 484–10 491. (doi:10.1073/pnas.2004978117)

13. Pastor-Satorras R, Castellano C, Van Mieghem P, Vespignani A. 2015 Epidemic processes in complex networks. *Rev. Mod. Phys.* **87**, 925. (doi:10.1103/RevModPhys.87.925)
14. Bansal S, Grenfell BT, Meyers LA. 2007 When individual behaviour matters: homogeneous and network models in epidemiology. *J. R. Soc. Interface* **4**, 879–891. (doi:10.1098/rsif.2007.1100)
15. Connell R, Dawson P, Skvortsov A. 2009 Comparison of an agent-based model of disease propagation with the generalised SIR epidemic model. Defence Science and Technology Organisation Victoria (Australia).
16. Stroud PD, Sydoriak SJ, Riese JM, Smith JP, Mniszewski SM, Romero PR. 2006 Semi-empirical power-law scaling of new infection rate to model epidemic dynamics with inhomogeneous mixing. *Math. Biosci.* **203**, 301–318. (doi:10.1016/j.mbs.2006.01.007)
17. Novozhilov AS. 2012 Epidemiological models with parametric heterogeneity: deterministic theory for closed populations. *Math. Model. Nat. Phenom.* **7**, 147–167. (doi:10.1051/mmnp/20127310)
18. Ajelli M, Gonçalves B, Balcan D, Colizza V, Hu H, Ramasco JJ, Merler S, Vespignani A. 2010 Comparing large-scale computational approaches to epidemic modeling: agent-based versus structured metapopulation models. *BMC Infect. Dis.* **10**, 190. (doi:10.1186/1471-2334-10-190)
19. Cliff OM, Harding N, Piraveenan M, Erten EY, Gambhir M, Prokopenko M. 2018 Investigating spatiotemporal dynamics and synchrony of influenza epidemics in Australia: an agent-based modelling approach. *Simul. Model. Pract. Theory* **87**, 412–431. (doi:10.1016/j.simpat.2018.07.005)
20. Zachreson C, Fair KM, Cliff OM, Harding N, Piraveenan M, Prokopenko M. 2018 Urbanization affects peak timing, prevalence, and bimodality of influenza pandemics in Australia: results of a census-calibrated model. *Sci. Adv.* **4**, eaau5294. (doi:10.1126/sciadv.aau5294)
21. Fair KM, Zachreson C, Prokopenko M. 2019 Creating a surrogate commuter network from Australian Bureau of Statistics census data. *Sci. Data* **6**, 1–14. (doi:10.1038/s41597-018-0005-2)
22. Australian Statistical Geography Standard (ASGS); 2020. [online; accessed 01-Jun-2021]. See https://www.abs.gov.au/websitedbs/D3310114.nsf/home/Australian+Statistical+Geography+Standard+(ASGS).
23. Sattenspiel L, Dietz K. 1995 A structured epidemic model incorporating geographic mobility among regions. *Math. Biosci.* **128**, 71–91. (doi:10.1016/0025-5564(94)00068-B)
24. Zachreson C, Fair KM, Harding N, Prokopenko M. 2020 Interfering with influenza: nonlinear coupling of reactive and static mitigation strategies. *J. R. Soc. Interface* **17**, 20190728. (doi:10.1098/rsif.2019.0728)
25. Ma J. 2020 Estimating epidemic exponential growth rate and basic reproduction number. *Infect. Dis. Model.* **5**, 129–141. (doi:10.1016/j.idm.2019.12.009)
26. Chang SL, Cliff OM, Zachreson C, Prokopenko M. 2021 Simulating transmission scenarios of the Delta variant of SARS-CoV-2 in Australia. (http://arxiv.org/abs/210706617).
27. Rasul G *et al.* 2021 Socio-economic implications of COVID-19 pandemic in South Asia: emerging risks and growing challenges. *Front. Sociol.* **6**, 629693. (doi:10.3389/fsoc.2021.629693)
28. Aneja R, Ahuja V. 2021 An assessment of socioeconomic impact of COVID-19 pandemic in India. *J. Public Affairs* **21**, e2266. (doi:10.1002/pa.2266)
29. Glennan S. 2005 Modeling mechanisms. *Stud. Hist. Phil. Sci. Part C* **36**, 443–464.
30. Maziarz M, Zach M. 2020 Agent-based modelling for SARS-CoV-2 epidemic prediction and intervention assessment: a methodological appraisal. *J. Eval. Clin. Pract.* **26**, 1352–1360. (doi:10.1111/jep.13459)
31. Hoertel N, Blachier M, Blanco C, Olfson M, Massetti M, Rico MS, Limosin F. 2020 A stochastic agent-based model of the SARS-CoV-2 epidemic in France. *Nat. Med.* **26**, 1417–1421. (doi:10.1038/s41591-020-1001-6)
32. Arino J, Van den Driessche P. 2003 A multi-city epidemic model. *Math. Popul. Stud.* **10**, 175–193. (doi:10.1080/08898480306720)
33. Belik V, Geisel T, Brockmann D. 2011 Natural human mobility patterns and spatial spread of infectious diseases. *Phys. Rev. X* **1**, 011001. (doi:10.1103/physrevx.1.011001)
34. Zachreson C, Shearer FM, Price DJ, Lydeamore MJ, McVernon J, McCaw J, Geard N. 2021 COVID-19 in low-tolerance border quarantine systems: impact of the Delta variant of SARS-CoV-2. (http://arxiv.org/abs/210912799).
35. Endo A, Abbott S, Kucharski AJ, Funk S. 2020 Estimating the overdispersion in COVID-19 transmission using outbreak sizes outside China. *Wellcome Open Res.* **5**, 67. (doi:10.12688/wellcomeopenres.15842.3)
36. Pluchino A, Biondo AE, Giuffrida N, Inturri G, Latora V, Le Moli R, Rapisarda A, Russo G, Zappalá C. 2021 A novel methodology for epidemic risk assessment of COVID-19 outbreak. *Sci. Rep.* **11**, 1–20. (doi:10.1038/s41598-020-79139-8)
37. Nyberg T *et al.* 2021 Risk of hospital admission for patients with SARS-CoV-2 variant B.1.1.7: cohort analysis. *BMJ* **373**, n1412. (doi:10.1136/bmj.n1412)
38. Levin AT, Hanage WP, Owusu-Boaitey N, Cochran KB, Walsh SP, Meyerowitz-Katz G. 2020 Assessing the age specificity of infection fatality rates for COVID-19: systematic review, meta-analysis, and public policy implications. *Eur. J. Epidemiol.* **35**, 1123–1138. (doi:10.1007/s10654-020-00698-1)
39. Zachreson C, Mitchell L, Lydeamore MJ, Rebuli N, Tomko M, Geard N. 2021 Risk mapping for COVID-19 outbreaks in Australia using mobility data. *J. R. Soc. Interface* **18**, 20200657. (doi:10.1098/rsif.2020.0657)
40. Gauvin L, Bajardi P, Pepe E, Lake B, Privitera F, Tizzoni M. 2020 Socioeconomic determinants of mobility responses during the first wave of COVID-19 in Italy: from provinces to neighbourhoods. *medRxiv*.
41. Chowell G. 2017 Fitting dynamic models to epidemic outbreaks with quantified uncertainty: a primer for parameter uncertainty, identifiability, and forecasts. *Infect. Dis. Model.* **2**, 379–398. (doi:10.1016/j.idm.2017.08.001)
42. Novozhilov AS. 2008 On the spread of epidemics in a closed heterogeneous population. *Math. Biosci.* **215**, 177–185. (doi:10.1016/j.mbs.2008.07.010)
43. Zachreson C, Chang S, Harding N, Prokopenko M. 2021 Supplementary code and data for 'The effects of local homogeneity assumptions in metapopulation models of infectious disease'. *Zenodo*. https://doi.org/10.5281/zenodo.6486795.
44. Cliff OM, Zachreson C, Harding N, Fair K, Prokopenko M, Piraveenan M. 2021 ACEMod (v19_12_2019) Source Code: Australian Census-based Epidemic Model. *Zenodo*. https://doi.org/10.5281/zenodo.5773908.